Contents

Original Papers

Estimating the Epidemic Size of Superspreading Coronavirus Outbreaks in Real Time: Quantitative Study (e46687)
Kitty Lau, Jian Kang, Minah Park, Gabriel Leung, Joseph Wu, Kathy Leung ................................................................. 17

Association of Rest-Activity Rhythm and Risk of Developing Dementia or Mild Cognitive Impairment in the Middle-Aged and Older Population: Prospective Cohort Study (e55211)
Shahab Haghashegh, Chenlu Gao, Elizabeth Sugg, Xi Zheng, Hui-Wen Yang, Richa Saxena, Martin Rutter, Michael Weedon, Agustin Ibanez, David Bennett, Peng Li, Lei Gao, Kun Hu ................................................................. 32

Exploring Dynamic Changes in HIV-1 Molecular Transmission Networks and Key Influencing Factors: Cross-Sectional Study (e56593)
Yan He, Ying Tang, Qun Hua, Xin Li, You Ge, Yangyang Liu, Rong Tang, Ye Tian, Wei Li ................................................................. 46

A Metric of Societal Burden Based on Virus Succession to Determine Economic Losses and Health Benefits of China’s Lockdown Policies: Model Development and Validation (e48043)
Wenxiu Chen, Bin Zhang, Chen Wang, Wei An, Shashika Guruge, Ho-kwong Chui, Min Yang ................................................................. 60

Optimal Systolic Blood Pressure for the Prevention of All-Cause and Cardiovascular Disease Mortality in Older Adults With Hypertension: Nationwide Population-Based Cohort Study (e52182)
Jae Lee, Kyung Lee .................................................................................................................................................. 70

Sleep Health Analysis Through Sleep Symptoms in 35,808 Individuals Across Age and Sex Differences: Comparative Symptom Network Study (e51585)
Christophe Gauld, Sarah Hartley, Jean-Arthur Micoulaud-Franchi, Sylvie Royant-Parola ........................................................................ 80

Early Detection of 5 Neurodevelopmental Disorders of Children and Prevention of Postnatal Depression With a Mobile Health App: Observational Cross-Sectional Study (e58565)
Fabrice Denis, Florian Le Goff, Madhu Desbois, Agnes Gepner, Guillaume Feliciano, Denise Silber, Jean-David Zeitoun, Guedalia Assuied ................................................................. 9

Predicting the Population Risk of Suicide Using Routinely Collected Health Administrative Data in Quebec, Canada: Model-Based Synthetic Estimation Study (e52773)
JianLi Wang, Fatemeh Kharrat, Geneviève Gariépy, Christian Gagné, Jean-François Pelletier, Victoria Massamba, Pascale Lévesque, Mada Mohammed, Alain Lesage ........................................................................................................ 106

Factors Influencing Data Quality in Electronic Health Record Systems in 50 Health Facilities in Rwanda and the Role of Clinical Alerts: Cross-Sectional Observational Study (e49127)
Hamish Fraser, Michael Mugisha, Ian Bacher, Joseph Ngenzi, Christopher Seebregts, Aline Umuhire, Jeanine Condo ........................................................................................................ 118
Discovering Subgroups of Children With High Mortality in Urban Guinea-Bissau: Exploratory and Validation Cohort Study (e48060)
Andreas Rieckmann, Sebastian Nielsen, Piotr Dworzynski, Heresh Amini, Søren Mogensen, Isabel Silva, Angela Chang, Onyebuchi Aran, Wojciech Samek, Naja Rod, Claus Ekstrom, Christine Benn, Peter Aaby, Ane Fisker ................................................................. 602

Data-Driven Identification of Potentially Successful Intervention Implementations Using 5 Years of Opioid Prescribing Data: Retrospective Database Study (e51323)
Lisa Hopcroft, Helen Curtis, Richard Croker, Felix Pretis, Peter Inglesby, David Evans, Sebastian Bacon, Ben Goldacre, Alex Walker, Brian MacKenna ................................................................. 618

Development of Interoperable Computable Phenotype Algorithms for Adverse Events of Special Interest to Be Used for Biologies Safety Surveillance: Validation Study (e49811)
Ashley Holdefer, Jeno Pizarro, Patrick Saunders-Hastings, Jeffrey Beers, Arianna Sang, Aaron Hettinger, Joseph Blumenthal, Eírík Martínez, Lance Jones, Matthew Deady, Hussein Ezzeldin, Steven Anderson ................................................................. 628

Global, Regional, and National Burden of Myocarditis in 204 Countries and Territories From 1990 to 2019: Updated Systematic Analysis (e46635)
Qingyu Kong, Xue Xu, Meng Li, Xiao Meng, Cuifen Zhao, Xiaorong Yang ................................................................. 684

Challenges and Implications for Menopausal Health and Help-Seeking Behaviors in Midlife Women From the United States and China in Light of the COVID-19 Pandemic: Web-Based Panel Surveys (e46538)
Bobo Lau, Catherine Tang, Eleanor Holroyd, William Wong ................................................................. 704

The Framing Effect of Digital Textual Messages on Uptake Rates of Medical Checkups: Field Study (e45379)
Amnon Maltz, Stella Rashkovich, Adi Sarid, Yafit Cohen, Tamar Landau, Elina Saifer, Neta Amoral Belkin, Tamar Alcalay ................................................................. 718

Association Between Adherence to 24-Hour Movement Guidelines and Noncommunicable Disease Risk in Chinese Adults: Prospective Cohort Study (e47517)
Si Yi Huang, Yuxuan Gu, Shahmir Ali, Jingjing Xue, Ronghua Zhang, Xu Wen ................................................................. 758

Evaluating the Knowledge of and Behavior Toward COVID-19 and the Possibility of Isolating at a City Level: Survey Study (e47170)
Elise Verot, Robin Chaux, Julie Gagnaire, Paul Bonjean, Amandine Gagneux-Brunon, Philippe Berthelot, Carole Pelissier, Billal Boulama, Franck Chauvin, Bruno Pozzetto, Elisabeth Botelho-Nevers ................................................................. 771

Relationship Between Physical Exercise and Cognitive Function Among Older Adults in China: Cross-Sectional Population-Based Study (e49790)
Fubaihui Wang, Changqing Gao, Yantao Wang, Zhuo Li, Feiran Zheng, Yanan Luo ................................................................. 802

Impact of Healthy Lifestyle Factors on Life Expectancy and Lifetime Health Care Expenditure: Nationwide Cohort Study (e57045)
Wei-Cheng Lo, Tsuey-Hwa Hu, Cheng-Yu Shih, Hsien-Ho Lin, Jing-Shiang Hwang ................................................................. 836

Advancing Colorectal Cancer Detection With Blood-Based Tests: Qualitative Study and Discrete Choice Experiment to Elicit Population Preferences (e53200)
Clarence Ong, Alex Cook, Ker-Kan Tan, Yi Wang ................................................................. 854

Newspaper Coverage of Hospitals During a Prolonged Health Crisis: Longitudinal Mixed Methods Study (e48134)
Frank van de Baan, Rachel Gifford, Dirk Ruwaard, Bram Fleuren, Daan Westra ................................................................. 871

Web-Based Information on Spinal Cord Stimulation: Qualitative Assessment of Publicly Accessible Online Resources (e50031)
Tiev Miller, Ali Hossinzadeh, Thomas Thordarson, Tamila Kalimullina, Soshi Samejima, Claire Shackleton, Raza Malik, Martin Calderón-Juárez, Rahul Sachdeva, Andrei Krassioukov ................................................................. 884
Impacts of Smoking Ban Policies on Billiard Hall Sales in South Korea Using Objective Sales Information of a Credit Card Company: Quasi-Experimental Study (e50466)  
Jin-Won Noh, Jooyoung Cheon, Hohyun Seong, Young Kwon, Ki-Bong Yoo.  
895

CD4/CD8 Ratio Recovered as a Predictor of Decreased Liver Damage in Adults Infected With HIV: 16-Year Observational Cohort Study (e45818)  
Bingyu Liang, Rujing Sun, Yan Yan Liao, Aidan Nong, Jinfeng He, Mengqiu Zhu, Hui Hu, Li Ye, Hao Liang.  
917

AIDSVu Cities’ Progress Toward HIV Care Continuum Goals: Cross-Sectional Study (e49381)  
Nicole Hood, Nanette Benbow, Chandni Jaggi, Shamaya Whitby, Patrick Sullivan, AIDSVu Health Department Collaborators.  
929

Characteristics of Hepatitis B Virus, Hepatitis C Virus, and Syphilis Coinfection in People With HIV/AIDS Contracted Through Different Sources: Retrospective Study (e46750)  
Rongrong Yang, Rui Yuan, Xien Gui, Hengning Ke, Ke Zhuang, Hui Hu, Ling Li, Ling Feng, Xingxia Yu, Yajun Yan, Mingqi Luo.  
939

SARS-CoV-2 Infection, Hospitalization, and Associated Factors Among People Living With HIV in Southeastern China From December 2022 to February 2023: Cross-Sectional Survey (e51449)  
Wei Cheng, Yun Xu, Haibo Jiang, Jun Li, Zhigang Hou, Haibin Meng, Wei Wang, Chengliang Chai, Jianmin Jiang.  
951

A Machine Learning Model for Identifying Sexual Health Influencers to Promote the Secondary Distribution of HIV Self-Testing Among Gay, Bisexual, and Other Men Who Have Sex With Men in China: Quasi-Experimental Study (e50656)  
Yuxin Ni, Ying Lu, Fengshi Jing, Tianyuan Wang, Yewei Xie, Xi He, Dan Wu, Rayner Tan, Joseph Tucker, Xumeng Yan, Jason Ong, Qingpeng Zhang, Hongbo Jiang, Wencan Dai, Liqun Huang, Wenhua Mei, Yi Zhou, Weiming Tang.  
971

Attrition Rates in HIV Viral Load Monitoring and Factors Associated With Overdue Testing Among Children Within South Africa’s Antiretroviral Treatment Program: Retrospective Descriptive Analysis (e40796)  
Ahmad Haeri Mazanderani, Lebohang Radebe, Gayle Sherman.  
986

Examining HIV Testing Coverage and Factors Influencing First-Time Testing Among Men Who Have Sex With Men in Zhejiang Province, China: Cross-Sectional Study Based on a Large Internet Survey (e56906)  
Lin He, Tingting Jiang, Wanjun Chen, Shaoqiang Jiang, Jinlei Zheng, Weiyoug Chen, Hui Wang, Qiaoqin Ma, Chengliang Chai.  
999

Sex Differences in Clustering Unhealthy Lifestyles Among Survivors of COVID-19: Latent Class Analysis (e50189)  
Lan Le, Thi Hoang, Tan Nguyen, Tien Dao, Binh Do, Khue Pham, Vinh Vu, Linh Pham, Lien Nguyen, Hoang Nguyen, Tuan Tran, Trung Nguyen, Anh Nguyen, Hoan Nguyen, Phuc Nguyen, Hoai Nguyen, Thu Pham, Thuy Le, Thao Nguyen, Cuong Tran, Ha-Linh Quach, Kien Nguyen, Tuyen Duong.  
1022

Digital Dietary Behaviors in Individuals With Depression: Real-World Behavioral Observation (e47428)  
Yue Zhu, Ran Zhang, Shuluo Yin, Yihui Sun, Fay Women, Rongxun Liu, Sheng Zeng, Xizhe Zhang, Fei Wang.  
1039

Understanding Risk Factors for Oropharyngeal Gonorrhea Among Sex Workers Attending Sexual Health Clinics in 2 Australian Cities: Mixed Methods Study (e46845)  
Tiffany Phillips, Christopher Fairley, Kate Maddaford, Anna McNulty, Basil Donovan, Rebecca Gu, Ruth McIver, Rebecca Wigan, Rick Varma, Jason Ong, Denton Callander, Gabrielle Skelsey, My Myon Myong, Dylan O’Hara, Jade Bilardi, Eric Chow.  
1051

The Impact of Rapid Handpump Repairs on Diarrhea Morbidity in Children: Cross-Sectional Study in Kwale County, Kenya (e42462)  
Patrick Thomson, Justin Stoler, Michelle Byford, David Bradford.  
1117
Smoking Behavior Change and the Risk of Heart Failure in Patients With Type 2 Diabetes: Nationwide Retrospective Cohort Study (e46450)
Jung Yoo, Su-Min Jeong, Kyu Lee, Heesun Lee, Ji Yoon, Kyungdo Han, Dong Shin. 1307

The Impact of Prostate-Specific Antigen Screening on Prostate Cancer Incidence and Mortality in China: 13-Year Prospective Population-Based Cohort Study (e47161)
Xiaohao Ruan, Ning Zhang, Dawei Wang, Jingyi Huang, Jinlun Huang, Da Huang, Tsun Chun, Brian Ho, Ada Ng, James Tsu, Yongle Zhan, Rong Na. 1321

Assessing Global, Regional, and National Time Trends and Associated Risk Factors of the Mortality in Ischemic Heart Disease Through Global Burden of Disease 2019 Study: Population-Based Study (e46821)
Tingting Shu, Ming Tang, Bo He, Xiaozhu Liu, Yu Han, Chang Liu, Pedro Jose, Hongyong Wang, Qing-Wei Zhang, Chunyu Zeng. 1333

Depression and Anxiety Symptoms Among Cisgender Gay and Bisexual Men During the Onset of the COVID-19 Pandemic: Time Series Analysis of a US National Cohort Study (e47048)
Chloe Mirzayi, Drew Westmoreland, Matthew Stief, Christian Grov. 1353

Global Trends in the Incidence of Anxiety Disorders From 1990 to 2019: Joinpoint and Age-Period-Cohort Analysis Study (e49609)
Huiru Cao, Yang Wu, Hui Yin, Yanqi Sun, Hui Yuan, Mengjun Tao. 1361

Relationship Between Lipoprotein(a), Renal Function Indicators, and Chronic Kidney Disease: Evidence From a Large Prospective Cohort Study (e50415)
Yingxin Liu, Ruoting Wang, Shuai Li, Changfa Zhang, Gregory Lip, Lehana Thabane, Guowei Li. 1376

Impact of the COVID-19 Pandemic on the Personal Networks and Neurological Outcomes of People With Multiple Sclerosis: Cross-Sectional and Longitudinal Case-Control Study (e45429)
Clare Riley, Shruthi Venkatesh, Amar Dhand, Nandini Doshi, Katelyn Kavak, Elle Levit, Christopher Perrone, Bianca Weinstock-Guttman, Erin Longbrane, Philip De Jager, Zongqi Xia. 1391

Association of Visceral Obesity Indices With Incident Diabetic Retinopathy in Patients With Diabetes: Prospective Cohort Study (e48120)
Jiaheng Chen, Yu Li, Zimin Niu, Zhanpeng He, Yao Xie, Jose Hernandez, Wenyong Huang, Harry Wang, Guangzhou Diabetic Eye Study Group. 1410

Age Differences in the Association of Sleep Duration Trajectory With Cancer Risk and Cancer-Specific Mortality: Prospective Cohort Study (e50836)
Chenan Liu, Qingsong Zhang, Chenning Liu, Tong Liu, Mengmeng Song, Qi Zhang, Haijun Xie, Shiqi Lin, Jiangshan Ren, Yue Chen, Xin Zheng, Jinyu Shi, Li Deng, Hanping Shi, Shouling Wu. 1425

Arterial Stiffness and Obesity as Predictors of Diabetes: Longitudinal Cohort Study (e46088)
Canccan Cui, Zhiyuan Wu, Jia Zheng, Jiaqi Chu, Jiajia Xu, Yitian Qi, Zhenming Zhang, Zhihui Wang, Mengchao Zhang, Zhanhao Mo, Liangkai Cheng, Lei Ding, Xiaoping Kang, Lin Liu, Te Zhang. 1435

The Effect of the COVID-19 Pandemic on Non–COVID-19 Deaths: Population-Wide Retrospective Cohort Study (e41792)
Abraham Wai, Tsz Yip, Yui Wong, Chun Chu, Teddy Lee, Ken Yu, Wang So, Janet Wong, Carlos Wong, Joshua Ho, Timothy Rainer. 1448

Exploration of the Healthy Donor Effect Among 0.6 Million Blood Donors in China: Longitudinal Study (e48617)
Shu Su, Yang Sun, Xiaoyun Gu, Wenjie Wu, Xiaodong Su, Ting Ma, Aowei Song, Xinxin Xie, Liqin Wang, Qianke Cheng, Lingxia Guo, Lei Zhang, Jiangcun Yang. 1461
The Use of ICD-9-CM Coding to Identify COVID-19 Diagnoses and Determine Risk Factors for 30-Day Death Rate in Hospitalized Patients in Italy: Retrospective Study (e44062)
Barbara Giordani, Alessandra Burgio, Francesco Grippo, Alessandra Barone, Erica Eugeni, Giovanni Baglio. 1473

Effects of Face Mask Mandates on COVID-19 Transmission in 51 Countries: Retrospective Event Study (e49307)
Anatol-Fiete Näher, Matthias Schulte-Althoff, Marvin Kopka, Felix Balzer, Francisco Pozo-Martin. 1484

Temporal Trends of Asthma Among Children in the Western Pacific Region From 1990 to 2045: Longitudinal Observational Study (e55327)
Cheng-hao Yang, Xin-yu Li, Jia-jie Lv, Meng-jie Hou, Ru-hong Zhang, Hong Guo, Chu Peng. 1496

Investigating Nonspecific Effects of the Live-Attenuated Japanese Encephalitis Vaccine on Lower Respiratory Tract Infections in Children Aged 25-35 Months: Retrospective Cohort Study (e53040)
Siyi Zhan, Hongbo Lin, Yingying Yang, Tao Chen, Sheng Mao, Chuanxi Fu. 1513

Association Between Long-Term Exposure to Air Pollution and the Rate of Mortality After Hip Fracture Surgery in Patients Older Than 60 Years: Nationwide Cohort Study in Taiwan (e46591)
Shu-Han Chuang, Yi-Jie Kuo, Shu-Wei Huang, Han-Wei Zhang, Hsiao-Ching Peng, Yu-Pin Chen. 1526

Individual and Geospatial Determinants of Health Associated With School-Based Human Papillomavirus Immunization in Alberta: Population-Based Cohort Study (e45508)
Jennifer Malkin, Genevieve Jessiman-Perreault, Amanda Alberga Machado, Gary Teare, Joanne Snider, Syed Tirmizi, Erik Youngson, Ting Wang, Jessica Law, Thilina Bandara, Mika Rathwell, Cordell Neudorf, Lisa Allen Scott. 1549

Timely Pulmonary Tuberculosis Diagnosis Based on the Epidemiological Disease Spectrum: Population-Based Prospective Cohort Study in the Republic of Korea (e47422)
Yousang Ko, Jae Park, Jinsoo Min, Hyung Kim, Hyeon-Kyoung Koo, Jee Oh, Yun-Jeong Jeong, Eunhye Lee, Bumhee Yang, Ju Kim, Sung-Soon Lee, Yunhyung Kwon, Jiyeon Yang, Ji Han, You Jang, Jinsob Kim. 1565

Preeclampsia Onset, Days to Delivery, and Autism Spectrum Disorders in Offspring: Clinical Birth Cohort Study (e47396)
Sarah Carter, Jane Lin, Ting Chow, Mayra Martinez, Chunyuan Qiu, R Feldman, Rob McConnell, Anny Xiang. 1579

Mutual Associations of Exposure to Ambient Air Pollutants in the First 1000 Days of Life With Asthma/Wheezing in Children: Prospective Cohort Study in Guangzhou, China (e52456)
Fenglin Tian, Xingxi Zhong, Yufeng Ye, Xiaohan Liu, Guanhaoyi He, Culing Wu, Zhiqing Chen, Qijiong Zhu, Siwen Yu, Jingjie Fan, Huan Yao, Wenjun Ma, Xiaomei Dong, Tao Liu. 1589

Defining the Subtypes of Long COVID and Risk Factors for Prolonged Disease: Population-Based Case-Crossover Study (e49841)
Skyler Resendez, Steven Brown, Hugo Ruiz Ayala, Prahalad Rangan, Jonathan Nebecker, Diane Montella, Peter Elkin. 1604

Association of the Type of Public Pension With Mental Health Among South Korean Older Adults: Longitudinal Observational Study (e49129)
Seung Kim, Hyunkyu Kim, Sung Jeong, Eun-Cheol Park. 1617

Risk of Subsequent Primary Cancers Among Adult-Onset 5-Year Cancer Survivors in South Korea: Retrospective Cohort Study (e48380)
Yoon Choi, Myeonghee Lee, Eun Kim, Jae Lee, Inkyung Jung, Jae-Ho Cheong. 1631

Metabolic Score for Insulin Resistance and New-Onset Type 2 Diabetes in a Middle-Aged and Older Adult Population: Nationwide Prospective Cohort Study and Implications for Primary Care (e49617)
Hui Cheng, Zhishui Jia, Yu Li, Xiao Yu, Jia Wang, Yao Xie, Jose Hernandez, Harry Wang. 1644
The Incidence and Outcomes of Out-of-Hospital Cardiac Arrest During the COVID-19 Pandemic in South Korea: Multicenter Registry Study (e52402)
Heekyung Lee, Jae-hoon Oh, Hyuk Choi, Hyungoo Shin, Yongil Cho, Juncheol Lee. ................................................................. 1658

Association Between Socioeconomic Inequalities in Pain and All-Cause Mortality in the China Health and Retirement Longitudinal Study: Longitudinal Cohort Study (e54309)
Zhuo Zhang, Dongmei Xue, Ying Bian. ......................................................... 1669

A Multimorbidity Analysis of Hospitalized Patients With COVID-19 in Northwest Italy: Longitudinal Study Using Evolutionary Machine Learning and Health Administrative Data (e52353)
Dayana Benny, Mario Giacobini, Alberto Catalano, Giuseppe Costa, Roberto Gnavi, Fulvio Ricceri. .................................................. 1680

Spatiotemporal Pattern and Its Determinants for Newly Reported HIV/AIDS Among Older Adults in Eastern China From 2004 to 2021: Retrospective Analysis Study (e51172)
Gang Huang, Wei Cheng, Yun Xu, Jiezhe Yang, Jun Jiang, Xiaohong Pan, Xin Zhou, Jianmin Jiang, Chengliang Chai. ................................................. 1702

Association of Child Growth Failure Indicators With Household Sanitation Practices in India (1998-2021): Spatiotemporal Observational Study (e41567)
Lovely Jain, Sreya Pradhan, Arun Aggarwal, Bijaya Padhi, Ramaiah Itumalla, Mahalaqua Khatib, Shilpa Gadhane, Quazi Zahiruddin, Celso Santos, Khalid AL-Mugheed, Tahani Alrahbeni, Neelima Kukreti, Prakashini Satapathy, Sarvesh Rustagi, Petra Heidler, Roy Marzo. ........................................... 1716

Spatiotemporal Analysis of HIV/AIDS Incidence in China From 2009 to 2019 and Its Association With Socioeconomic Factors: Geospatial Study (e56229)
Ziyi Xie, Bowen Chen, Zhizhuang Duan. ..................................................... 1734

Pulmonary Tuberculosis Notification Rate Within Shenzhen, China, 2010-2019: Spatial-Temporal Analysis (e57209)
Peixuan Lai, Weicong Cai, Lin Qu, Chuangyue Hong, Kailin Lin, Weiguo Tan, Zhiguang Zhao. ......................................................... 1748

Spatiotemporal Epidemiological Trends of Mpox in Mainland China: Spatiotemporal Ecological Comparison Study (e57807)
Shuli Ma, Jie Ge, Lei Qin, Xiaoting Chen, Linlin Du, Yanbo Qi, Li Bai, Yunfeng Han, Zhiping Xie, Jiaxin Chen, Yuehui Jia. ........................................... 1762

Association Between Nitrogen Dioxide Pollution and Cause-Specific Mortality in China: Cross-Sectional Time Series Study (e44648)
Jie Zeng, Guozhen Lin, Hang Dong, Mengmeng Li, Honglian Ruan, Jun Yang. ................................................................. 1790

Eating Habits and Lifestyle Factors Related to Childhood Obesity Among Children Aged 5-6 Years: Cluster Analysis of Panel Survey Data in Korea (e51581)
Heemoon Lim, Hyejung Lee. ................................................................. 1804

Association Between Early-Life Exposure to Antibiotics and Development of Child Obesity: Population-Based Study in Italy (e51734)
Anna Cantarutti, Paola Rescigno, Claudia Da Borso, Joaquin Gutierrez de Rubalcava Doblas, Silvia Bressan, Elisa Barbieri, Carlo Giaquinto, Cristina Canova. ................................................................. 1816

Predicting the Transition to Metabolically Unhealthy Obesity Among Young Adults With Metabolically Healthy Obesity in South Korea: Nationwide Population-Based Study (e52103)
HyunHae Lee, Ji-Su Kim, HyeRin Shin. ................................................................. 1828

A Comprehensive Youth Diabetes Epidemiological Data Set and Web Portal: Resource Development and Case Studies (e53330)
Catherine McDonough, Yan Li, Nita Vangeepuram, Bian Liu, Gaurav Pandey. ................................................................. 1840
Effectiveness, Safety, and Acceptability of Primaquine Mass Drug Administration in Low-Endemicity Areas in Southern Thailand: Proof-of-Concept Study (e51993)  

Participatory Disease Surveillance for the Early Detection of Cholera-Like Diarrheal Disease Outbreaks in Rural Villages in Malawi: Prospective Cohort Study (e49539)  
Mariana Valerio, Beverly Laher, John Phuka, Guilherme Lichand, Daniela Paolotti, Onicio Neto. 1891

Estimation of the Population Size of Street- and Venue-Based Female Sex Workers and Sexually Exploited Minors in Rwanda in 2022: 3-Source Capture-Recapture (e50743)  
Elysee Tuyishime, Eric Remera, Catherine Kayitesi, Samuel Malamba, Beata Sangwayire, Ignace Habimana Kabano, Horacio Ruisenor-Escudero, Tom Olouch, Angela Umbo Chukwu. 1905

Triangulating Truth and Reaching Consensus on Population Size, Prevalence, and More: Modeling Study (e48738)  
Ian Fellows, Carl Corcoran, Anne McIntyre. 1917

Leveraging Routinely Collected Program Data to Inform Extrapolated Size Estimates for Key Populations in Namibia: Small Area Estimation Study (e48963)  
Talia Loeb, Kalai Willis, Frans Velishavo, Daniel Lee, Amrita Rao, Stefan Baral, Katherine Rucinski. 1930

Estimated Number of Injection-Involved Overdose Deaths in US States From 2000 to 2020: Secondary Analysis of Surveillance Data (e49527)  
Eric Hall, Patrick Sullivan, Heather Bradley. 1943

Out-of-Hospital Cardiac Arrest Before and During the COVID-19 Pandemic in Hong Kong: Registry-Based Study From 2017 to 2023 (e56054)  
Richard Xu, Ruiqi Sun, Siu-Ngor Fu. 1957

Brady Bennett, Stephanie DuBose, Ya-Lin Huang, Christopher Johnson, Karen Hoover, Jeffrey Wiener, David Purcell, Patrick Sullivan. 1972

Investigating the Interrelationships Among Mental Health, Substance Use Disorders, and Suicidal Ideation Among Lesbian, Gay, and Bisexual Adults in the United States: Population-Based Statewide Survey Study (e48776)  
Alex Chan, Hon Tam, Florence Wong, Gordon Wong, Lok Leung, Jacqueline Ho, Patrick Tang, Elsie Yan. 2028

Improving the Efficiency of Inferences From Hybrid Samples for Effective Health Surveillance Studies: Comprehensive Review of Quantitative Methods (e48186)  
Mansour Fahimi, Elizabeth Hair, Elizabeth Do, Jennifer Kreslake, Xiaolu Yan, Elisa Chan, Frances Barlas, Abigail Giles, Larry Osborn. 2039

Projected Time for the Elimination of Cervical Cancer Under Various Intervention Scenarios: Age-Period-Cohort Macrosimulation Study (e46360)  
Yi-Chu Chen, Yun-Yuan Chen, Shih-Yung Su, Jing-Rong Jhuang, Chun-Ju Chiang, Ya-Wen Yang, Li-Ju Lin, Chao-Chun Wu, Wen-Chung Lee. 2051

Predicting Lung Cancer Survival to the Future: Population-Based Cancer Survival Modeling Study (e46737)  
Fan-Tsui Meng, Jing-Rong Jhuang, Yan-Teng Peng, Chur-Ju Chiang, Ya-Wen Yang, Chi-Yen Huang, Kuo-Ping Huang, Wen-Chung Lee. 2072

Early Results of an Initiative to Assess Exposure to Firearm Violence in Ambulatory Care: Descriptive Analysis of Electronic Health Record Data (e47444)  
Nicole Cook, Megan Hoopes, Frances Biel, Natalie Cartwright, Michelle Gordon, Marion Sills. 2072
Revealing the Mysteries of Population Mobility Amid the COVID-19 Pandemic in Canada: Comparative Analysis With Internet of Things–Based Thermostat Data and Google Mobility Insights (e46903)
Kirti Sahu, Joel Dubin, Shannon Majowicz, Sam Liu, Plinio Morita. ................................................................. 2079

Search Engines and Generative Artificial Intelligence Integration: Public Health Risks and Recommendations to Safeguard Consumers Online (e53086)
Amir Ashraf, Tim Mackey, András Fittler. ................................................................. 2093

Ambulance Services Attendance for Mental Health and Overdose Before and During COVID-19 in Canada and the United Kingdom: Interrupted Time Series Study (e46029)
Graham Law, Rhiannon Cooper, Melissa Pirrie, Richard Ferron, Brent McLeod, Robert Spaight, A Siriwardena, Gina Agarwal, UK Canada Emergency Calls Data Analysis and GEospatial Mapping (EDGE) Consortium. ................................................................. 2104

Epidemiological Characteristics of Varicella in Anhui Province, China, 2012-2021: Surveillance Study (e50673)
Kun Xuan, Ning Zhang, Tao Li, Xingya Pang, Qingru Li, Tianming Zhao, Binbing Wang, Zhenqiu Zha, Jihai Tang. ................................................................. 2135

Epidemic Characteristics and Meteorological Risk Factors of Hemorrhagic Fever With Renal Syndrome in 151 Cities in China From 2015 to 2021: Retrospective Analysis (e52221)
Yizhe Luo, Longyao Zhang, Yameng Xu, Qiyuan Kuai, Wenhao Li, Yifan Wu, Licheng Liu, Jiaron Ren, Lingling Zhang, Qiufang Shi, Xiaobo Liu, Weilong Tan. ................................................................. 2149

Population Behavior Changes Underlying Phasic Shifts of SARS-CoV-2 Exposure Settings Across 3 Omicron Epidemic Waves in Hong Kong: Prospective Cohort Study (e51498)
Chin Chan, Shui Lee, Tsz Kwan, Samuel Wong, Eng-Kiong Yeoh, Ngai Wong. ................................................................. 2162

A Prediction Model for Identifying Seasonal Influenza Vaccination Uptake Among Children in Wuxi, China: Prospective Observational Study (e56064)
Qiang Wang, Liqiong Yang, Shixin Xu, Yuan Shen, Hui Jin, Leesa Lin. ................................................................. 2179

The Migrant-Local Difference in the Relationship Between Social Support, Sleep Disturbance, and Loneliness Among Older Adults in China: Cross-Sectional Study (e49253)
Mingli Pang, Jieru Wang, Mingyue Zhao, Rui Chen, Hui Liu, Xixing Xu, Shixue Li, Fanlei Kong. ................................................................. 2190

Impact of the Narcotics Information Management System on Opioid Use Among Outpatients With Musculoskeletal and Connective Tissue Disorders: Quasi-Experimental Study Using Interrupted Time Series (e47130)
Iyn-Hyang Lee, So Kim, Susin Park, Jae Ryu, Nam Je. ................................................................. 2204

Use of Sentinel Surveillance Platforms for Monitoring SARS-CoV-2 Activity: Evidence From Analysis of Kenya Influenza Sentinel Surveillance Data (e50799)
Daniel Owusu, Linus Ndegwa, Jorim Ayugi, Peter Kiruthia, Rosalia Kalani, Mary Okeyo, Nancy Otieno, Gilbert Kikwai, Bonventure Juma, Peninah Munyua, Francis Kuria, Emmanuel Okunga, Ann Moen, Gideon Emukule. ................................................................. 2216

Development of New Stringency Indices for Nonpharmacological Social Distancing Policies Implemented in Korea During the COVID-19 Pandemic: Random Forest Approach (e47099)
Catherine Apio, Kyulhee Han, Dooen Lee, Bogeom Lee, Taesung Park. ................................................................. 2229

Impact of the COVID-19 Pandemic on Health Care Utilization in the Vaccine Safety Datalink: Retrospective Cohort Study (e48159)
Lei Qian, Lina Sy, Vennis Hong, Sungching Glenn, Denison Ryan, Jennifer Nelson, Simon Hambidge, Bradley Crane, Ousseny Zerbo, Malini DeSilva, Jason Glanz, James Donahue, Elizabeth Liles, Jonathan Duffy, Stanley Xu. ................................................................. 2247
Impact of the COVID-19 Pandemic on People Living With Rare Diseases and Their Families: Results of a National Survey (e48430)
Maurizio Macaluso, Marc Rothenberg, Thomas Ferkol, Pierce Kuhnell, Henry Kaminski, David Kimberlin, Michael Benatar, Mirna Chehade, The Principal Investigators of the Rare Diseases Clinical Research Network – Cycle 4 .......................................................... 2257

Perspectives of European Patient Advocacy Groups on Volunteer Registries and Vaccine Trials: VACCELERATE Survey Study (e47241)

The Changing Landscape of Respiratory Viruses Contributing to Hospitalizations in Quebec, Canada: Results From an Active Hospital-Based Surveillance Study (e40792)
Rodica Gilca, Rachid Amini, Sara Carazo, Radhouene Doggui, Charles Frenette, Guy Boivin, Hugues Charest, Jeannot Dumaresq .......................................................... 2289

Updated Surveillance Metrics and History of the COVID-19 Pandemic (2020-2023) in Latin America and the Caribbean: Longitudinal Trend Analysis (e44398)
Lori Post, Scott Wu, Alan Soetikno, Egon Ozer, Yingxuan Liu, Sarah Welch, Claudia Hawkins, Charles Moss, Robert Murphy, Maryann Mason, Robert Havey, Alexander Lundberg .......................................................... 2306

The Impact of Wireless Emergency Alerts on a Floating Population in Seoul, South Korea: Panel Data Analysis (e43554)
Sungwook Yoon, Hyungsoo Lim, Sungho Park .......................................................... 2326

Assessing SARS-CoV-2 Testing Adherence in a University Town: Recurrent Event Modeling Analysis (e48784)
Yury García, Alec Schmidt, Leslie Solis, Maria Daza-Torres, J Montesinos-López, Brad Pollock, Miriam Nuño .......................................................... 2335

Persistence and Variation of the Indirect Effects of COVID-19 Restrictions on the Spectrum of Notifiable Infectious Diseases in China: Analysis of National Surveillance Among Children and Adolescents From 2018 to 2021 (e47626)
Li Chen, Liping Wang, Yi Xing, Junqing Xie, Binbin Su, Mengjie Geng, Xiang Ren, Yi Zhang, Jieyu Liu, Tao Ma, Manman Chen, Jessica Miller, Yanhui Dong, Yi Song, Jun Ma, Susan Sawyer .......................................................... 2362

Exploring the Association Between Structural Racism and Mental Health: Geospatial and Machine Learning Analysis (e52691)
Fahimeh Mohebbi, Amir Forati, Lucas Torres, Terri deRoon-Cassini, Jennifer Harris, Carissa Tomas, John Mantsch, Rina Ghose .......................................................... 2380

An Accessible Web-Based Survey to Monitor the Mental Health of People With Mild Intellectual Disability or Low Literacy Skills During the COVID-19 Pandemic: Comparative Data Analysis (e44827)
Monique Koks-Leensen, Anouk Menko, Fieke Raaijmakers, Gerdine Fransen-Kuppens, Kirsten Bevelander .......................................................... 2392

Ethnic Disparities in COVID-19 Vaccine Mistrust and Receipt in British Columbia, Canada: Population Survey (e48466)
Bushra Mahmood, Prince Adu, Geoffrey McKee, Aamir Bharma, James Wilton, Naveed Janjua .......................................................... 2419

Durability of the Effectiveness of Heterologous COVID-19 Vaccine Regimens in Thailand: Retrospective Cohort Study Using National Registration Data (e48255)
Ponlagrit Kumwichar, Chittawan Poonsiri, Siobhan Botwright, Natchalaikorn Sinchumroonwitt, Bootsakorn Lohanjun, Supharerak Thawillarp, Nontawit Cheewaruangroj, Amorn Chokchaisiripakdee, Yot Teerawattananon, Virasakdi Chongsuvivatwong .......................................................... 2434

Predicting COVID-19 Vaccination Uptake Using a Small and Interpretable Set of Judgment and Demographic Variables: Cross-Sectional Cognitive Science Study (e47979)
Nicole Vike, Sunra Bari, Leandros Stefanopoulos, Shamal Lalvani, Byoung Kim, Nicos Maglaveras, Martin Block, Hans Breiter, Aggelos Katsaggelos .......................................................... 2449
Motivators and Demotivators for COVID-19 Vaccination Based on Co-Occurrence Networks of Verbal Reasons for Vaccination Acceptance and Resistance: Repetitive Cross-Sectional Surveys and Network Analysis (e50958)
Qiuyan Liao, Jiehu Yuan, Irene Wong, Michael Ni, Benjamin Cowling, Wendy Lam. ................................................................. 2474

Ischemic Stroke After Bivalent COVID-19 Vaccination: Self-Controlled Case Series Study (e53807)
Stanley Xu, Lina Sy, Vennis Hong, Kimberly Holmquist, Lei Qian, Paddy Farrington, Katia Bruxvoort, Nicola Klein, Bruce Fireman, Bing Han, Bruno Lewin. ................................................................. 2490

Adverse Events of COVID-19 Vaccines in the United States: Temporal and Spatial Analysis (e51007)
Yiming Li, Jianfu Li, Yifang Dang, Yong Chen, Cui Tao. ................................................................. 2503

The Effectiveness of National Expanded Program on Immunization With Hepatitis A Vaccines in the Chinese Mainland: Interrupted Time-Series Analysis (e53982)
Bo-Wen Ming, Li Li, Hao-Neng Huang, Jia-Jun Ma, Chen Shi, Xiao-Han Xu, Zhou Yang, Chun-Quan Ou. ................................................................. 2547

Laboratory Data Timeliness and Completeness Improves Following Implementation of an Electronic Laboratory Information System in Côte d'Ivoire: Quasi-Experimental Study on 21 Clinical Laboratories From 2014 to 2020 (e50407)
Yao He, Yves-Rolland Kouabenan, Paul Assoa, Nancy Puttkammer, Bradley Wagenaar, Hong Xiao, Stephen Gloyd, Noah Hoffman, Pascal Komena, N'Zi Kamelan, Casey Iams-Hauser, Adama Pongathe, Alain Kouakou, Jan Flowers, Nadine Abiola, Natacha Kohemun, Jean-Bernard Amani, Christiane Adje-Toure, Lucy Perrone. ................................................................. 2587

Spatial Access to Continuous Maternal and Perinatal Health Care Services in Low-Resource Settings: Cross-Sectional Study (e49367)
Qin Li, Elsa Kanduma, Isaías Ramiro, Dong Xu, Rosa Cuco, Eusebio Chaguissa, Yili Yang, Xiuli Wang, Jay Pan. ................................................................. 2599

Evidence From the China Family Panel Studies Survey on the Effect of Integrating the Basic Medical Insurance System for Urban and Rural Residents on the Health Equity of Residents: Difference-in-Differences Analysis (e50622)
Yingying Meng, Ran Yu, Huixin Bai, Junqiang Han. ................................................................. 2626

**Viewpoints**

Combining Digital and Molecular Approaches Using Health and Alternate Data Sources in a Next-Generation Surveillance System for Anticipating Outbreaks of Pandemic Potential (e47673)
Pablo Ramos, Izabel Marcilio, Ana Bento, Gerson Penha, Juliane de Oliveira, Ricardo Khouri, Roberto Andrade, Roberto Carreiro, Vinicius Oliveira, Luiz Galvão, Luiz Landa, Mauricio Barreto, Kay van der Horst, Manoel Barral-Netto, AESOP Collaborating Teams. ................................................................. 147

Digital Transformation of Public Health for Noncommunicable Diseases: Narrative Viewpoint of Challenges and Opportunities (e49575)
Onicio Leal Neto, Viktor Von Wyl. ................................................................. 156

Defining the Dimensions of Diversity to Promote Inclusion in the Digital Era of Health Care: A Lexicon (e51980)
Yashoda Sharma, Anindita Saha, Jennifer Goldsack. ................................................................. 167

Designing Electronic Data Capture Systems for Sustainability in Low-Resource Settings: Viewpoint With Lessons Learned From Ethiopia and Myanmar (e47703)
Natalie Benda, Kylie Dougherty, Abebe Gebremariam Gobezayehu, John Cranmer, Sakie Zawtha, Katerina Andreadis, Heran Biza, Ruth Masterson Creber. ................................................................. 174
Lara Kufiof, Rana Hajjeh, Mohannad Al Nsour, Randa Saad, Victoria Bélorgeot, Abdinasir Abubakar, Yousef Khader, Salman Rawaf. ................................. 186

Untapped Potential of Unobtrusive Observation for Studying Health Behaviors (e46638)
Jack Benton, David French. .................................................................................. 194

Data Sharing in a Decentralized Public Health System: Lessons From COVID-19 Syndromic Surveillance (e52587)
Ryan Rigby, Alva Ferdinand, Hye-Chung Kum, Cason Schmit. ........................... 202

Responding to the Return of Influenza in the United States by Applying Centers for Disease Control and Prevention Surveillance, Analysis, and Modeling to Inform Understanding of Seasonal Influenza (e54340)
Rebecca Borchering, Matthew Biggerstaff, Lynnette Brammer, Alicia Budd, Shikha Garg, Alicia Fry, A Iuliano, Carrie Reed. ........................................... 211

From the Public Health Perspective: a Scalable Model for Improving Epidemiological Testing Efficacy in Low- and Middle-Income Areas (e55194)
Xuefeng Huang, Qian-Yi Kong, Xiaowen Wan, Yaling Huang, Rongrong Wang, Xiaoxue Wang, Yingying Li, Yuqing Wu, Chongyuan Guan, Junyang Wang, Yuanjuan Zhang. ............................................................... 218

Development of a Consolidated Health Facility Masterlist Using Data From Polio Electronic Surveillance in the World Health Organization African Region (e54250)

Now Is the Time to Strengthen Government-Academic Data Infrastructures to Jump-Start Future Public Health Crisis Response (e51880)
Jian-Sin Lee, Allison Tyler, Tiffany Veinot, Elizabeth Yakel. ................................. 2347

Reviews

HIV Prevention and Treatment Interventions for Black Men Who Have Sex With Men in Canada: Scoping Systematic Review (e40493)
Jemal Demeke, Pascal Djiaedou, Abban Yusuf, Darren Whitfield, David Lightfoot, Fiqir Worku, Gamji Abu-Ba'are, Lawrence Mbuagbaw, Sulaimon Giwa, LaRon Nelson. ................................................................. 238

Changes to Public Health Surveillance Methods Due to the COVID-19 Pandemic: Scoping Review (e49185)
Emily Clark, Sophie Neumann, Stephanie Hopkins, Alyssa Kostopoulos, Leah Hagerman, Maureen Dobbins. ................................................................. 252

Counseling Supporting HIV Self-Testing and Linkage to Care Among Men Who Have Sex With Men: Systematic Review and Meta-Analysis (e45647)
Siyu Chen, Yuan Fang, Paul Chan, Joseph Kawuki, Phoenix Mo, Zixin Wang. ................................................................. 281

Understanding Gaps in the Hypertension and Diabetes Care Cascade: Systematic Scoping Review (e51802)
Jie Wang, Fangqin Tan, Zhanchong Wang, Yiwen Yu, Jingsong Yang, Yueqing Wang, Ruitai Shao, Xuejun Yin. ................................................................. 309

Optimization of Screening Strategies for COVID-19: Scoping Review (e44349)
Yuanhua Liu, Yun Yin, Michael Ward, Ke Li, Yue Chen, Mengwei Duan, Paulina Wong, Jie Hong, Jiaqi Huang, Jin Shi, Xuan Zhou, Xi Chen, Jiayao Xu, Rui Yuan, Lingcai Kong, Zhijie Zhang. ................................................................. 329

Short- and Long-Term Predicted and Witnessed Consequences of Digital Surveillance During the COVID-19 Pandemic: Scoping Review (e47154)
Leigha Comer, Lorie Donelle, Bradley Hibbert, Maxwell Smith, Anita Kothari, Saverio Stranges, Jason Gilliland, Jed Long, Jacqueline Burkell, Jacob Shelley, Jodi Hall, James Shelley, Tommy Cooke, Marionette Ngole Dione, Danica Facca. ................................................................. 353
<table>
<thead>
<tr>
<th>Title</th>
<th>DOI</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug-Resistant Tuberculosis Case-Finding Strategies: Scoping Review</td>
<td>(e46137)</td>
<td>374</td>
</tr>
<tr>
<td>Susanna van Wyk, Marriott Niwasa, Fang-Wen Lu, Chih-Chan Lan, James</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seddon, Graeme Hoddinott, Lario Viljoen, Gunar Günther, Numurai</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruswa, N Shah, Mareli Claassens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of Physical Exercise on Telomere Length in Healthy Adults:</td>
<td>(e46019)</td>
<td>667</td>
</tr>
<tr>
<td>Systematic Review, Meta-Analysis, and Meta-Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juan Sánchez-González, Juan Sánchez-Rodríguez, Sergio Varela-Rodríguez,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rogelio González-Sarmiento, Cristina Rivera-Pizón, Raúl Juárez-Vela,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clara Tejada-Garrido, Javier Martín-Vallejo, Víctor Navarro-López</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Association Between Physical Activity and the Risk of Burnout in Health</td>
<td>(e49772)</td>
<td>735</td>
</tr>
<tr>
<td>Care Workers: Systematic Review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pierpaolo Mincarone, Antonella Bodini, Maria Tumolo, Saverio Sabina,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riccardo Colella, Linda Mannini, Eugenio Sabato, Carlo Leo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community Engagement in Vaccination Promotion: Systematic Review and</td>
<td>(e49695)</td>
<td>784</td>
</tr>
<tr>
<td>Meta-Analysis (e49695)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yao Xie, Xiaoli Liao, Meijuan Lin, Lin Yang, Kin Cheung, Qingpeng</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang, Yan Li, Chun Hao, Harry Wang, Yang Gao, Dexing Zhang,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alex Molassiotis, Gilman Siu, Angela Leung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participatory Approaches in the Context of Research Into Workplace</td>
<td>(e50195)</td>
<td>815</td>
</tr>
<tr>
<td>Health Promotion to Improve Physical Activity Levels and Reduce</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sedentary Behavior Among Office-Based Workers: Scoping Review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aidan Buffey, Christina Langley, Brian Carson, Alan Donnelly, Jon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salsberg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessing the Labeling Information on Drugs Associated With Suicide</td>
<td>(e49755)</td>
<td>1105</td>
</tr>
<tr>
<td>Risk: Systematic Review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soo Jeon, HyunJoo Lim, Hyo-bin Cheon, Juhee Ryu, Jin-Won Kwon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation of Socioeconomic and Environmental Factors With</td>
<td>(e48682)</td>
<td>1856</td>
</tr>
<tr>
<td>Incidence of Crohn Disease in Children and Adolescents: Systematic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review and Meta-Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jens Weidner, Ingrm Glauche, Ulf Manuwald, Ivana Kern, Ines Reinecke,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franziska Bathelt, Makan Amin, Fan Dong, Ulrike Rothe, Joachim</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kugler</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated Network Analysis of Symptom Clusters Across Monkeypox</td>
<td>(e49285)</td>
<td>2521</td>
</tr>
<tr>
<td>Epidemics From 1970 to 2023: Systematic Review and Meta-Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shu Su, Meng Jia, Yingni Yu, Hu Li, Wenwei Yin, Yi Lu, Rongzhong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huang, Rong Xiang, Huizhe Huang, Peng Hu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Impact of Optimal Glycemic Control on Tuberculosis Treatment</td>
<td>(e53948)</td>
<td>2559</td>
</tr>
<tr>
<td>Outcomes in Patients With Diabetes Mellitus: Systematic Review and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meta-Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li Zhao, Feng Gao, Chunlan Zheng, Xuezhi Sun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19 Vaccine Hesitancy: Umbrella Review of Systematic Reviews and</td>
<td>(e54769)</td>
<td>2569</td>
</tr>
<tr>
<td>Meta-Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tahani Rahbeni, Prakasini Satapathy, Ramaiah Itumalia, Roy Marzo,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khalid Mugheed, Mahalaqua Khatib, Shilpa Gaidhane, Quazi Zahiruddin,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ali Rabaan, Hayam Airasheed, Maha Al-Subaie, Nawal Al Kaabil,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mohammed Alissa, Amani Ibrahim, Hussain Alsaif, Israa Naser, Sarvesh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rustagi, Neelima Kukreti, Arkadiusz Dziedzic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapid Surveillance Report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend of Mortality Due to Congenital Anomalies in Children Younger</td>
<td>(e53860)</td>
<td>426</td>
</tr>
<tr>
<td>Than 5 Years in Eastern China, 2012-2021: Surveillance Data Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wen-Hong Dong, Jun-Xia Guo, Lei Wang, Shuang-Shuang Zheng, Bing-Quan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhu, Jie Shao</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Protocols

Xinchun Gu, Conall Watson, Utkarsh Agrawal, Heather Whitaker, William Elson, Sneha Anand, Ray Borrow, Anna Buckingham, Elizabeth Button, Lottie Curtis, Dominic Dunn, Alex Elliot, Filipa Ferreira, Rosalind Goudie, Uy Hoang, Katja Hoschler, Gavin Jaminie, Debasish Kar, Beatrix Kele, Meredith Leston, Ezra Linley, Jack Macartney, Gemma Marsden, Cecilia Okusi, Omid Parvizi, Catherine Quinot, Praveen Sebastianpillai, Vanashree Sexton, Gillian Smith, Timea Suli, Nicholas Thomas, Catherine Thompson, Daniel Toddill, Rashmi Wimalaratna, Matthew Inada-Kim, Nick Andrews, Victoria Tzortziou-Brown, Rachel Byford, Maria Zambon, Jamie Lopez-Bernal, Simon de Lusignan. 1071

Assessing Outcomes in HIV Prevention and Treatment Programs With Female Sex Workers and Men Who Have Sex With Men: Expanded Polling Booth Survey Protocol (e54313)
Parinita Bhattacharjee, Leigh McClarty, Joshua Kimani, Shajy Isac, Rhoda Wanjiru Kabuti, Antony Kinyua, Jaffred Karakaja Okoyana, Virjinia Njeri Ndukuuyu, Helgar Musyoki, Anthony Kiplagat, Peter Arimi, Souradet Shaw, Faran Emmanuel, Monica Gandhi, Marissa Becker, James Blanchard. 1091

Corrigenda and Addendas

Correction: Dashboard With Bump Charts to Visualize the Changes in the Rankings of Leading Causes of Death According to Two Lists: National Population-Based Time-Series Cross-Sectional Study (e56014)

Correction: Comparison of Four Active SARS-CoV-2 Surveillance Strategies in Representative Population Sample Points: Two-Factor Factorial Randomized Controlled Trial (e57203)

Corrigendum to: COVID-19 Vaccine Hesitancy: Umbrella Review of Systematic Reviews and Meta-Analysis (e64080)
Tiffany Leung, Travis Sanchez, Gunther Eysenbach. 1985

Letters to the Editor

The Need for a Bleed Type–Specific Annual Bleeding Rate in Hemophilia Studies (e51372)
Kun Huang. 1987

Authors' Reply: “The Need for a Bleed Type–Specific Annual Bleeding Rate in Hemophilia Studies” (e54756)
Limin Wang, Shimeng Liu, Shan Jiang, Chaofan Li, Liyong Lu, Yunhai Fang, Shunping Li. 1989

Tutorials

Generating Contextual Variables From Web-Based Data for Health Research: Tutorial on Web Scraping, Text Mining, and Spatial Overlay Analysis (e50379)
Pablo Galvez-Hernandez, Angelina Gonzalez-Viana, Luis Gonzalez-de Paz, Ketan Shankardass, Carles Muntaner. 1992

Prospective Spatiotemporal Cluster Detection Using SaTScan: Tutorial for Designing and Fine-Tuning a System to Detect Reportable Communicable Disease Outbreaks (e50653)
Alison Levin-Rector, Martin Kulldorff, Eric Peterson, Scott Hostovich, Sharon Greene. 2011
Research Letters

U-Shaped Relationship Between Fibrinogen Level and 10-year Mortality in Patients With Acute Coronary Syndrome: Prospective Cohort Study (e54485)
Yi Li, Yuheng Jia, Lin Bai, Bosen Yang, Mao Chen, Yong Peng. 2538

Geographic Disparities in Access to Assisted Reproductive Technology Centers in China: Spatial-Statistical Study (e55418)
Qingqing Zhou, Huatang Zeng, Liqun Wu, Kaichuan Diao, Rongxin He, Bin Zhu. 2542
Original Paper

Estimating the Epidemic Size of Superspreading Coronavirus Outbreaks in Real Time: Quantitative Study

Kitty Y Lau1,2, BEng; Jian Kang2, MS; Minah Park3, PhD; Gabriel Leung1,2, MD; Joseph T Wu1,2,4, PhD; Kathy Leung1,2,4, PhD

1Laboratory of Data Discovery for Health Limited (D24H), Hong Kong Science Park, China (Hong Kong)
2WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, China (Hong Kong)
3Department of Health Convergence, Ewha Womans University, Seoul, Republic of Korea
4The University of Hong Kong – Shenzhen Hospital, Shenzhen, China

Corresponding Author:
Kathy Leung, PhD
WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health
Li Ka Shing Faculty of Medicine
1/F North Wing, Patrick Manson Building
7 Sassoon Road, Pok Fu Lam
The University of Hong Kong
China (Hong Kong)
Phone: 852 3917 6735
Email: ksmleung@hku.hk

Abstract

Background: Novel coronaviruses have emerged and caused major epidemics and pandemics in the past 2 decades, including SARS-CoV-1, MERS-CoV, and SARS-CoV-2, which led to the current COVID-19 pandemic. These coronaviruses are marked by their potential to produce disproportionally large transmission clusters from superspreading events (SSEs). As prompt action is crucial to contain and mitigate SSEs, real-time epidemic size estimation could characterize the transmission heterogeneity and inform timely implementation of control measures.

Objective: This study aimed to estimate the epidemic size of SSEs to inform effective surveillance and rapid mitigation responses.

Methods: We developed a statistical framework based on back-calculation to estimate the epidemic size of ongoing coronavirus SSEs. We first validated the framework in simulated scenarios with the epidemiological characteristics of SARS, MERS, and COVID-19 SSEs. As case studies, we retrospectively applied the framework to the Amoy Gardens SARS outbreak in Hong Kong in 2003, a series of nosocomial MERS outbreaks in South Korea in 2015, and 2 COVID-19 outbreaks originating from restaurants in Hong Kong in 2020.

Results: The accuracy and precision of the estimation of epidemic size of SSEs improved with longer observation time; larger SSE size; and more accurate prior information about the epidemiological characteristics, such as the distribution of the incubation period and the distribution of the onset-to-confirmation delay. By retrospectively applying the framework, we found that the 95% credible interval of the estimates contained the true epidemic size after 37% of cases were reported in the Amoy Garden SARS SSE in Hong Kong, 41% to 62% of cases were observed in the 3 nosocomial MERS SSEs in South Korea, and 76% to 86% of cases were confirmed in the 2 COVID-19 SSEs in Hong Kong.

Conclusions: Our framework can be readily integrated into coronavirus surveillance systems to enhance situation awareness of ongoing SSEs.

(JMIR Public Health Surveill 2024;10:e46687) doi:10.2196/46687

KEYWORDS

coronavirus; superspreading event; SSE; epidemic size; severe acute respiratory syndrome; SARS; Middle East respiratory syndrome; MERS; coronavirus disease 2019; COVID-19
Introduction

In the past 2 decades, 3 coronaviruses have emerged and caused widespread public anxiety: SARS-CoV-1 in 2002; MERS-CoV in 2012; and most recently, SARS-CoV-2 in 2019. As of November 2023, the new COVID-19 caused by SARS-CoV-2 has led to more than 771 million confirmed cases and almost 7 million deaths [1]. Despite the differences in epidemiologic characteristics, all 3 coronaviruses are marked by transmission heterogeneity and the ability to produce disproportionally large clusters via superspreading events (SSEs) [2-10]. For instance, it is estimated that 19% of the cases caused 80% of the local transmission in the first wave of COVID-19 epidemic in Hong Kong by April 28, 2020 [11]. The overdispersed transmission pattern was also previously reported for both SARS and MERS [5,8,12].

SSEs are shaped by a combination of determinants originating from the virus, host, and environment [13]. The emergence of mutations and variants with increased transmissibility and greater immune escape mechanisms contribute substantially to the occurrence of SSEs [14]. Superspreaders, who are known to be unusually infectious, were reported to shed a higher viral load over an extended duration [15], which was exacerbated in the context of COVID-19 because significant proportion of these individuals who are infectious could transmit the virus before symptom onset (presymptomatic) or without showing any symptoms (asymptomatic) [16-20]. Certain environments facilitate the transmission, as evidenced by most SSEs of SARS and MERS being nosocomial outbreaks [2,21].

Although the exact mechanism of SSEs is not well established, monitoring SSEs in real time is critical to contain and mitigate epidemics of coronaviruses [22]. It is estimated that implementing control measures a week earlier would lead to a 2.6 times decrease in average epidemic size and a reduction of 4 weeks in average epidemic duration [21,23]. Real-time estimation of the epidemic size of SSEs allows us to proactively take timely measures, including resource planning and mobilization, testing, contact tracing, and implementing targeted interventions in high-risk settings. However, given the inherent heterogeneity of an epidemic, population-level measurements of the average transmissibility of a typical individual who is infectious, such as basic reproductive number ($R_0$) or effective reproductive number ($R_e$), are not appropriate for estimating the transmission potential of SSEs [5]. Here, we developed a framework based on back-calculation for SSE epidemic size estimation, allowing for a more comprehensive understanding of the transmission dynamics.

Methods

Ethical Considerations

This study involves secondary analysis of existing aggregate research data, including SARS-CoV-1 data from Leung et al [24], MERS data from Cowling et al [7], and COVID-19 data that are publicly available [25]. All data were deidentified and only aggregate data were used. The University of Hong Kong/Hospital Authority Hong Kong West Institutional Review Board approved the secondary analysis without requiring further consent. As participants were not directly involved in the current research activities, no compensation was provided to the individuals.

Model Assumptions

Back-calculation was first designed for short-term prediction of diseases with a long incubation period [26] and has also been applied to point-source outbreaks with limited onward transmission, such as Legionnaires disease [27]. We extended the method and focused on estimating the final size of symptomatic or laboratory-confirmed cases at the early stage of a coronavirus SSE, and asymptomatic cases were not considered in the estimation if they were not confirmed and reported.

We assumed that (1) all cases in a given SSE were infected at around the same time (eg, infected by the same superspreader or environmental exposure within a short period of time); (2) all cases followed the same probability distribution function (PDF) of the incubation period, the same PDF of the onset-to-confirmation delay, and the same PDF of the generation time; and (3) there was limited onward secondary transmission when the SSEs were controlled by sufficient contact tracing and timely responses. For SARS and MERS, we assumed that the $R_t$ of secondary transmission was 0, given their relatively lower potential of person-to-person transmission [2]. For COVID-19, which has much higher transmissibility and significant presymptomatic transmission, we assumed that the $R_t$ of secondary transmission could be greater than 1, but the transmission was restricted within 1 disease generation. We assumed that testing and contact tracing policies were the same within 14 days, given that the generation time for SARS-CoV-1, MERS, and ancestral SAR-CoV-2 are around 7 days (Table 1).
Table 1. Parameters used in the simulations.

<table>
<thead>
<tr>
<th>Simulated scenarios</th>
<th>Incubation period (lognormal distribution): mean range (SD range)</th>
<th>Onset-to-confirmation delay (lognormal distribution): mean range (SD range)</th>
<th>Secondary transmission(^a); (R_t)(^b)</th>
<th>Generation time (lognormal distribution): mean range (SD range)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS and MERS</td>
<td>3-15 (1-10)</td>
<td>_c</td>
<td>0</td>
<td>—</td>
<td>[28-32]</td>
</tr>
<tr>
<td>SARS and MERS</td>
<td>3-15 (1-10)</td>
<td>0-7 (1-7)</td>
<td>0</td>
<td>—</td>
<td>[33,34]</td>
</tr>
<tr>
<td>COVID-19</td>
<td>5-8 (2-5)</td>
<td>—</td>
<td>1.5</td>
<td>5-7 (5-7)</td>
<td>[19,35-38]</td>
</tr>
<tr>
<td>COVID-19</td>
<td>5-8 (2-5)</td>
<td>3-7 (2-5)</td>
<td>1.5</td>
<td>5-7 (5-7)</td>
<td>[19]</td>
</tr>
<tr>
<td>COVID-19 adjusted for presymptomatic tran-</td>
<td>5-8 (2-5)</td>
<td>—</td>
<td>1.5</td>
<td>5-7 (5-7)</td>
<td>[19,35-38]</td>
</tr>
<tr>
<td>mission(^d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19 adjusted for presymptomatic tran-</td>
<td>5-8 (2-5)</td>
<td>3-7 (2-5)</td>
<td>1.5</td>
<td>5-7 (5-7)</td>
<td>[19]</td>
</tr>
</tbody>
</table>

\(^a\)Secondary transmission can be measured by \(R_t\) within 1 disease generation. 

\(^b\)\(R_t\): effective reproductive number. 

\(^c\)Not applicable. 

\(^d\)The 60% presymptomatic transmission was included in the prior information.

Figure 1A-B summarizes the characteristics of coronavirus SSEs in disease surveillance. Here, the incubation period is the time interval between infection and symptom onset; in practice, there is a delay from symptom onset to case confirmation, and the time of symptom onset can be unavailable for asymptomatic infections or due to underreporting.
Figure 1. Characteristics of coronavirus SSEs. (A) Time of exposure, symptom onset, and case confirmation. The incubation period is the time interval between infection and symptom onset. The onset-to-confirmation delay is the time interval between symptom onset and confirmation. (B) Transmission between an infector-infectee pair: (a) transmission occurs after symptom onset of the infector (symptomatic transmission); (b) transmission occurs before symptom onset of the infector (presymptomatic transmission). The generation time is the interval between the exposure time of the infector and the infectee. (C) Approximate distribution of presymptomatic, asymptomatic, and symptomatic transmission for SARS, MERS, and COVID-19. (D) Simulated COVID-19 SSE. The undetected presymptomatic transmission pairs within 1 disease generation and R_t = 1.5 in a simulated COVID-19 SSE with 10 cases. R_t: effective reproductive number; SSE: superspreading event.

Estimating the Size of SSEs

We formulated the model to estimate the epidemic size of SSEs as follows. Let $X$ denote the incubation period with PDF $g(x)$ and $Y$ denote the onset-to-confirmation with PDF $h(y)$. For an SSE $j$, let $n_j$ be the total number of people infected at time 0 and $k_j$ be the number of symptomatic cases that have been confirmed up to (and including) time $t_j$ since the exposure time for this SSE.

Let $x_{ij}$ and $y_{ij}$ be the values of $X$ and $Y$ for the $i$th case in the $j$th SSE. Let $f(X,Y|\theta)$ be the joint probability distribution of $X$ and $Y$ with $\theta$ representing the parameters. We assume that $X$ and $Y$ are independent such that $f(X,Y|\theta) = g(x)h(y)$, and $F(X,Y|\theta)$ is the cumulative probability distribution of $f(X,Y|\theta)$. The number of symptomatic cases $k_j$ has a binomial distribution with SSE size $n_j$ and success probability $F(X+Y>r; \theta)$. Therefore, the likelihood function for $j$th SSE is:
If the time of symptom onset is known, let $M_{j,i}$ be the set of confirmed cases with known time from exposure to symptom onset and known time from symptom onset to confirmation with a size of $m_{j,i}$. If the time of symptom onset is unknown, let $M_{j,2}$ be the set of confirmed cases with unknown time from exposure to symptom onset but known time from exposure to confirmation with a size of $m_{j,2}$. Let $Z$ denote $X+Y$ with the probability distribution $q(z|\theta)$. The likelihood function for the $j$th SSE becomes:

The overall likelihood function for $w$ SSEs is:

We estimated the SSE size $n_j$ and other parameters $\theta$ using Markov chain Monte Carlo with Gibbs sampling and noninformative flat priors. We considered an estimate to be accurate if the 95% credible interval (CRI) covered the true value. An estimate was precise if the CRI had a relative error within 1.

### Model Validation

We simulated SSE scenarios based on the epidemiological characteristics of SARS, MERS, and COVID-19 (Figure 1C). For SARS and MERS to be considered under control, we assumed that there was a single exposure and no secondary transmission. For COVID-19, we assumed that 60% of the transmission occurred during the presymptomatic phase [19,20,35,39,40]. For example, if we simulated a single-exposure COVID-19 SSE with 10 cases, 6 (60%) second-generation cases were infected by the other 4 first-generation cases in the SSE during the presymptomatic phase. However, the 6 second-generation cases would be taken as first-generation infections directly linked to the primary exposure when the second-generation cases would be taken as first-generation during the presymptomatic phase. Thus, we assumed that all cases in the Amoy Gardens SSE were exposed to SARS-CoV-1 on March 15, and was discharged on March 19. He visited and stayed 1 night at the Amoy Gardens apartment complex, where he had diarrhea. Thus, we assumed that all cases in the Amoy Gardens SSE were exposed to SARS-CoV-1 on March 19. We used noninformative prior distribution for all parameters in the inference since no information about the incubation period and onset-to-confirmation delay could be estimated by bootstrapping from 50 or 100 cases before the occurrence of the simulated SSEs.

The SSE size estimation was performed starting from the time when 5% of all cases were observed until the time when 95% of all cases were observed. To compare scenarios with different epidemic sizes, we used relative size (ie, percentage of the cases observed in an SSE) and SSE duration (ie, percentage of the duration between the time of confirmation of the first and last case) in the assessment of the accuracy of the estimation.

### Case Studies

We applied our framework to the Amoy Gardens SSE for SARS in Hong Kong, nosocomial SSEs for MERS in South Korea, and 2 SSEs from the COVID-19 pandemic in Hong Kong. An alert would be issued when 10 or more cases were confirmed to link with the same index case, by which we assumed that we had minimal data to start the SSE size estimation.

#### Hong Kong Amoy Gardens SARS SSE

The Amoy Gardens SSE was the largest cluster in the 2003 Hong Kong SARS outbreak [4,24]. The index case developed symptoms on March 13-14, was admitted to the Prince Wales Hospital on March 15, and was discharged on March 19. He visited and stayed 1 night at the Amoy Gardens apartment complex, where he had diarrhea. Thus, we assumed that all cases in the Amoy Gardens SSE were exposed to SARS-CoV-1 on March 19. We used noninformative prior distribution for all parameters in the inference since no information about the incubation period and the onset-to-confirmation delay could be estimated before the Amoy Gardens SSE (Table 2). A total of 311 laboratory-confirmed cases were retrospectively confirmed and reported in the Amoy Gardens SSE.

### Table 2. Prior distributions for parameters.

<table>
<thead>
<tr>
<th>Case studies</th>
<th>Prior distribution for the incubation period (lognormal distribution): mean (SD) or mean range (SD range)</th>
<th>Prior distribution for the onset-to-confirmation delay (lognormal distribution): mean (SD) or mean range (SD range)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS Amoy Gardens SSE\textsuperscript{a} in Hong Kong</td>
<td>3-15 (1-10)</td>
<td>0-7 (1-7)</td>
<td>[28-34]</td>
</tr>
<tr>
<td>MERS nosocomial SSEs in South Korea</td>
<td>6.3 (4.3)</td>
<td>5.6 (5.4)</td>
<td>[41-44]</td>
</tr>
<tr>
<td>COVID-19 restaurant SSEs in Hong Kong</td>
<td>5.2 (3.9)</td>
<td>5.3 (4.0)</td>
<td>[45]</td>
</tr>
</tbody>
</table>

\textsuperscript{a}SSE: superspreading event.
South Korea Nosocomial MERS SSEs

The MERS SSEs in South Korea in 2015 were the largest MERS outbreak outside of the Middle East, with 82.3% of all confirmed infections caused by only 5 (2.7%) cases [41,46]. The index case visited multiple clinics after traveling to the Middle East. He first developed fever and myalgia illness on May 11 and visited the same clinic on May 12, 14, and 15. The symptoms were not resolved and a cough developed on May 15. The index case was admitted to a secondary hospital (Cluster 1) on May 15 and transferred to a tertiary hospital in Seoul (Cluster 2) on May 17, where he was later diagnosed with MERS-CoV on May 20. Meanwhile, another large SSE took place in other hospitals and was traced back to this same index case (Cluster 3). We took May 15, the day that the index case visited Cluster 1, as the time of exposure of all cases in Cluster 1. In the series of nosocomial SSEs, the time of exposure would be different for each confirmed case depending on when they visited the contaminated hospitals. Therefore, 3 major nosocomial clusters (Cluster 1-3) with sizes of 29, 125, and 27 were studied retrospectively [7,41]. We used MERS data from the previous Middle East outbreaks as the prior information for parameter estimation (Table 2).

Hong Kong COVID-19 SSEs

For the COVID-19 case study, 2 linked Hong Kong restaurant SSEs in 2020 with relatively exclusive contact tracing were selected [25]. The first SSE took place at a restaurant of Tao Heung Holdings on July 9 where people gathered in a celebration event, resulting in an SSE of 42 cases. Soon, another SSE took place at another restaurant of Fulum Holdings on July 11, where hundreds of people attended a birthday party, resulting in an SSE of 44 cases. We assumed the time of exposure to be July 9 and July 11 for the 2 SSEs, respectively. Prior information on the incubation period and onset-to-confirmation delay from previous COVID-19 confirmed cases were used in the parameter estimation (Table 2).

Results

Model Validation

Inferred from the model simulation, we found that the accuracy and precision of our estimation increased with the size of the SSE (Figures 2 and 3). For SARS and MERS, with and without onset-to-confirmation delay, all Crls covered the true SSE size and the relative error was less than 1 for all simulated scenarios, when the estimation was performed at the time exceeding 50% of the duration of the SSE or when more than 60% of the cases were observed (Figure 2A-B). With the presence of onset-to-confirmation delay, the range of relative error doubled when the estimation was performed before 50% of the duration of the SSE or when less than 60% of the cases were observed (Figure 2C-D). Our method tended to overestimate the size in the early stage of an SSE when less than 20% of cases were reported, especially in the scenarios with onset-to-confirmation delay.
Figure 2. The accuracy and precision of SSE size estimation for simulated SARS and MERS scenarios. Scenario 1: (A) simulated SARS and MERS SSEs with no onset-to-confirmation delay, with size estimation performed at the time as the percentage of the duration of the SSE; and (B) simulated SARS and MERS SSEs with no onset-to-confirmation delay, with size estimation performed at the time by the percentage of cases observed and reported. Scenario 2: (C) simulated SARS and MERS SSEs with onset-to-confirmation delay, with size estimation performed at the time as the percentage of the duration of the SSE; and (D) simulated SARS and MERS SSEs with onset-to-confirmation delay, with size estimation performed at the time by the percentage of cases observed and reported. SSE: superspreading event.

If the presymptomatic transmission was not adjusted in the COVID-19 simulated SSEs, there would always be an underestimation even after 50% of the duration of the SSE or when more than 60% of the cases were observed (Figure 3A-D). For a typical COVID-19 SSE with 50 cases, the model would predict the final size to be around 30 when half of the cases were observed if second-generation infections cannot be distinguished from first-generation infections. If we could detect the transmission pattern of COVID-19 in time and adjust for presymptomatic transmission as per our prior information, the estimation would perform similarly well to the simulated scenarios of SARS and MERS when presymptomatic transmission was not included (Figure 3E-H).
Figure 3. The accuracy and precision of SSE size estimation for simulated COVID-19 scenarios. Scenario 3: (A) simulated COVID-19 SSEs with no onset-to-confirmation delay but 60% presymptomatic transmission, with size estimation performed at the time as the percentage of the duration of the SSE; and (B) simulated COVID-19 SSEs with no onset-to-confirmation delay but 60% presymptomatic transmission, with size estimation performed at the time by the percentage of cases observed and reported. Scenario 4: (C) simulated COVID-19 SSEs with both onset-to-confirmation delay and 60% presymptomatic transmission, with size estimation performed at the time as the percentage of the duration of the SSE; and (D) simulated COVID-19 SSEs with both onset-to-confirmation delay and 60% presymptomatic transmission, with size estimation performed at the time by the percentage of cases observed and reported. Scenario 5: (E) simulated COVID-19 SSEs with no onset-to-confirmation delay but 60% presymptomatic transmission adjusted in prior information, with size estimation performed at the time as the percentage of the duration of the SSE; and (F) simulated COVID-19 SSEs with no onset-to-confirmation delay but 60% presymptomatic transmission adjusted in prior information, with size estimation performed at the time by the percentage of cases observed and reported. Scenario 6: (G) simulated COVID-19 SSEs with onset-to-confirmation delay and 60% presymptomatic transmission adjusted in prior information, with size estimation performed at the time as the percentage of the duration of the SSE; and (H) simulated COVID-19 SSEs with onset-to-confirmation delay and 60% presymptomatic transmission adjusted in prior information, with size estimation performed at the time by the percentage of cases observed and reported. SSE: superspreading event.
Case Studies

In the case study of Amoy Gardens SARS SSE, the Crl covered the final SSE size and had a relative error within 1 in 4 days after the alert was issued, which was 25 days before the last case of the SSE was observed (Figure 4). At the time when the alert was issued, only 36.7% (114/311) of all cases were observed. The SSE size estimation became more accurate and precise thereafter.

Figure 4. Real-time size estimation of Amoy Gardens SARS SSE in Hong Kong in 2003. (A) Epidemic curve by time of symptom onset. (B) Epidemic curve by time of confirmation. (C) Real-time estimation of SSE size by the time since the exposure (May 19, 2003). The dash line indicates the actual epidemic size of the SSE. Crl: credible interval; SSE: superspreading event.

For the series of South Korea MERS SSEs, the Crls of Cluster 1, Cluster 2, and Cluster 3 covered the final SSE sizes with a relative error within 1 in 2 days, 3 days, and 0 days after the alert was issued, respectively, which were 4 days, 15 days, and 9 days before the last case of the respective SSE was observed (Figure 5). At the time when alerts were issued, 62% (18/29), 52% (65/125), and 41% (11/27) of all cases in the respective SSEs were observed, and the estimation converged sooner for later SSEs.
Figure 5. Real-time size estimation of nosocomial MERS SSEs in South Korea in 2015. (A) Epidemic curve by time of confirmation for Cluster 1, Cluster 2, and Cluster 3 denoted by blue, red, and yellow, respectively. (B) Real-time estimation of the SSE size of Cluster 1. (C) Real-time estimation of the SSE size of Cluster 2. (D) Real-time estimation of the SSE size of Cluster 3 by the time since the index case visited Cluster 1 (May 15, 2015). The dash line indicates the actual epidemic size of the SSEs. Crl: credible interval; SSE: superspreading event.

For the Tao Heung COVID-19 SSE in Hong Kong, the Crl covered the final SSE size and had a relative error within 1 in 5 days after the alert was issued, which was 20 days before the last case of the SSE was observed. At this time, 76% (32/42) of all cases were observed (Figure 6). Similarly, for the Fulum COVID-19 SSE, the Crl covered the final SSE size and had a relative error within 1 in 7 days after the alert was issued and 5 days before the last case was observed. At the time of issuing the alert, 86% (38/44) of all cases had been observed (Figure 6).
**Figure 6.** Real-time size estimation of COVID-19 SSEs in 2 restaurants in Hong Kong in 2020. (A) Epidemic curve by time of symptom onset of the Tao Heung SSE. (B) Epidemic curve by time of confirmation of the Tao Heung SSE. (C) Real-time estimation of the size of the Tao Heung SSE by the time since the social gathering on July 9, 2020. (D) Epidemic curve by time of symptom onset of the Fulum SSE. (E) Epidemic curve by time of confirmation of the Fulum SSE. (F) Real-time estimation of the size of the Fulum SSE by the time since the social gathering on July 11, 2020. The dashed line indicates the actual epidemic size of the SSEs. CrI: credible interval; SSE: superspreading event.

**Discussion**

In this study, we developed a framework based on back-calculation to estimate the epidemic size of coronavirus SSEs in real time. Our method bypassed the complexity of developing a transmission model and gave accurate estimates when there were limited secondary transmission within the SSEs. As expected, we found that the precision and accuracy of the size estimation increased with more observed cases, where large SSEs could be spotted earlier, such that immediate responses could be taken. The estimation performed better with the absence of onset-to-confirmation delay, and prior information on parameters improved the performance of the model. In the case study of MERS nosocomial SSEs, we used data from previous MERS cases from the Middle East to inform the estimation in South Korea. Accurate estimation was obtained sooner for Cluster 2 and Cluster 3 with more information about the distribution of the incubation period and the distribution of the onset-to-confirmation delay obtained after the occurrence of Cluster 1.

We showed that the epidemic size of coronavirus SSEs could be accurately estimated before 50% of the cases were reported when there was no undetected secondary transmission, such as in the SARS and MERS simulated scenarios, or when the secondary transmission was adjusted, such as in the COVID-19 simulated scenarios. In the retrospective study on the Amoy Gardens SARS SSE, we got a robust estimation as early as having only 37% of all the cases identified, and in practice, hospitals could prepare resources beforehand for such a large SSE. The estimation of SARS and MERS SSE sizes were more accurate compared with that of COVID-19 because most symptomatic infections would be confirmed and reported. For example, in the case study of Amoy Gardens, it was suspected that SAR-CoV-1 was excreted in the stool and transmitted through sewerage, and the SSE was identified early when many Amoy Gardens cases were traced back to the block where the index case had stayed [4,24].
However, our method tended to overestimate the size early in simulated SARS and MERS SSEs when the observed case number was low. For simulated COVID-19 SSEs, our method always underestimated the size when secondary transmission within the SSE was not detected or observed until later in the SSE. The real-time daily estimation could also be unstable, especially when little was known about the distribution of the incubation period and reporting delays. For example, in the Amoy Gardens SARS SSE, the time of exposure of some cases could be later than the common exposure time that we assumed (May 19, 2003), depending on the time they came into contact with the contaminated areas. The SARS-CoV-1 virus could also be transmitted via person-to-person contact or contamination of communal facilities such as elevators and doors other than sewage [47]. In the MERS case study, we bracketed the window of the exposure time and grouped the cases into 3 hospital SSEs based on the time of visiting specific hospitals.

The recall bias can be another source of errors, especially for COVID-19 infections where symptoms were relatively mild for many cases. Although we explored the source of uncertainty in our estimation, the required accuracy and precision depends on the potential impact of the SSEs, the available resources for control measures, and the requirement for decision-making. For example, to plan resource allocation and better prepare the health care system, the 95% upper bound of the size estimate should be used. For surveillance purposes, the 95% lower bound of the estimate can be used as a threshold to issue alerts and initiate actions. It could also be used as a conservative estimate of the minimal number of infections.

In summary, our framework can be applied to coronavirus SSEs when there is limited undetected secondary transmission or when secondary infection is accounted for. The emergence of more transmissible variants further complicates the situation [48], and intensive contact tracing and testing might be required to alert SSEs in time. Currently, with the COVID-19 pandemic transiting to the endemic phase and society returning to normalcy, our method can be integrated into coronavirus surveillance systems to monitor potential SSEs in large social gatherings.

Acknowledgments
This research was supported by General Research Fund (grant 17110020), Health and Medical Research Fund (grants 21200122, COVID19F05, COVID19126, and CID-HKU2), Health and Medical Research Fund Research Fellowship Scheme (grant 0620097), and the AIR@InnoHK administered by Innovation and Technology Commission of The Government of the Hong Kong Special Administrative Region. KL was supported by the Enhanced New Staff Start-up Research Grant from Li Ka Shing Faculty of Medicine, The University of Hong Kong. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Data Availability
All the epidemiological information we used is available in the main text or Multimedia Appendix 1. The data, codes, and parameters used to plot Figures 2, 3, 5, and 6 are shared on a GitHub repository [49]. The data (the time of exposure, the time of symptom onsets, the time of case confirmation, the hospital cluster, the calculated incubation period, and the calculated onset-to-confirmation delay) of MERS and COVID-19 case studies were from the Korea Centers for Disease Control and Prevention (KCDC), World Health Organization (WHO), and Centre for Health Protection (CHP) in Hong Kong, which are publicly available. Requests for data of the SARS case study should be sent to the corresponding author.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Details about the estimation of COVID-19 superspreading events and the adjustment for presymptomatic transmission.

References


47. Lee SH. The SARS epidemic in Hong Kong. J Epidemiol Community Health 2003 Sep 01;57(9):652-654 [FREE Full text] [doi: 10.1136/jech.57.9.652] [Medline: 12933765]


Abbreviations

CrI: credible interval
PDF: probability distribution function
R₀: basic reproductive number
Rₐ: effective reproductive number
SSE: superspreading event

Edited by A Mavragani; submitted 21.02.23; peer-reviewed by N Yamada, D Thomas, A Akhmetzhanov; comments to author 11.10.23; revised version received 01.12.23; accepted 10.01.24; published 12.02.24.

Please cite as:
Estimating the Epidemic Size of Superspreading Coronavirus Outbreaks in Real Time: Quantitative Study
JMIR Public Health Surveill 2024;10:e46687
URL: https://publichealth.jmir.org/2024/1/e46687
doi: 10.2196/46687
PMID: 38345850

© Kitty Y Lau, Jian Kang, Minah Park, Gabriel Leung, Joseph T Wu, Kathy Leung. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 12.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Association of Rest-Activity Rhythm and Risk of Developing Dementia or Mild Cognitive Impairment in the Middle-Aged and Older Population: Prospective Cohort Study

Shahab Haghayegh\textsuperscript{1,2,3,4}, MSc, PhD; Chenlu Gao\textsuperscript{1,2,3,4}, PhD; Elizabeth Sugg\textsuperscript{1}, BA; Xi Zheng\textsuperscript{4}, MSc; Hui-Wen Yang\textsuperscript{4}, PhD; Richa Saxena\textsuperscript{1,2,3}, PhD; Martin K Rutter\textsuperscript{5,6}, MD; Michael Weedon\textsuperscript{7}, PhD; Agustin Ibanez\textsuperscript{8}, PhD; David A Bennett\textsuperscript{9}, MD; Peng Li\textsuperscript{1,2,3,4}, PhD; Lei Gao\textsuperscript{1,2*}, MD; Kun Hu\textsuperscript{1,2,3,4*}, PhD

\textsuperscript{1}Massachusetts General Hospital, Boston, MA, United States
\textsuperscript{2}Harvard Medical School, Boston, MA, United States
\textsuperscript{3}Broad Institute, Cambridge, MA, United States
\textsuperscript{4}Brigham and Women's Hospital, Boston, MA, United States
\textsuperscript{5}Faculty of Medicine, Biology and Health, University of Manchester, Manchester, United Kingdom
\textsuperscript{6}Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Centre, Manchester, United Kingdom
\textsuperscript{7}University of Exeter, Exeter, United Kingdom
\textsuperscript{8}Adolfo Ibañez University, Santiago, Chile
\textsuperscript{9}RUSH University, Chicago, IL, United States
* these authors contributed equally

Corresponding Author:
Shahab Haghayegh, MSc, PhD
Massachusetts General Hospital
149 13th Street
4th Floor
Boston, MA, 02129
United States
Phone: 1 5129543436
Email: shaghaye\textsuperscript{gh@mgh.harvard.edu}

Abstract

Background: The relationship between 24-hour rest-activity rhythms (RARs) and risk for dementia or mild cognitive impairment (MCI) remains an area of growing interest. Previous studies were often limited by small sample sizes, short follow-ups, and older participants. More studies are required to fully explore the link between disrupted RARs and dementia or MCI in middle-aged and older adults.

Objective: We leveraged the UK Biobank data to examine how RAR disturbances correlate with the risk of developing dementia and MCI in middle-aged and older adults.

Methods: We analyzed the data of 91,517 UK Biobank participants aged between 43 and 79 years. Wrist actigraphy recordings were used to derive nonparametric RAR metrics, including the activity level of the most active 10-hour period (M10) and its midpoint, the activity level of the least active 5-hour period (L5) and its midpoint, relative amplitude (RA) of the 24-hour cycle [RA=(M10-L5)/(M10+L5)], interdaily stability, and intradaily variability, as well as the amplitude and acrophase of 24-hour rhythms (cosinor analysis). We used Cox proportional hazards models to examine the associations between baseline RAR and subsequent incidence of dementia or MCI, adjusting for demographic characteristics, comorbidities, lifestyle factors, shiftwork status, and genetic risk for Alzheimer's disease.

Results: During the follow-up of up to 7.5 years, 555 participants developed MCI or dementia. The dementia or MCI risk increased for those with lower M10 activity (hazard ratio [HR] 1.28, 95\% CI 1.14-1.44, per 1-SD decrease), higher L5 activity (HR 1.15, 95\% CI 1.10-1.21, per 1-SD increase), lower RA (HR 1.23, 95\% CI 1.16-1.29, per 1-SD decrease), lower amplitude (HR 1.32, 95\% CI 1.17-1.49, per 1-SD decrease), and higher intradaily variability (HR 1.14, 95\% CI 1.05-1.24, per 1-SD increase) as well as advanced L5 midpoint (HR 0.92, 95\% CI 0.85-0.99, per 1-SD advance). These associations were similar in people
aged <70 and >70 years, and in non–shift workers, and they were independent of genetic and cardiovascular risk factors. No significant associations were observed for M10 midpoint, interdaily stability, or acrophase.

**Conclusions:** Based on findings from a large sample of middle-to-older adults with objective RAR assessment and almost 8-years of follow-up, we suggest that suppressed and fragmented daily activity rhythms precede the onset of dementia or MCI and may serve as risk biomarkers for preclinical dementia in middle-aged and older adults.

(JMIR Public Health Surveill 2024;10:e55211) doi:10.2196/55211

**KEYWORDS**
circadian rhythm; dementia; actigraphy; cognitive decline; RAR; rest-activity rhythms; cognitive impairment

### Methods

**Study Population and Data Source**

We used longitudinal data on UKB participants (age range at baseline 43-79 years; 54% female) [28]. Upon enrollment, UKB participants completed a series of questionnaires that collected their demographic, lifestyle, and medical history information. Participants consented to releasing their electronic health records from the United Kingdom’s centralized National Health Service (NHS), which were then stored in the UKB’s Hospital Inpatient Data library [29]. In the UKB cohort, 103,711 participants completed actigraphy assessments between 2013 and 2015 (2.8 to 9.7 years after enrollment) [29]—the baseline of this study. We used follow-up data until September 2021 (maximum and median follow-up after actigraphy: 7.5 and 5 years, respectively). After excluding those participants with poor calibration of activity counts, significant gaps in data likely due to off-wrist periods, <6 days of collected data, dementia or MCI at baseline, or any missing covariate, 91,517 participants were included in this study.

**Ethical Considerations**

The UKB received approval from the North West Multi-centre Research Ethics Committee (11/NW/03820; 16/NW/0274; 21/NW/0157). This study was conducted under the terms of the UKB (33883) and Mass General Brigham Institutional Review Board (#2018P000356).

**Assessment of RARs**

Participants wore triaxial accelerometer devices (Axivity AX3; Axivity Ltd) for up to 7 days during the collection period. Prior actigraphy assessment of older adults [30,31] and existing criteria from the UKB [32] were used to perform quality checks. Activity counts in each 15-second epoch were derived from accelerometer data sampled at ~100 Hz (see Multimedia Appendix 1 [29,33]). The first 6 days of activity counts were used to obtain the following nonparametric RAR measures [34]: (1) activity counts during the most active 10-hour period of the 24-hour cycle (M10) and (2) the midpoint of the M10 period, (3) activity counts during the least active 5-hour period of the 24-hour cycle (L5)—likely representing hours during sleep, (4) the timing midpoint of the L5 period (L5 midpoint), (5) relative amplitude (RA) calculated as (M10-L5)/(M10+L5)—representing the robustness of a 24-hour rest-activity cycle; (6) interdaily stability (IS) that quantifies the stability of the 24-hour rhythm between different days (Multimedia Appendix 2); and (7) intradaily variability (IV) that describes the fragmentation of the rhythm (Multimedia Appendix 2; and (7) intradaily variability (IV) that describes the fragmentation of the rhythm (Multimedia Appendix 2).
Appendix 2). Cosinor analysis was also performed to derive 2 additional measures of 24-hour activity rhythms: the amplitude (midline to peak) and acrophase (time of the peak) of the 24-hour rhythm. All RAR data analyses were performed using the eZActi2 software [35,36].

Assessment of Dementia and MCI

Study participant hospitalization records were kept within the UK’s NHS during the follow-up period before being released by the UKB. The UKB provided algorithmically defined incidence of health matters from ICD-10 (International Classification of Disease, 10th Revision) codes. We obtained data from clinical coding of dementia (ICD-10 code: F05) and MCI (ICD-10 code: F0.67), and from the UKB algorithm “date of all-cause dementia” (field 42018). Age at death or the date of death was based on the death certificates in the NHS. The first occurrence of dementia or MCI (time-to-event) was the first date of diagnosis relative to the actigraphy assessment date.

Assessment of Covariates

The following covariates that may affect RARs were considered in this study: (1) demographics, including age at actigraphy, male or female designated sex of individuals, self-reported ethnicity as European or non-European, college-level education (reported as yes or no), and the Townsend deprivation index; (2) comorbidities, including sleep apnea (based on ICD-10 code G47.30), circulatory disease (based on reports of high cholesterol, diabetes, hypertension, ischemic heart disease, smoking, and peripheral vascular disease), BMI >30, and morbidity burden (classified at the time of actigraphy as none, moderate, or high based on previously used methods that summed the presence of diseases or disorders of the endocrine, connective tissue, gastrointestinal, hematological, musculoskeletal, immune, renal, and respiratory systems as well as any cancers) [37-40]; (3) lifestyle, including alcohol intake (categorized by daily use, 3-4 times per week, 1-2 times per week, a few times per month, and never), smoking status (categorized as current, previous, and never); (4) shiftwork (yes or no); and (5) genetics based on the polygenic risk score (PRS) for Alzheimer disease. We calculated the single PRS for Alzheimer disease using the PRS continuous shrinkage [41] method. This method calculates posterior effect size from genome-wide association study summary statistics with models comprising information of local linkage disequilibrium patterns, and thus reduces PRS error and improves performance. In this study we used genome-wide association study summary statistics from a recent study for Alzheimer disease [42], and the linkage disequilibrium reference panel matrices from the UKB. The PRS continuous shrinkage default settings were used, and after deriving the posterior summary statistics, we used PLINK2 [43,44].

Statistical Analysis

Cox proportional hazard models were used to assess the associations of RAR measures with the subsequent incidence of dementia or MCI. The results were reported as hazard ratios (HRs) with corresponding 95% CIs. For each RAR measure (except M10 midpoint, L5 midpoint, and phase), participants were divided into 4 quartiles (Q1-Q4). The highest quartile (Q4) was used as a reference for RA, M10, IS, and 24-hour amplitude; the lowest quartile (Q1) was used as a reference for L5 and IV. These reference levels were chosen based on prior findings and our hypothesis regarding the direction of the association of each RAR measure with the risk of dementia and MCI, that is, the level of the RAR measure that is hypothesized to be linked to the lowest risk of developing dementia was considered as the reference [37]. Acrophase, M10 midpoint, and L5 midpoint were categorized into 3 tertiles: earlier (6:00 AM-1:34 PM for acrophase, 6:00 AM-12:58 PM for M10 midpoint, and noon-2:47 AM for L5 midpoint), middle (1:34 PM-2:25 PM for acrophase, 12:58 PM-2:01 PM for M10 midpoint, and 2:47 AM-3:44 AM for L5 midpoint), and later (2:25 AM-6:00 AM for acrophase, 2:01 AM-6:00 AM for M10 midpoint, and 3:44 AM-noon for L5 midpoint) groups, and the middle groups were used as the reference [14,15,17]. Separate Cox models were also used to obtain HRs for 1 SD change in each RAR measure. Secondary analyses were performed to investigate (1) the associations between RAR measures and the risk of developing dementia (by excluding participants who only developed MCI); (2) the associations between RAR measures and risk of developing dementia or MCI after excluding those participants who were shift workers at baseline; (3) the interaction effects of Alzheimer disease PRS (<median PRS vs >median PRS) and RAR measures on the risk of developing dementia or MCI (by including interaction terms and also stratifying participants based on their PRS values); and (4) the interaction effects of age (<70 vs ≥70 years) and RAR measures on the risk of developing dementia or MCI (by including interaction terms and also stratifying participants based on their age). All statistical analyses were performed using JMP Pro (version 16, SAS Institute).

Results

Participant Characteristics

Table 1 describes the demographic, lifestyle, and clinical comorbidity data gathered from the 91,517 UKB participants who were included in this study. Most participants were of White European descent (>95%). In comparison to the 90,962 participants who did not develop dementia or MCI, those who did (n=555) were older (69.6 vs 62.4 years) and more likely to be male (n=308, 55.5% vs n=39,758, 43.7%), had lower levels of education (n=215, 38.7% vs n=39,372, 43.3% attended college), had a higher prevalence of sleep apnea (n=10, 1.8% vs n=771, 0.8%), and circulatory system disease (n=261, 47% vs n=22,273, 24.4%), had a higher morbidity burden (2.0, SD 1.8 vs 1.1, SD 1.3), and were more likely to be current or past smokers (n=40, 7.2% vs n=6271, 6.9% and n=342, 61.6% vs n=48,428, 53.2%).
Table 1. Baseline demographics, lifestyle, and clinical comorbidities of participants (n=91,517), by dementia or MCI status at follow-up.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Participants who developed dementia or MCI (n=555)</th>
<th>Participants who did not develop dementia or MCI (n=90,962)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at actigraphy (years), mean (SD)</td>
<td>69.6 (5.4)</td>
<td>62.4 (7.8)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>308 (55.5)</td>
<td>39,758 (43.7)</td>
</tr>
<tr>
<td>Female</td>
<td>247 (44.5)</td>
<td>51,204 (56.3)</td>
</tr>
<tr>
<td>Attended college, n (%)</td>
<td>215 (38.7)</td>
<td>39,372 (43.3)</td>
</tr>
<tr>
<td>Townsend deprivation index (higher), n (%)</td>
<td>274 (49.4)</td>
<td>45,007 (49.5)</td>
</tr>
<tr>
<td>European ethnic background</td>
<td>535 (96.4)</td>
<td>87,951 (96.7)</td>
</tr>
<tr>
<td><strong>Rest-activity rhythmicity characteristics, mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative amplitude</td>
<td>0.956 (0.035)</td>
<td>0.956 (0.035)</td>
</tr>
<tr>
<td>Amplitude (24-hour, AU)</td>
<td>28.0 (15.9)</td>
<td>33.3 (15.9)</td>
</tr>
<tr>
<td>M10 (count)</td>
<td>127,003 (58,873)</td>
<td>149,468 (61,082)</td>
</tr>
<tr>
<td>L5 (count)</td>
<td>2708 (2622)</td>
<td>2367 (2565)</td>
</tr>
<tr>
<td>Phase (hours after midnight)</td>
<td>13.79 (1.18)</td>
<td>14.02 (1.24)</td>
</tr>
<tr>
<td>IV, AU</td>
<td>0.95 (0.24)</td>
<td>0.91 (0.24)</td>
</tr>
<tr>
<td>IS, AU</td>
<td>0.54 (0.13)</td>
<td>0.52 (0.13)</td>
</tr>
<tr>
<td><strong>Comorbidities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sleep apnea, n (%)</td>
<td>10 (1.8)</td>
<td>771 (0.8)</td>
</tr>
<tr>
<td>Circulatory system disease, n (%)</td>
<td>261 (47)</td>
<td>22,273 (24.4)</td>
</tr>
<tr>
<td>BMI &gt; 30 kg/m², n (%)</td>
<td>125 (22.5)</td>
<td>17,586 (19.3)</td>
</tr>
<tr>
<td>Morbidity burden (number of diagnoses), mean (SD)</td>
<td>2.0 (1.8)</td>
<td>1.1 (1.3)</td>
</tr>
<tr>
<td><strong>Alcohol intake, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>160 (28.8)</td>
<td>20,833 (22.9)</td>
</tr>
<tr>
<td>3 to 4 times per week</td>
<td>117 (21.1)</td>
<td>23,643 (26.0)</td>
</tr>
<tr>
<td>Once or twice per week</td>
<td>104 (18.7)</td>
<td>22,830 (25.1)</td>
</tr>
<tr>
<td>Few times per month</td>
<td>114 (20.5)</td>
<td>18,522 (20.4)</td>
</tr>
<tr>
<td>Never</td>
<td>60 (10.8)</td>
<td>5134 (5.6)</td>
</tr>
<tr>
<td><strong>Smoking status, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>40 (7.2)</td>
<td>6271 (6.9)</td>
</tr>
<tr>
<td>Previous</td>
<td>342 (61.6)</td>
<td>48,428 (53.2)</td>
</tr>
<tr>
<td>Never</td>
<td>173 (31.2)</td>
<td>36,263 (39.9)</td>
</tr>
<tr>
<td>Shiftwork, n (%)</td>
<td>18 (3.2)</td>
<td>7351 (8.1)</td>
</tr>
</tbody>
</table>

aMCI: mild cognitive impairment.
bCardiovascular disease means the presence of any of the following: hypertension, high cholesterol, smoking, diabetes, ischemic heart disease, and peripheral vascular disease.
cData come from recruitment between 2.8 and 9.7 years before actigraphy.
dParticipants that scored above the median Townsend deprivation index.
eAU: arbitrary unit.
fM10: activity level of the most active 10-hour period.
gCount: relative mean change in acceleration.
hL5: activity level of the least active 5-hour period.

https://publichealth.jmir.org/2024/1/e55211
Table 2 presents multivariable-adjusted HRs for dementia or MCI associated with RAR metrics when considered as quartiles of exposure or per SD difference. Figure 1 shows survival plots for incident dementia or MCI associated with RAR metrics. The risk of dementia or MCI was statistically higher in those with more suppressed and fragmented 24-hour activity rhythms as quantified by lower RA (multivariable-adjusted HR per 1-SD decrease=1.23, 95% CI 1.16-1.29; Q1 vs Q4, HR 1.88, 95% CI 1.46-2.41; Figure 1A), lower M10 (multivariable-adjusted HR per 1-SD decrease=1.28, 95% CI 1.14-1.44; Q1 vs Q4, HR 1.69, 95% CI 1.30-2.19; Figure 1B), higher L5 (multivariable-adjusted HR per 1-SD increase=1.15, 95% CI 1.10-1.21; Q4 vs Q1, HR 1.51, 95% CI 1.19-1.91; Figure 1C), and larger IV (multivariable-adjusted HR per 1-SD increase=1.14, 95% CI 1.05-1.24; Q4 vs Q1, HR 1.56, 95% CI 1.20-2.02; Figure 1D). Consistently, the risk of dementia or MCI was statistically higher in those with smaller 24-hour amplitude based on cosinor analysis (multivariable-adjusted HR per 1-SD decrease=1.32, 95% CI 1.17-1.49; Q1 vs Q4, HR 1.86, 95% CI 1.42-2.42; Figure 1E). In addition, participants with delayed L5 midpoint had a lower risk for dementia or MCI (multivariable-adjusted HR per 1-SD increase or ~75 min delay in L5 midpoint=0.92, 95% CI 0.85-0.99; Figure 1G). IS (Figure 1F), M10 midpoint (Figure 1H), and acrophase (Figure 1I) had no significant associations with the risk of dementia or MCI.

In addition, we found that the risk was higher for older participants (multivariable-adjusted HR for each year older at baseline=1.17, 95% CI 1.15-1.19) and male participants (multivariable-adjusted HR 1.33, 95% CI 1.11-1.58).

The associations between RAR measures and incident dementia or MCI remained similar when including only participants who developed dementia (Multimedia Appendix 3) or when excluding those shift workers (Multimedia Appendix 4). In addition, the associations between RAR metrics and risk of dementia or MCI were independent of PRS (P values >.10 for the interaction terms of PRS and all RAR measures; Multimedia Appendix 5) while higher PRS was associated with an increased risk of dementia or MCI (multivariable-adjusted HR per 1-SD increase=1.48, 95% CI 1.36-1.60). Moreover, the RAR-dementia or MCI associations appeared to be similar for the younger (<70 years old) and older (≥70 years old) participants (P values >.10 for the interaction terms of age group and all RAR measures, except M10; Multimedia Appendix 6).
Table 2. Relationships of RAR\(^a\) measures with risk of developing dementia or MCI\(^b,c\).

<table>
<thead>
<tr>
<th>RAR characteristics</th>
<th>Adjusted hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relative amplitude</strong></td>
<td></td>
</tr>
<tr>
<td>Q1(^d)</td>
<td>1.88 (1.46-2.41)</td>
</tr>
<tr>
<td>Q2</td>
<td>1.29 (0.99-1.67)</td>
</tr>
<tr>
<td>Q3</td>
<td>0.94 (0.71-1.26)</td>
</tr>
<tr>
<td>Q4</td>
<td>Reference</td>
</tr>
<tr>
<td>Per 1-SD decrease</td>
<td>1.23 (1.16-1.29)</td>
</tr>
<tr>
<td><strong>M10(^e)</strong></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>1.69 (1.30-2.19)</td>
</tr>
<tr>
<td>Q2</td>
<td>1.11 (0.83-1.47)</td>
</tr>
<tr>
<td>Q3</td>
<td>1.03 (0.76-1.38)</td>
</tr>
<tr>
<td>Q4</td>
<td>Reference</td>
</tr>
<tr>
<td>Per 1-SD decrease</td>
<td>1.28 (1.14-1.44)</td>
</tr>
<tr>
<td><strong>M10 midpoint</strong></td>
<td></td>
</tr>
<tr>
<td>Earlier</td>
<td>1.10 (0.90-1.34)</td>
</tr>
<tr>
<td>Middle</td>
<td>Ref</td>
</tr>
<tr>
<td>Later</td>
<td>1.03 (0.83-1.27)</td>
</tr>
<tr>
<td>Per 1-SD increase</td>
<td>0.93 (0.84-1.03)</td>
</tr>
<tr>
<td><strong>L5(^f)</strong></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>Reference</td>
</tr>
<tr>
<td>Q2</td>
<td>1.23 (0.96-1.57)</td>
</tr>
<tr>
<td>Q3</td>
<td>1.00 (0.78-1.30)</td>
</tr>
<tr>
<td>Q4</td>
<td>1.51 (1.19-1.91)</td>
</tr>
<tr>
<td>Per 1-SD increase</td>
<td>1.15 (1.10-1.21)</td>
</tr>
<tr>
<td><strong>L5 midpoint</strong></td>
<td></td>
</tr>
<tr>
<td>Earlier</td>
<td>1.06 (0.87-1.30)</td>
</tr>
<tr>
<td>Middle</td>
<td>Reference</td>
</tr>
<tr>
<td>Later</td>
<td>0.86 (0.70-1.06)</td>
</tr>
<tr>
<td>Per 1-SD increase</td>
<td>0.92 (0.85-0.99)</td>
</tr>
<tr>
<td><strong>IV(^g)</strong></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>Reference</td>
</tr>
<tr>
<td>Q2</td>
<td>1.64 (1.27-2.12)</td>
</tr>
<tr>
<td>Q3</td>
<td>1.37 (1.05-1.78)</td>
</tr>
<tr>
<td>Q4</td>
<td>1.56 (1.20-2.02)</td>
</tr>
<tr>
<td>Per 1-SD increase</td>
<td>1.14 (1.05-1.24)</td>
</tr>
<tr>
<td><strong>IS(^b)</strong></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>1.00 (0.78-1.27)</td>
</tr>
<tr>
<td>Q2</td>
<td>0.82 (0.64-1.05)</td>
</tr>
<tr>
<td>Q3</td>
<td>1.01 (0.81-1.26)</td>
</tr>
<tr>
<td>Q4</td>
<td>Reference</td>
</tr>
<tr>
<td>Per 1-SD decrease</td>
<td>0.98 (0.89-1.07)</td>
</tr>
</tbody>
</table>

Amplitude
<table>
<thead>
<tr>
<th>RAR characteristics</th>
<th>Adjusted hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>1.86 (1.42-2.42)</td>
</tr>
<tr>
<td>Q2</td>
<td>1.28 (0.96-1.70)</td>
</tr>
<tr>
<td>Q3</td>
<td>1.22 (0.91-1.63)</td>
</tr>
<tr>
<td>Q4</td>
<td>Reference</td>
</tr>
<tr>
<td>Per 1-SD decrease</td>
<td>1.32 (1.17-1.49)</td>
</tr>
</tbody>
</table>

**Acrophase**

<table>
<thead>
<tr>
<th>Category</th>
<th>Adjusted hazard ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earlier</td>
<td>0.99 (0.81-1.20)</td>
</tr>
<tr>
<td>Middle</td>
<td>Reference</td>
</tr>
<tr>
<td>Later</td>
<td>0.90 (0.73-1.13)</td>
</tr>
<tr>
<td>Per 1-SD increase</td>
<td>0.93 (0.85-1.02)</td>
</tr>
</tbody>
</table>

*a* RAR: rest-activity rhythm.

*b* MCI: mild cognitive impairment.

*c* Models are adjusted for age at the time of actigraphy, sex, education, Townsend deprivation index, ethnic background, obesity, sleep apnea, morbidity burdens, circulatory disorders, night shiftwork status, alcohol intake, smoking status, and polygenic risk score of Alzheimer disease.

*d* Q: quartile.

*e* M10: activity level of the most active 10-hour period.

*f* L5: activity level of the least active 5-hour period.

*g* IV: intradaily variability.

*h* IS: interdaily stability.
Figure 1. Rest-activity rhythm and risk of developing dementia or MCI. Survival curves for dementia or MCI since baseline (actigraphy assessment) for (A) participants with lower RA (1st quartile) and higher RA (4th quartile), (B) participants with lower M10 (1st quartile) and higher M10 (4th quartile), (C) participants with lower L5 (1st quartile) and higher L5 (4th quartile), (D) participants with lower IV (1st quartile) and higher IV (4th quartile), (E) participants with lower amplitude (1st quartile) and higher amplitude (4th quartile), (F) participants with lower IS (1st quartile) and higher IS (4th quartile), (G) participants with earlier L5 midpoint (1st tertile) and later L5 midpoint (3rd tertile), (H) participants with earlier M10 midpoint (1st tertile) and later M10 midpoint (3rd tertile), and (I) participants with earlier acrophase (1st tertile) and later acrophase (3rd tertile). IS: interdaily stability; IV: intradaily variability; L5: activity level of the least active 5-hour period; M10: activity level of the most active 10-hour period; MCI: mild cognitive impairment; RA: relative amplitude.

Discussion

Principal Findings

In this large, prospective cohort study, we evaluated the association between RAR metrics, derived from wrist actigraphy, and incidence of dementia or MCI during a follow-up of up to 7.5 years. Our results underscore the significance of specific RAR metrics, notably RA, M10, amplitude, L5, L5 midpoint, and IV in delineating the dementia or MCI risk, independent of previously identified risk factors for dementia or cognitive decline.

Unlike previous studies, our analysis is based on a notably large sample size of >94,000 participants, using objective actigraphy assessments for RAR and spanning a broad age range of 43-79 years. This study adopted a longitudinal design with nearly 8 years of follow-up and comprehensively adjusted for known confounders, including demographic, shiftwork status, lifestyle, comorbidity, and genetics, to enhance the robustness of the results. Specifically, we showed that suppressed 24-hour rhythmicity (lower RA and 24-hour amplitude), accompanied by reduced activity levels (M10) during the active phase and increased activity levels during the resting phase (L5) as well as fragmented 24-hour rhythms (greater IV), were linked to the higher risk for dementia or MCI. These results are consistent with previous studies. For instance, using the Rush Memory and Aging Project data, the risk for Alzheimer dementia was higher in those with lower 24-hour amplitude and greater IV.
findings might be explained by unadjusted confounders that delay acrophase, M10, and L5 midpoints with a higher risk [14]. Xiao et al [17] reported a significant linear association of lower L5 midpoint (but not M10 midpoint) with higher risk for dementia (but not MCI) [14] observed an association between earlier L5 midpoint (but not M10 midpoint) and acrophase or M10 mid-time. L5 mid-time is usually related to the timing of sleep, whereas M10 mid-time and acrophase are related to the timing of peak activity. Previous research has similarly reported a lack of significant associations [13,14,17], our research contributes to more definitive insights into these complex associations with a substantially larger sample size. In a related study, Park et al [50] reported higher IS in older adults when compared to younger adults, and interpreted the results as the consequence of changes in daily schedules. This study raises a potential concern about the masking effect of daily schedules on RAR measures, especially IS [51]. An important follow-up question is how reliable IS can be in reflecting intrinsic changes in circadian regulation or predicting or capturing the long-term impacts of acute disturbed 24-hour behavioral cycles on circadian health and related cognitive changes. Future studies including circadian rhythms of other physiological variables or mathematical modeling for estimation of circadian rhythms [52-56] are needed to address the question.

The relationship between the RAR phase and dementia or MCI risk is still inconclusive. Our study identified earlier L5 mid-time as a risk factor for dementia or MCI, but not changes in acrophase or M10 mid-time. L5 mid-time is usually related to the timing of sleep, whereas M10 mid-time and acrophase are usually related to the timing of peak activity. Previous research has yielded varied results. For example, Lysen et al [13] did not observe any association between the circadian phase measured by L5 onset and risk of dementia or MCI, whereas Posner et al [14] observed an association between earlier L5 midpoint (but not M10 midpoint) with higher risk for dementia (but not MCI) [14]. Xiao et al [17] reported a significant linear association of delayed acrophase, M10, and L5 midpoints with a higher risk for dementia or MCI in older women. These inconsistent findings might be explained by unadjusted confounders that influence sleep timing, such as chronotype (ie, preferred sleep time), sleep disorders (eg, insomnia), use of sleep medication, and photoperiod. Future studies should consider controlling for such factors when clarifying relationships between the RAR phase and dementia or MCI.

Clinical Implications and Future Research

Our findings provide insights into the clinical practice and future research in dementia and MCI prevention, screening, and intervention. Specifically, incorporating assessments of sleep and rest-activity patterns into routine health evaluations might be beneficial for middle-aged and older adults. In geriatric care, routine monitoring and management of RAR may help evaluate the factors affecting cognitive health. Educating caregivers and family members about the importance of consistent rest-activity patterns could be incorporated into the home-based care for individuals at risk. Tailoring preventive and therapeutic strategies to individuals based on their RAR characteristics, especially in populations like shift workers, could also be effective. Future research should further clarify the causality of the associations between RAR and cognition, and test whether interventions that improve sleep hygiene, modify light exposure, or adjust physical activity levels can positively impact RAR and, consequently, help prevent or slow cognitive decline.

Strengths and Limitations

The strengths of this study include having a large sample size of more than 94,000 participants; using objective assessments of RAR using actigraphy; controlling for a large number of confounders, including demographic, lifestyle, comorbidity, genetics, and morbidity burden; large age range of participants (aged between 43 and 79 years); and the longitudinal study design with nearly 8 years of follow-up. Limitations of our study are as follows: (1) the majority of participants were of White European descent (>95%), limiting our ability to investigate racial or ethnic differences in the associations; (2) the rate of dementia or MCI events (~550 out of 94,000) appeared to be relatively low due to the overall young age of the participants (median age 63.5 years); (3) we were unable to differentiate between different types of dementia. However, this provides a great opportunity for future studies, when the participants became older, to investigate the long-term association of RAR and different types of dementia or MCI; (4) single-time assessment of actigraphy and covariates did not allow us to examine changes in RAR and dementia risk; (5) internal circadian clocks and environmental factors such as light exposure and social obligations were not assessed or controlled such that it is not possible to separate intrinsic and extrinsic influences on RAR.

Conclusions

We found that altered daily rest-activity patterns were linked to future risk of dementia or MCI, independent of other known risk factors. Monitoring of ambulatory daily motor activity or rest-activity patterns with wearable devices may provide a unique opportunity to identify people at higher risk of dementia or MCI.
Acknowledgments

This research has been conducted using the UK Biobank Resource (33883). This work is supported by the National Institutes of Health (NIH) grant RF1AG059867 to KH; KH is also partially supported by RF1AG064312; SH is partially supported by 5T32HL007901-25; PL is supported by the BrightFocus Foundation (A20208865); LG is supported by the Alzheimer’s Association Clinician Scientist Fellowship (AACSF-23-1148490); CG is supported by the Alzheimer’s Association (AARFD-22-928372) and the American Academy of Sleep Medicine Foundation (290-FF-22). MKR was in part supported by the National Institute for Health and Care Research Manchester Biomedical Research Centre. AI is partially supported by the Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat, supported by Fogarty International Center and the NIH National Institutes of Aging [R01 AG057234, R01 AG075775, R01 AG21051, R01 AG083799, CARDS-NIH], Alzheimer’s Association [SG-20-725707], Rainwater Charitable foundation—Tau Consortium, the Bluefield Project to Cure Frontotemporal Dementia, and Global Brain Health Institute), grants from ANID/FONDECYT Regular (1210195, 1210176, and 1220995; ANID/FONDAP/15150012; and ANID/PIA/ANILLOS ACT210096. The contents of this publication are solely the responsibility of the authors and do not represent the official views of these institutions. The funders had no role in the study design, data collection and analysis, and decision to publish or preparation of this paper. LG and KH are co-senior authors of the manuscript.

Conflicts of Interest

SH reports receiving consulting fees from Achaemenid LLC, unrelated to this project. The other authors declare no competing interests.

Multimedia Appendix 1

Activity count calculation.

[PDF File (Adobe PDF File), 429 KB - publichealth_v10i1e55211_app1.pdf ]

Multimedia Appendix 2

Nonparametric analysis of circadian rest-activity rhythms.

[PDF File (Adobe PDF File), 38 KB - publichealth_v10i1e55211_app2.pdf ]

Multimedia Appendix 3

Effects of rest-activity rhythm (RAR) measures on the risk for developing dementia (excluding those with only mild cognitive impairment; MCI). Among those with no dementia or MCI at baseline (N=91,517), 489 participants developed dementia during the follow-up.

[PDF File (Adobe PDF File), 111 KB - publichealth_v10i1e55211_app3.pdf ]

Multimedia Appendix 4

Effects of rest-activity rhythm (RAR) measures on risk of developing dementia or mild cognitive impairment (MCI) in non–shift workers.

[PDF File (Adobe PDF File), 113 KB - publichealth_v10i1e55211_app4.pdf ]

Multimedia Appendix 5

Effects of rest-activity rhythm (RAR) measures on risk of developing dementia or mild cognitive impairment (MCI) in participants with lower polygenic risk score (PRS) for Alzheimer disease (<median PRS) and participants with higher PRS (>median PRS).

[PDF File (Adobe PDF File), 123 KB - publichealth_v10i1e55211_app5.pdf ]

Multimedia Appendix 6

Effects of rest-activity rhythm (RAR) measures on risk of developing dementia or mild cognitive impairment (MCI) in the younger group (<70 years) and older group (≥70 years).

[PDF File (Adobe PDF File), 123 KB - publichealth_v10i1e55211_app6.pdf ]

References


44. Purcell S, Chang C. Cog Genomics. URL: https://www.cog-genomics.org/plink/2.0 [accessed 2024-04-10]


Abbreviations

HR: hazard ratio
ICD-10: International Classification of Disease, 10th Revision
IS: interdaily stability
IV: intradaily variability
LS: activity level of the least active 5-hour period
M10: activity level of the most active 10-hour period
MCI: mild cognitive impairment
NHS: National Health Service
PRS: polygenic risk score
Q: quartile
RA: relative amplitude
RAR: rest-activity rhythm
UKB: UK Biobank

©Shahab Haghayegh, Chendu Gao, Elizabeth Sugg, Xi Zheng, Hui-Wen Yang, Richa Saxena, Martin K Rutter, Michael Weedon, Agustin Ibanez, David A Bennett, Peng Li, Lei Gao, Kun Hu. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 07.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).
Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Exploring Dynamic Changes in HIV-1 Molecular Transmission Networks and Key Influencing Factors: Cross-Sectional Study

Yan He\textsuperscript{1*}, MSc; Ying Tang\textsuperscript{2*}, MSc; Qun Hua\textsuperscript{2*}, MSc; Xin Li\textsuperscript{3}, MSc; You Ge\textsuperscript{4}, MSc; Yangyang Liu\textsuperscript{4}, MSc; Rong Tang\textsuperscript{5}, MPH; Ye Tian\textsuperscript{2}, MSc; Wei Li\textsuperscript{2}, PhD

\textsuperscript{1}Department of Infection Management, Nanjing Drum Tower Hospital, Nanjing, China
\textsuperscript{2}Children's Hospital of Nanjing Medical University, Nanjing, China
\textsuperscript{3}Nanjing Municipal Center for Disease Control and Prevention, Nanjing, China
\textsuperscript{4}School of Public Health, Southeast University, Nanjing, China
\textsuperscript{5}Nanjing Qixia District Center for Disease Control and Prevention, Nanjing, China
\textsuperscript{*}these authors contributed equally

Corresponding Author:
Wei Li, PhD
Children's Hospital of Nanjing Medical University
No.72, Guangzhou Road
Gulou District
Nanjing, 210008
China
Phone: 86 13851522683
Email: weili126@126.com

Abstract

Background: The HIV-1 molecular network is an innovative tool, using gene sequences to understand transmission attributes and complementing social and sexual network studies. While previous research focused on static network characteristics, recent studies’ emphasis on dynamic features enhances our understanding of real-time changes, offering insights for targeted interventions and efficient allocation of public health resources.

Objective: This study aims to identify the dynamic changes occurring in HIV-1 molecular transmission networks and analyze the primary influencing factors driving the dynamics of HIV-1 molecular networks.

Methods: We analyzed and compared the dynamic changes in the molecular network over a specific time period between the baseline and observed end point. The primary factors influencing the dynamic changes in the HIV-1 molecular network were identified through univariate analysis and multivariate analysis.

Results: A total of 955 HIV-1 polymerase fragments were successfully amplified from 1013 specimens; CRF01_AE and CRF07_BC were the predominant subtypes, accounting for 40.8\% (n=390) and 33.6\% (n=321) of the specimens, respectively. Through the analysis and comparison of the basic and terminal molecular networks, it was discovered that 144 sequences constituted static molecular networks, and 487 sequences contributed to the formation of dynamic molecular networks. The findings of the multivariate analysis indicated that the factors occupation as a student, floating population, Han ethnicity, engagement in occasional or multiple sexual partnerships, participation in anal sex, and being single were independent risk factors for the dynamic changes observed in the HIV-1 molecular network, and the odds ratio (OR; 95\% CIs) values were 2.63 (1.54-4.47), 1.83 (1.17-2.84), 2.91 (1.09-7.79), 1.75 (1.06-2.90), 4.12 (2.48-6.87), 5.58 (2.43-12.80), and 2.10 (1.25-3.54), respectively. Heterosexuality and homosexuality seem to exhibit protective effects when compared to bisexuality, with OR values of 0.12 (95\% CI 0.05-0.32) and 0.26 (95\% CI 0.11-0.64), respectively. Additionally, the National Eight-Item score and sex education experience were also identified as protective factors against dynamic changes in the HIV-1 molecular network, with OR values of 0.12 (95\% CI 0.05-0.32) and 0.26 (95\% CI 0.11-0.64), respectively.

Conclusions: The HIV-1 molecular network analysis showed 144 sequences in static networks and 487 in dynamic networks. Multivariate analysis revealed that occupation as a student, floating population, Han ethnicity, and risky sexual behavior were independent risk factors for dynamic changes, while heterosexuality and homosexuality were protective compared to bisexuality. A higher National Eight-Item score and sex education experience were also protective factors. The identification of HIV dynamic
molecular networks has provided valuable insights into the characteristics of individuals undergoing dynamic alterations. These findings contribute to a better understanding of HIV-1 transmission dynamics and could inform targeted prevention strategies.

(JMIR Public Health Surveill 2024;10:e56593) doi: 10.2196/56593

KEYWORDS
HIV; dynamic; molecular transmission network; influence factors; HIV-1; molecular network; real-time analysis; transmission; Nanjing; infection; gene distance; heterosexual; homosexual; dynamic alteration; dynamic alterations; risk factor

Introduction

AIDS, a sexually transmitted disease, is significantly influenced by the social and sexual network structure and characteristics of people living with HIV [1,2]. Understanding the social and sexual network structure of individuals infected with HIV is important for public health, as it facilitates comprehension of transmission dynamics within the population and informs targeted prevention and control measures [3-5]. Prior research examining the social and sexual network structure of people living with HIV has predominantly used field epidemiological techniques, including questionnaire surveys, peer tracing, and community follow-ups, to delineate the attributes of social and sexual transmission networks [6-9]. However, the extended latency period of HIV and the time gap between infection and diagnosis create a challenge for social communication networks that rely on self-reported data from individuals who are infected [10]. The quality of self-reported data provided by individuals who are infected may be influenced by various factors such as social discrimination, stigma, and privacy concerns. These factors can create barriers to accurate reporting and may result in underreporting or misrepresentation of information. Consequently, conventional epidemiological investigations, including behavioral surveys, exposure history assessments, and contact tracing, encounter difficulties in analyzing the structure and attributes of HIV transmission networks [11]. Alternative approaches and methodologies may need to be considered to overcome these difficulties and obtain a more comprehensive understanding of the HIV transmission network.

The HIV-1 molecular network has garnered extensive acknowledgment as an innovative technique that uses the gene sequences of individuals who are infected with HIV-1 to examine its transmission attributes [12]. It functions as a valuable supplement to the study of HIV-1 social and sexual networks [13-16]. The implementation and progression of the HIV-1 molecular network has introduced novel viewpoints for the prevention and control of HIV/AIDS [17-20]. The molecular network has been used to assess the efficacy of antiretroviral therapy in preventing secondary HIV transmission [21]. This approach is deemed a precise mechanism for HIV/AIDS prevention and control, enabling a more effective use of public health resources [22]. By identifying individuals or groups who are at a higher risk of transmitting HIV-1, targeted interventions can be implemented to prevent the spread of the virus and improve treatment outcomes. Furthermore, extensive investigations can promptly identify individuals with an undiagnosed infection within the transmission network. Moreover, the implementation of timely interventions among individuals who are uninfected in risk networks can effectively impede further dissemination and spread of HIV-1, thereby holding promising prospects for curtailing the spread of HIV and reducing the incidence of new infections. It becomes possible to optimize prevention efforts and achieve better outcomes in controlling the spread of HIV-1 [23].

Prior research on the impact of molecular networks on HIV risk has predominantly concentrated on static network characteristics, assessing the structural attributes of the network at a particular moment. Factors such as network size, density, and individual characteristics (eg, age, sex, occupation, risky sexual behavior, HIV-related knowledge, and attitudes) have been identified as important considerations in understanding HIV transmission patterns and implementing effective prevention and control measures [24,25].

Recent research has underscored the importance of dynamic features within the HIV molecular network [26,27]. Identified actively growing clusters, despite demographic and risk characteristics that may diverge from the overall population, warrant focused intervention prioritization. The dynamics of cluster growth offer valuable guidance for the allocation and prioritization of public health resources, thereby bolstering the utility of networks in assessing intervention efficacy. These dynamic changes intricately intertwine with the personal characteristics and behaviors of individuals who are infected. Real-time analysis of HIV-1 molecular network characteristics provides a framework for resource allocation toward swiftly evolving molecular clusters. This approach facilitates the identification and comprehension of ongoing transformations within molecular networks, thus enabling prompt interventions and response strategies [26,28]. However, prevailing research on molecular networks in China predominantly focuses on static networks at a single timepoint, with limited exploration of dynamic changes [29-32]. Our study specifically targets newly formed and dynamically growing molecular clusters to elucidate the dynamic changes within HIV-1 molecular transmission networks. By analyzing the primary influencing factors driving these dynamics, we aim to unravel the intricate nature of HIV molecular networks. Ultimately, this research endeavor seeks to empower public health initiatives with the capability to more effectively target HIV spread and implement tailored prevention measures through the active monitoring and analysis of molecular network dynamics.

Methods

Participants

From September 1, 2015, to June 30, 2019, this research was conducted in five districts of Nanjing: Qinhuai, Xuanwu, Qixia, Jiangning, and Gulou. The study focused on individuals who

https://publichealth.jmir.org/2024/1/e56593
were newly diagnosed with an HIV infection during this specific time frame. The Nanjing Center for Disease Control and Prevention was responsible for conducting confirmation testing on all individuals infected with HIV included in the study.

**Questionnaire Survey**

A structured questionnaire was used for the one-on-one survey, which took place in a separate room. The privacy of the participants was prioritized throughout the questionnaire survey procedure, and strict measures were implemented to uphold the confidentiality of participants’ personal data. The questionnaire covered various topics related to the participants. These included demographic information such as age, gender, occupation as a student or not, marital status, religion, ethnicity, transmission route, sexual orientation, geographical localities (floating population or not), and sex transmission disease history. Additionally, the questionnaire explored participants’ knowledge and behaviors related to HIV, including condom use, casual sexual partners, multiple sexual partners, anal sex, use of enhancers, experiences with sex education, and HIV/AIDS knowledge, using an 8-item questionnaire (National Eight-Item), which was designed by the Chinese Center for Disease Control and Prevention [33,34] (Multimedia Appendix 1).

**Specimen Collection and Storage**

Blood samples were collected from survey participants using EDTA anticoagulant tubes and transported to the Nanjing Center for Disease Control and Prevention, and the samples were processed within 12 hours. Plasma, lymphocyte enrichment solution, and red blood cells were separated and divided into separate tubes. The aliquoted cryopreservation tubes were labeled and stored at –80 °C.

**HIV-1 Polymerase Region Fragment Amplification**

The amplification of the HIV-1 virus was executed through the use of reverse transcription–polymerase chain reaction and nested polymerase chain reaction methodologies, in accordance with the protocols described earlier [35]. The amplification process was directed toward a segment of the polymerase (Pol) region of the specimen, with a specific focus on the 1-99 amino acids of the protease region and the 1-254 amino acids of the reverse transcriptase region (HXB2: 2253-3312). The length of the amplified sequence was 1060 base pairs. In cases where the amplified sequence was shorter than 1060 base pairs or samples failed to amplify, reamplification and sequencing were performed. HIV drug resistance was demonstrated as with our previous study [35].

**Identification of HIV-1 Subtypes in Nanjing**

The gene sequences that were sequenced with success underwent sorting and splicing procedures, while the possibility of contamination was addressed through the use of the web-based plagiarism detection tool, ElimDupes [36]. This study used the web-based HIV basic local alignment search tool (BLAST) for the identification and genotyping of HIV-1 in Nanjing. Furthermore, gene subtype reference sequences, such as CRF01_AE, CRF07_BC, CRF08_BC, CRF5501_B, B, CRF6701_B, and CRF6801_B, were procured from the HIV sequence database website, and a reference sequence data set was constructed in conjunction with the Nanjing HIV-1 gene sequence data. The gene evolution tree was constructed using an approximately maximum likelihood method in FastTree v2.1 (Morgan N Price) for comparative analysis, and the distribution of HIV-1 genotypes in Nanjing was determined based on the findings of the evolution tree.

**The HIV-1 Molecular Network in Nanjing**

An approximately maximum likelihood phylogenetic tree was constructed using FastTree v2.1. The tree was built under the general time reversible + gamma distribution + proportion of invariable sites nucleotide substitution model. The Shimodaira-Hasegawa test was performed to assess the support value for each node in the phylogenetic tree. A support value of 90% was used as the threshold for considering a node well supported. The genetic distance between pairs of sequences was calculated using the Tamura-Nei 93 (TN93) method; a genetic distance threshold ≤0.045 was used to identify potential transmission clusters [37,38]. The HIV-1 molecular network in Nanjing was constructed by using a combination of phylogenetic tree and gene distance (90% + 0.045) methods, which was demonstrated in our previous study [39]. The visual editing of HIV molecular transmission networks was accomplished using Cytoscape 3.10.1 (National Resource for Network Biology), a widely used platform for constructing molecular networks [40,41].

**Identification of Dynamic Changes in HIV-1 Molecular Networks**

The molecular network constructed from 2015 to 2017 served as the baseline molecular network (initial time point), while the molecular network observed from 2015 to 2019 represented the observed end point molecular network (final time point).

By comparing the baseline and observed end point molecular networks, the following definitions were derived. The static molecular network consists of unchanged molecular clusters and molecular clusters that do not appear in the observed end point molecular network. This means that the structure and members of these clusters remain constant during the observation period. The dynamic molecular network refers to the newly formed molecular clusters observed in the observed end point molecular network and the dynamically growing molecular clusters based on the baseline molecular network (Figure 1). This indicates that the clusters were formed within the observation period and may continue to grow or change over time.
Figure 1. The dynamic changes in molecular networks recognition. (A) shows the baseline molecular network, and (B) shows the observed end point molecular network. The molecular cluster composed of A, B, and C is the unchanged molecular cluster. The molecular cluster composed of H and I do not appear in the observed end point molecular network. The molecular cluster composed of D, E, F, J, and K is a dynamically growing molecular cluster. The molecular cluster composed of L, M, and N is the newly formed molecular cluster. The red spheres represent the dynamic molecular network, while the green spheres represent a statical molecular network.

Statistical Analysis

EpiData software (v3.1; Odense) was used to construct a database. Two individuals each input the questionnaire data into a computer, and the questionnaires’ completeness were verified using telephone follow-ups or secondary surveys to supplement any missing information. The classification data in this study were expressed as a percentage, and the chi-square test was used to compare various groups. Additionally, multivariate analysis was conducted using logistic regression analysis, with a test level of α=.05.

Ethics Approval

Written informed consent was obtained from all patients participating in the study, and the research protocol was approved by the Medical Ethics Committee of Zhongda Hospital of Southeast University (2017ZDKYSB045). Our survey is not anonymous; however, all investigators have signed confidentiality agreements to strictly protect the personal privacy information of the participants. The data used for statistical analysis has been deidentified, with the personal privacy information of the investigators removed. Additionally, each participant will receive a compensation of ¥100 (US $15).

Results

Subtypes of HIV-1 Pol Fragment Genome in Nanjing From 2015 to 2019

In this study, 955 HIV-1 Pol fragments were successfully amplified from 1013 specimens; the success rate of amplification was 94.27%. Subtype identification revealed that CRF01_AE and CRF07_BC were the predominant subtypes, accounting for 40.8% (n=390) and 33.6% (n=321), respectively. The remaining subtypes included URF, CRF67_01B, CRF68_01B, B, CRF55_01B, CRF08_BC, CRF58_01B, CRF59_01B, CRF87_cpx, and C, which accounted for 9% (n=86), 4.6% (n=44), 3.7% (n=35), 3.5% (n=33), 2.3% (n=22), 1.3% (n=12), 0.6% (n=6), 0.4% (n=4), 0.1% (n=1), and 0.1% (n=1), respectively.

Identification of HIV-1 Dynamic Molecular Network in Nanjing From 2015 to 2019

The baseline molecular network is the HIV-1 molecular network in Nanjing from 2015 to 2017 (Figure 2). A total of 295 HIV-1 sequences entered the network, with a clustering ratio of 56.8% (295/591), forming 75 molecular clusters. Among them, there were 38 CRF01_AE molecular clusters (122 sequences); 23 CRF07_BC molecular clusters (90 sequences); 2 CRF08_BC molecular clusters (5 sequences); 2 CRF6701_B molecular clusters (23 sequences); and 1 CRF6801_B molecular cluster (17 sequences), 1 CRF5501_B molecular cluster (8 sequences), 3 B subtype molecular clusters (8 sequences), 4 URF molecular clusters (19 sequences), and 1 CRF5901_B molecular cluster (3 sequences).

The observation end point molecular network is the HIV-1 molecular network in Nanjing from 2015 to 2019 (Figure 3). A total of 565 HIV-1 sequences were included in the network, with a clustering ratio of 59.2% (565/955), forming 124 molecular clusters. Among them, there were 57 CRF01_AE molecular clusters (233 sequences); 34 CRF07_BC molecular clusters (188 sequences); 2 CRF08_BC molecular clusters (6 sequences); 5 CRF6701_B molecular clusters (34 sequences); and 3 CRF6801_B molecular clusters (19 sequences), 3 CRF5501_B molecular clusters (14 sequences), 8 subtype B molecular clusters (20 sequences), 10 URF molecular clusters (46 sequences), and 2 other molecular clusters including 5 sequences (CRF5801_B 2 pieces, CRF5901_B 3 pieces).
By comparing the baseline and observed end point molecular networks, it was discovered that 144 sequences remained static or disappeared; these sequences formed molecular clusters that constituted a static molecular network. On the other hand, 487 sequences were found in newly formed or dynamically growing molecular clusters, contributing to the formation of dynamic molecular networks within the clusters where these sequences were located. In addition, the remaining 324 gene sequences were excluded from the analysis of the dynamic molecular network because they were absent from both the basic molecular network and the end point molecular network.

**Figure 2.** The baseline molecular network. The red spheres represent a dynamic molecular network, while the green spheres represent a statical molecular network. The number in the center of each sphere represents the degree of the molecular network.
The Impact of General Personal Characteristics on the Composition of Molecular Networks

The molecular network composition (dynamic and static) was taken as the dependent variable for univariate analysis. The findings indicated that the proportion of male individuals who were infected (460/596, 77.2%) entering the dynamic molecular network was marginally higher than that of female individuals (180/266, 69%), although no statistically significant difference was observed. Notably, a significantly higher proportion of infected individuals 20 years and younger entered the dynamic molecular network, with 89.2% (58/65) of such individuals being represented, compared to those older than 20 years (420/557, 75.4%); the difference was statistically significant ($P=0.02$). A significantly higher proportion of students who were infected (154/186, 82.8%) were found to have entered the dynamic molecular network compared to nonstudents who were infected (324/436, 74.3%; $P=0.02$). The study found significant variations in the proportion of infected individuals with different sexual orientations entering the dynamic molecular network. Specifically, the proportion of bisexual individuals who were infected entering the network was the highest (214/262, 81.7%). Additionally, the proportion of individuals who were infected from the floating population was significantly higher than that of the nonmigrant population (224/277, 80.9% vs 254/345, 73.6%), and the proportion of Han individuals who were infected entering the dynamic molecular network was higher than that of minority individuals who were infected. The study also observed a higher proportion of infection among single individuals, but there were no statistically significant differences in the proportion of people who were infected with different education levels, infection routes, religion, and history of sexuality (Table 1).
### Table 1. The impact of general personal characteristics on the composition of molecular networks.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Static molecular network (n=144, n (%))</th>
<th>Dynamic molecular network (n=487, n (%))</th>
<th>Chi-square test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8 (30.8)</td>
<td>18 (69.2)</td>
<td>0.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Male</td>
<td>136 (22.8)</td>
<td>460 (77.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age groups (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤20</td>
<td>7 (10.8)</td>
<td>58 (89.2)</td>
<td>6.26 (1)</td>
<td>.02</td>
</tr>
<tr>
<td>&gt;20</td>
<td>137 (24.6)</td>
<td>420 (75.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation as a student</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>112 (25.7)</td>
<td>324 (74.3)</td>
<td>5.27 (1)</td>
<td>.02</td>
</tr>
<tr>
<td>Yes</td>
<td>32 (17.2)</td>
<td>154 (82.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below college</td>
<td>52 (26.0)</td>
<td>148 (74.0)</td>
<td>1.35 (1)</td>
<td>.25</td>
</tr>
<tr>
<td>College and above</td>
<td>92 (21.8)</td>
<td>330 (78.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmission route</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homosexual behavior</td>
<td>114 (23.1)</td>
<td>379 (76.9)</td>
<td>0.01 (2)</td>
<td>.99</td>
</tr>
<tr>
<td>Heterosexual activity</td>
<td>26 (23.4)</td>
<td>85 (76.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4 (22.2)</td>
<td>14 (77.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexual orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>25 (24.3)</td>
<td>78 (75.7)</td>
<td>6.40 (2)</td>
<td>.04</td>
</tr>
<tr>
<td>Homosexual</td>
<td>71 (27.6)</td>
<td>186 (72.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bisexual</td>
<td>48 (18.3)</td>
<td>214 (81.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geographical localities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permanent residents</td>
<td>91 (26.4)</td>
<td>254 (73.6)</td>
<td>4.53 (1)</td>
<td>.03</td>
</tr>
<tr>
<td>Floating population</td>
<td>53 (19.1)</td>
<td>224 (80.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnic minority</td>
<td>10 (45.5)</td>
<td>12 (54.6)</td>
<td>6.38 (1)</td>
<td>.01</td>
</tr>
<tr>
<td>Han (Han Chinese)</td>
<td>134 (22.3)</td>
<td>466 (77.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious belief</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>125 (23.2)</td>
<td>414 (76.8)</td>
<td>0.01 (1)</td>
<td>.95</td>
</tr>
<tr>
<td>Yes</td>
<td>19 (22.9)</td>
<td>64 (77.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not single</td>
<td>111 (26.2)</td>
<td>312 (73.8)</td>
<td>7.10 (1)</td>
<td>.008</td>
</tr>
<tr>
<td>Single</td>
<td>33 (16.6)</td>
<td>166 (83.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex transmission disease history</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>104 (22.7)</td>
<td>355 (77.3)</td>
<td>0.24 (1)</td>
<td>.63</td>
</tr>
<tr>
<td>Yes</td>
<td>40 (24.5)</td>
<td>123 (75.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**The Impact of HIV-1–Related Knowledge and Risky Behaviors on the Molecular Network Composition of People Who Were Infected**

There was no significant difference in the proportion of people who were infected entering the dynamic molecular network with different condom use conditions ($P=.79$). However, the proportion of infected persons with casual sexual partners (381/479, 79.5%) was higher than that of patients without casual partners (97/143, 67.8%; $P=.004$). Moreover, the proportion of individuals who reported having multiple sexual partners and engaged in anal sex entering the dynamic molecular network was significantly higher compared to those who did not report such behaviors; these differences were statistically significant.
Additionally, the proportion of infected persons with qualified scores in the eight national items who had received sex education entering the dynamic molecular network was significantly lower than the proportion of infected persons with unqualified scores in the National Eight-Item and who had not received sex education. However, there was no statistical difference in the proportion of individuals who were infected who reported using enhancers compared to those who did not (P=.96; Table 2).

Table 2. The impact of HIV-1 infection knowledge and risky behaviors on molecular network composition.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Static molecular network (n=144, n (%))</th>
<th>Dynamic molecular network (n=487, n (%))</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condom use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>20 (21.7)</td>
<td>72 (78.3)</td>
<td>1.05 (3)</td>
<td>.79</td>
</tr>
<tr>
<td>Frequently</td>
<td>54 (22.0)</td>
<td>191 (78.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occasionally</td>
<td>53 (25.6)</td>
<td>154 (74.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>17 (21.8)</td>
<td>61 (78.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Casual sexual partners</td>
<td></td>
<td></td>
<td>8.49 (1)</td>
<td>.004</td>
</tr>
<tr>
<td>No</td>
<td>46 (32.2)</td>
<td>97 (67.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>98 (20.5)</td>
<td>381 (79.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple partners</td>
<td></td>
<td></td>
<td>19.54 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>56 (36.1)</td>
<td>99 (63.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>88 (18.8)</td>
<td>379 (81.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anal sex</td>
<td></td>
<td></td>
<td>8.29 (1)</td>
<td>.004</td>
</tr>
<tr>
<td>No</td>
<td>40 (33.1)</td>
<td>81 (66.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>104 (20.8)</td>
<td>397 (79.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of enhancers</td>
<td></td>
<td></td>
<td>0.00 (1)</td>
<td>.96</td>
</tr>
<tr>
<td>No</td>
<td>97 (23.1)</td>
<td>323 (76.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>47 (23.3)</td>
<td>155 (76.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex education</td>
<td></td>
<td></td>
<td>4.56 (1)</td>
<td>.03</td>
</tr>
<tr>
<td>No</td>
<td>38 (18.1)</td>
<td>172 (81.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>106 (25.7)</td>
<td>306 (74.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>National Eight-Item score</td>
<td></td>
<td></td>
<td>6.39 (1)</td>
<td>.01</td>
</tr>
<tr>
<td>Unqualified</td>
<td>25 (15.8)</td>
<td>133 (84.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualified</td>
<td>119 (25.6)</td>
<td>345 (74.4)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Effect of Subtype, Drug Resistance, and Clusters of Differentiation 4 Value on Molecular Network Composition**

The statistical analysis revealed a significant difference in the proportion of infected individuals with a clusters of differentiation (CD4) value <500 cells/mm$^3$ entering the dynamic molecular network (363/459, 79.1%) compared to those with CD4 value ≥500 cells/mm$^3$ (115/163, 70.6%; P=.03). However, no statistically significant differences were observed in the proportion of infected individuals with different genotypes or drug resistance entering the dynamic molecular network. Further details can be found in Table 3.
Table 3. The effect of subtype, drug resistance, and clusters of differentiation (CD4) value on molecular network composition.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Static molecular network (n=144), n (%)</th>
<th>Dynamic molecular network (n=487), n (%)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic subtypes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRF01_AE</td>
<td>62 (25.2)</td>
<td>184 (74.8)</td>
<td>3.00 (2)</td>
<td>.22</td>
</tr>
<tr>
<td>CRF07_BC</td>
<td>40 (19.1)</td>
<td>170 (81.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>42 (25.3)</td>
<td>124 (74.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug resistance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>120 (22.3)</td>
<td>419 (77.7)</td>
<td>1.79 (1)</td>
<td>.18</td>
</tr>
<tr>
<td>Yes</td>
<td>24 (28.9)</td>
<td>59 (71.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4+ (cells/mm³)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;500</td>
<td>96 (20.9)</td>
<td>363 (79.1)</td>
<td>4.92 (1)</td>
<td>.03</td>
</tr>
<tr>
<td>≥500</td>
<td>48 (29.5)</td>
<td>115 (70.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Logistic Regression Multivariate Analysis

Taking the molecular network status of individuals infected with HIV-1 as the dependent variable (dynamic=1, static=0), the variables with statistical significance in univariate analysis (occupation, age, ethnicity, sexual orientation, floating population, casual sexual partner, multiple sexual partners, National Eight-Item scores, anal intercourse, singleness, sex education experience, and CD4) were included in the multivariate analysis. The assignments were as follows: occupation (student=1, nonstudent=0), age (<20 years=1, >20 years=0), ethnicity (Han=1, minority=0), sexual orientation (heterosexual=1, same sex=2, bisexual=3), floating population (yes=1, no=0), casual partner (yes=1, no=0), multiple sexual partners (yes=1, no=0), National Eight-Item score (qualified=1, unqualified=0), anal intercourse (yes=1, no=0), marital status (single=1, not single=0), sex education experience (yes=1, no=0), and CD4 (≥500 cells/mm³=1, <500 cells/mm³=0).

The findings of the multivariate analysis indicate that various factors, including occupation as a student, migrant status, Han ethnicity, engagement in occasional or multiple sexual partnerships, participation in anal sex, and being single, are independent risk factors for the dynamic changes observed in the HIV-1 molecular network, and the odds ratio (OR) values were 2.63 (95% CI 1.54-4.47), 1.83 (95% CI 1.17-2.84), 2.91 (95% CI 1.09-7.79), 1.75 (95% CI 1.06-2.90), 4.12 (95% CI 2.48-6.87), 5.58 (95% CI 2.43-12.80), and 2.10 (95% CI 1.25-3.54), respectively. Conversely, heterosexual and homosexuality appear to be protective factors against such changes when compared to bisexuality, with OR values of 0.12 (95% CI 0.05-0.32) and 0.26 (95% CI 0.11-0.64), respectively. Additionally, the National Eight-Item score and experience with sex education were also identified as protective factors against dynamic changes in the HIV-1 molecular network; the OR values were 0.12 (95% CI 0.05-0.32) and 0.26 (95% CI 0.11-0.64), respectively. Further details can be found in Table 4.

Table 4. Logistic regression analysis of factors influencing the dynamic changes in HIV-1 molecular networks.

<table>
<thead>
<tr>
<th>Variables</th>
<th>β (SE)</th>
<th>Wald chi-square (df)</th>
<th>P values</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupation (student)</td>
<td>0.97 (0.27)</td>
<td>12.64 (1)</td>
<td>&lt;.001</td>
<td>2.63 (1.54-4.47)</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>-2.10 (0.49)</td>
<td>18.73 (1)</td>
<td>.003</td>
<td>0.12 (0.05-0.32)</td>
</tr>
<tr>
<td>Homosexual</td>
<td>-1.35 (0.46)</td>
<td>8.74 (1)</td>
<td>.002</td>
<td>0.26 (0.11-0.64)</td>
</tr>
<tr>
<td>Geographical localities (floating population)</td>
<td>0.60 (0.23)</td>
<td>7.14 (1)</td>
<td>.008</td>
<td>1.83 (1.17-2.84)</td>
</tr>
<tr>
<td>Ethnicity (Han Chinese)</td>
<td>1.07 (0.50)</td>
<td>4.54 (1)</td>
<td>.03</td>
<td>2.91 (1.09-7.79)</td>
</tr>
<tr>
<td>Casual partners (yes)</td>
<td>0.56 (0.26)</td>
<td>4.85 (1)</td>
<td>.03</td>
<td>1.75 (1.06-2.90)</td>
</tr>
<tr>
<td>Multiple sexual partners (yes)</td>
<td>1.42 (0.26)</td>
<td>29.62 (1)</td>
<td>&lt;.001</td>
<td>4.12 (2.48-6.87)</td>
</tr>
<tr>
<td>National Eight-Item score (qualified)</td>
<td>-1.58 (0.33)</td>
<td>23.7 (1)</td>
<td>&lt;.001</td>
<td>0.21 (0.11-0.39)</td>
</tr>
<tr>
<td>Anal sex (yes)</td>
<td>1.72 (0.42)</td>
<td>16.50 (1)</td>
<td>&lt;.001</td>
<td>5.58 (2.43-12.80)</td>
</tr>
<tr>
<td>Marital status (single)</td>
<td>0.74 (0.27)</td>
<td>7.83 (1)</td>
<td>.005</td>
<td>2.10 (1.25-3.54)</td>
</tr>
<tr>
<td>Sex education (yes)</td>
<td>-0.52 (0.24)</td>
<td>4.76 (1)</td>
<td>.03</td>
<td>0.59 (0.37-0.95)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.26 (0.59)</td>
<td>0.19 (1)</td>
<td>.66</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

The study revealed a diverse distribution of HIV-1 subtypes in Nanjing, with CRF01_AE and CRF07_BC being the predominant subtypes. This observation aligns with the global trend of HIV-1 subtype distribution, with CRF01_AE and CRF07_BC commonly found in Asian regions [39,42]. The identification of various other subtypes, including URFs and less prevalent subtypes, emphasizes the genetic complexity of the HIV epidemic in the studied population.

The dynamic changes in the molecular networks were analyzed to understand the dynamics of HIV-1 transmission [43]. The network’s dynamic changes are indicative of potential transmission associations within it. The addition of new gene sequences can disrupt the existing network and give rise to novel networks, thereby revealing the evolving transmission patterns with greater precision. This approach facilitates the real-time monitoring of the molecular network’s dynamic changes, thereby enabling the prompt identification of rapidly expanding or emerging molecular clusters. Such information can inform targeted interventions and the evaluation of prior prevention and control measures [44].

By conducting a multifactor analysis of the HIV-1 molecular network’s dynamic changes, it was observed that individuals who were bisexual, had multiple sexual partners, had casual sexual partners, and engaged in anal sex were more likely to be present in the dynamic molecular network. Men who have sex with men (MSM) in China may face social and familial pressures that lead them to choose to marry women [45] but engage in extramarital same-sex relationships, thereby acting as a bridge for HIV transmission from high-risk groups to the general population [46]. Bisexual individuals have a higher incidence of unprotected sex, lower risk perception, and weaker awareness of self-protection, making them more susceptible to HIV transmission. Furthermore, individuals constrained by family and involved in fixed marital relationships have limited opportunities to engage with fixed same-sex partners, leading them to seek commercial same-sex services when they have sexual needs. Additionally, individuals who engage in sexual activity with multiple partners are prone to inconsistent condom use in contrast to those who engage in sexual activity with only one partner. The emergence of online social platforms has facilitated casual sexual encounters, resulting in a surge of inadvertent sexual behaviors. The failure to consistently use condoms during casual sexual activity significantly heightens the risk of HIV infection and transmission. The physiological vulnerability of the anus renders anal sex a higher-risk activity, and the lack of condom use during same-sex anal intercourse further exacerbates the risk of infection and transmission [47].

The findings of the analysis indicate that individuals with dynamic network connections are more likely to be affiliated with migrant, single, and student populations. The floating population serves as a significant conduit for HIV transmission across various regions. Furthermore, the population in question predominantly comprises sexually active young adults who spend prolonged periods away from their families, resulting in a dearth of both physical and emotional support. This lack of restraint renders them susceptible to engaging in high-risk behaviors related to HIV. The unattached demographic, unencumbered by familial obligations, exhibits a greater propensity for engaging in precarious conduct, such as engaging in casual sexual liaisons and maintaining multiple sexual partners, thereby heightening their susceptibility to contracting HIV and propagating the virus. In contemporary times, college students have emerged as a crucial cohort for HIV prevention and management. Despite possessing a higher level of education, this group manifests a significant disparity between their awareness and conduct. Their inclination toward seeking novelty coupled with the proliferation of diverse social media and mobile apps, which makes finding sexual partners online extremely convenient, further exacerbates this phenomenon. A study conducted by our research team revealed that a majority of university students infected with HIV contracted the virus through male-to-male sexual behaviors, with mobile apps serving as the primary platform for locating male sexual partners. The male-to-male sexual contacts exhibited a lack of awareness regarding the high prevalence of HIV/AIDS in their population, and they do not consistently use condoms during sexual encounters [48]. Higher National Eight-Item scores and sex education experience also emerged as protective factors, suggesting the potential role of knowledge and awareness in reducing transmission dynamics.

In response to these findings, Nanjing has implemented targeted measures to enhance the accessibility of HIV/AIDS pre-exposure prophylaxis (PrEP) and postexposure prophylaxis (PEP) services among MSM, especially the floating population of Nanjing [49,50]. These measures include improving the availability of PrEP and PEP medication and related health care services, developing educational programs to increase awareness and knowledge about PrEP and PEP, and engaging with MSM communities through community-based organizations and support groups.

Conclusion

We conducted HIV molecular network analysis in Nanjing from 2015 to 2019, and CRF01_AE and CRF07_BC were identified as the predominant subtypes. Molecular network analysis revealed dynamic changes in the HIV molecular network over time, with 487 sequences contributing to newly formed or dynamically growing molecular clusters. Multivariate analysis confirmed occupation as a student, a floating population, Han ethnicity, engaging in casual or multiple sexual partnerships, participation in anal sex, and being single as independent risk factors for dynamic network changes, while heterosexuality and homosexuality appeared to be protective factors when compared with bisexuality. Additionally, higher National Eight-Item scores and sex education experience were identified as protective factors against dynamic network changes. The findings emphasize the importance of targeted interventions addressing specific risk factors identified in the study. Strategies focused on education, awareness, and behavioral interventions may contribute to stabilizing or reducing the dynamic changes observed in the HIV-1 molecular network. Understanding the intricate dynamics of transmission networks is crucial for...
designing effective public health measures to control and prevent further spread of HIV.

**Limitations**

This study has the following limitations. First, the definition of dynamic network in this study is only based on the comparison of two time point networks, and multiple time point comparisons should be used to discover nascent networks and dynamic growth networks. Second, although we used the HIV Pol gene sequence from 2015 to 2019 to analyze the dynamic changes in the molecular network, the sample size was not large enough. Third, despite conducting the investigation in a one-on-one format, there remains a possibility of recall bias attributable to HIV-related stigma, compounded by the fact that the time of infection is unknown for many individuals.

**Acknowledgments**

This work was supported by the China Postdoctoral Science Foundation (2022M721682).

**Data Availability**

We have submitted our sequences data to GenBank (submission 2796534), and the data sets are available from WL (weili126@126.com) after the coauthors approve the request.

**Authors' Contributions**

WL and Y Tian were responsible for the conception and design of the study. YH and Y Tang were major contributors to the writing of the original draft. QH and XL were responsible for the acquisition of data. YG and YL contributed to the investigation. YG and RT helped in the analysis and interpretation of data. WL and YH contributed to the final approval of the version to be submitted. All authors read and approved the final version of the manuscript. WL (weili126@126.com) and Y Tian (medtronic@163.com) are co-corresponding authors of this manuscript.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
National Eight-Item questionnaire regarding AIDS prevention.

[DOCX File, 16 KB - publichealth_v10i1e56593_app1.docx]

**References**


https://publichealth.jmir.org/2024/1/e56593

JMR Public Health Surveill 2024 | vol. 10 | e56593 | p.57

(page number not for citation purposes)


Abbreviations

**BLAST**: basic local alignment search tool
**CD**: clusters of differentiation
**MSM**: men who have sex with men
**OR**: odds ratio
**PEP**: postexposure prophylaxis
**Pol**: polymerase
**PrEP**: pre-exposure prophylaxis
**TN93**: Tamura-Nei 93

©Yan He, Ying Tang, Qun Hua, Xin Li, You Ge, Yangyang Liu, Rong Tang, Ye Tian, Wei Li. Originally published in Jmir Public Health and Surveillance (https://publichealth.jmir.org), 29.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in Jmir Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Original Paper

A Metric of Societal Burden Based on Virus Succession to Determine Economic Losses and Health Benefits of China’s Lockdown Policies: Model Development and Validation

Wenxiu Chen1,2, MS; Bin Zhang1, PhD; Chen Wang1, MS; Wei An1, PhD; Shashika Kumudumali Guruge1, MS; Ho-kwong Chui3, PhD; Min Yang1,2, Prof Dr

1National Engineering Research Center of Industrial Wastewater Detoxification and Resource Recovery, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China
2University of Chinese Academy of Sciences, College of Resources and Environment, Beijing, China
3Environmental Protection Department, Hong Kong SAR Government, Hong Kong, China

Corresponding Author:
Wei An, PhD
National Engineering Research Center of Industrial Wastewater Detoxification and Resource Recovery
Research Center for Eco-Environmental Sciences
18 Shuangqing Road
Beijing, 100085
China
Phone: 86 18001155476
Fax: 86 62923541
Email: anwei@rcees.ac.cn

Abstract

Background: The COVID-19 pandemic had a profound impact on the global health system and economic structure. Although the implementation of lockdown measures achieved notable success in curbing the spread of the pandemic, it concurrently incurred substantial socioeconomic costs.

Objective: The objective of this study was to delineate an equilibrium between the economic losses and health benefits of lockdown measures, with the aim of identifying the optimal boundary conditions for implementing these measures at various pandemic phases.

Methods: This study used a model to estimate the half-lives of the observed case fatality rates of different strains. It was based on global infection and death data collected by the World Health Organization and strain sequence time series data provided by Nextstrain. The connection between the health benefits and economic losses brought by lockdown measures was established through the calculation of disability-adjusted life years. Taking China’s city lockdowns as an example, this study determined the cost-benefit boundary of various lockdown measures during the evolution of COVID-19.

Results: The study reveals a direct proportionality between economic losses due to lockdowns and the observed case fatality rates of virus strains, a relationship that holds true irrespective of population size or per capita economic output. As SARS-CoV-2 strains evolve and population immunity shifts, there has been a notable decrease in the observed case fatality rate over time, exhibiting a half-life of roughly 8 months. This decline in fatality rates may offset the health benefits of maintaining unchanged lockdown measures, given that the resultant economic losses might exceed the health benefits.

Conclusions: The initial enforcement of lockdown in Wuhan led to significant health benefits. However, with the decline in the observed case fatality rate of the virus strains, the economic losses increasingly outweighed the health benefits. Consequently, it is essential to consistently refine and enhance lockdown strategies in accordance with the evolving fatality and infection rates of different virus strains, thereby optimizing outcomes in anticipation of future pandemics.

(JMIR Public Health Surveill 2024;10:e48043) doi:10.2196/48043

KEYWORDS
SARS-CoV-2; lockdown; virus succession; benefit; loss; fatality rate; pandemic; blanket lockdown; partial lockdown

https://publichealth.jmir.org/2024/1/e48043

XSL•FO

RenderX
Introduction

Throughout history, there have been significant impacts on the global economy and public health from infectious diseases, such as the Spanish flu, Middle East respiratory syndrome (MERS), the bubonic plague, and the COVID-19 pandemic. Among measures to control the spread of diseases, lockdown is the most effective, as it can slow down the spread of a virus and potentially prevent a surge in cases and deaths in a short time. However, it can also result in significant economic costs, particularly in terms of lost jobs, reduced economic output, and increased government spending on relief measures [1-4]. Therefore, it is necessary to develop a quantitative method to assess the economic losses and health benefits of strict lockdown measures and their social costs in order to determine which levels of lockdown are sustainable. This will allow us to better prepare for future pandemics.

There has been previous international research on the balance between the health benefits and economic losses of pandemic lockdowns [5,6]. Some studies attempted to correlate the characteristic R0 value (effective transmission number) of virus transmission, as calculated using the SIR (susceptible-infected-recovered) model, with economic behavior [7-10]. Additionally, some studies have used discrete selection experiments to explore the lockdown boundary conditions [11-15]. However, while both methods offer insights into achieving balance, the former only considers the impact of a single parameter, the R0, while the latter is influenced by public fatigue with lockdown when determining the balance boundary. China’s prolonged 3-year lockdown measures during the COVID-19 pandemic resulted in noticeable lockdown fatigue, making it impractical to determine the lockdown boundary conditions through discrete selection experiments.

The World Health Organization (WHO) uses the concept of disability-adjusted life years (DALYs) to quantify the disease burden of various symptoms, whereas lockdown strategies lead to a reduction in individual social and work time, leading to significant productivity losses. Thus, time emerges as a crucial metric for quantifying the impact of lockdowns, with all associated factors being translatable into economic losses by time [16,17]. This allows for the comparison of health benefits and economic losses of various epidemic control policies using time-based or economic metrics. On the other hand, as the virus undergoes evolution and mutation, the comprehensive benefits of lockdown measures will change. Accompanying the reduced virulence of the virus and enhanced human immunity, numerous countries progressively eased their lockdown policies, notably Sweden, which was the first to revoke restrictions, in February 2022. China, the first country to face the COVID-19 epidemic, quickly implemented a city lockdown in Wuhan on January 23, 2020. As of December 7, 2022, China adjusted its dynamic “zero-COVID” policy and no longer implemented lockdowns.

Thus, it would be useful to find a quantitative method to determine when and how to implement epidemic control policies.

This study uses an optimization model to estimate the half-lives of the observed case fatality rates (OCFRs) of different strains based on global infection and death data collected by the WHO and strain-sequence time-series data provided by Nextstrain. We establish the connection between the health benefits and economic losses brought by lockdown measures through the calculation of DALYs. Taking China’s city lockdowns as an example, this study determines the cost-benefit boundary of various lockdown measures during the evolution of COVID-19. The main innovations of this research include (1) examining the half-lives of the OCFRs of different strains of the COVID-19 virus, (2) establishing a method for quantifying the economic losses and health benefits of lockdown based on comprehensive calculation of disease burden, and (3) identifying factors influencing the benefits of lockdown. This research will provide an effective tool for the management of COVID-19 and potential future infectious diseases, offering valuable references for government and business decision-making.

Methods

Ethical Considerations

The data used in this article were all obtained from a public website, and the study did not involve animal or human experiments, so an ethics committee review statement was not required.

Data Source

This study sourced the data on the number of SARS-CoV-2 deaths and infections from the WHO [18]. Additionally, temporal evolutionary data on various SARS-CoV-2 strains were acquired from Nextstrain [19].

Estimation Method for the OCFRs of SARS-CoV-2 Strains

The phenomenon of virus mutation succession is dynamic, evolving over time and potentially leading to the emergence of various mutant strains within a specific timeframe. Throughout this timeframe, the total fatality rate is influenced by the OCFRs of diverse strains, as depicted in equation 1:

\[ \text{OCFR} = \frac{\text{Total Deaths}}{\text{Total Infections}} \]

The equation used to calculate the half-life of the OCFR is given below:

\[ t_{1/2} = \frac{\ln(2)}{\lambda} \]

In the above equation, \( t \) refers to a specific time period, \( N_t \) refers to the total population at the time period \( t \), \( P_t \) refers to the overall population infection rate at time period \( t \), \( S_t \) refers to the total OCFR at time period \( t \), \( n \) refers to the number of different strains, \( x_{it} \) refers to the proportion of the \( i \)-th strain during time period \( t \), and \( s_{it} \) refers to the OCFR of the \( i \)-th strain during time period \( t \). This parameter can be obtained through multiple linear regression analysis of the OCFR over time, and the general solution to the equation is \( S_t = s_{it} \). As the proportion of different strains changed over time, this study selected the overall OCFR.
of the virus when the strain reached the maximum prevalence rate as the maximum possible observed fatality rate of the strain.

**Calculation of DALYs**

Using the classic 2-stage disease model recommended by the WHO, DALYs were used to quantify the disease burden of COVID-19:

\[
\text{DALYs} = \sum_{i} \left( \left( P_{i} \times N \right) \times \lambda ight)
\]

In the above equations, DALYs are per person-year (ppy), life-lost years (YLLs) refer to premature mortality ppy, years lived with disability (YLDs) are given ppy, \( i \) refers to different age groups (the whole population is divided into 8 age groups: 0-9 y, 10-19 y, 20-29 y, etc, until 70-80 y and >80 y), \( N \) refers to population number, \( P \) refers to population infection rate, \( S \) refers to OCFR, \( L \) refers to loss of life expectancy caused by death, \( DW \) refers to disease disability weight, \( LA \) refers to duration of the disease, \( P_{seq} \) refers to probability of sequelae, \( DW_{seq} \) refers to disability weight of sequelae, \( LB \) refers to duration of sequelae, and \( \lambda \) refers to the proportion of DALY loss caused by death. Table 1 lists the parameter values.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reference value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population (( N ))</td>
<td>1,425,887,360</td>
<td>[20]</td>
</tr>
<tr>
<td>Virus infection rate (( P ))</td>
<td>( a )</td>
<td>—</td>
</tr>
<tr>
<td>Observed case fatality rate (( S ))</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Course of disease (years) (( LA ))</td>
<td>0.077</td>
<td>[21]</td>
</tr>
<tr>
<td>COVID-19 disability weight (( DW ))</td>
<td>0.051</td>
<td>[21]</td>
</tr>
<tr>
<td>Disability weight of sequelae (( DW_{seq} ))</td>
<td>0.219</td>
<td>[22]</td>
</tr>
<tr>
<td>The incidence of sequelae (( P_{seq} ))</td>
<td>0.058</td>
<td>[23]</td>
</tr>
<tr>
<td>Course of sequelae (years) (( LB ))</td>
<td>0.167</td>
<td>[23]</td>
</tr>
<tr>
<td>Lockdown time (days) (( T ))</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Per capita gross national product (dollars per year)</td>
<td>15,308.712</td>
<td>[20]</td>
</tr>
<tr>
<td>Control crowd ratio (( a ))</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Productivity weight (( \omega ))</td>
<td>—</td>
<td>[21]</td>
</tr>
</tbody>
</table>

\( a \) Not applicable.

**Lockdown Economic Calculations**

There are 2 types of lockdown methods: blanket lockdown and local or partial lockdown. The former covers a region like an entire province, city, or district to stop many social activities as a whole; the latter covers small-scale entities such as buildings, corridors, and individual families. When an individual lived under lockdown, without mobility or direct contact with the outside world, contact with the virus was effectively avoided. Therefore, the benefit of implementing lockdown measures for a certain period of time was to reduce the population’s disease burden caused by virus infection. The disease burden and economic burden caused by large-scale infection with different strains of SARS-CoV-2 among the population during the epidemic can be calculated using equations 4 and 8. Typically, the disease burden under the condition of total infection or herd immunity can be considered as the possible benefit gained from implementing a lockdown. The difference between a blanket lockdown and local or partial lockdown lies in the scope. For a partial lockdown, the minimum benefit is to prevent the disease burden caused by local population infection from reaching the level of herd immunity, while the maximum possible benefit is to reduce the disease burden within the entire jurisdiction of the government or even globally, as the lockdown can prevent the virus from spreading worldwide. Particularly for highly contagious and fatal viruses such as Ebola and MERS, localized lockdowns can effectively prevent widespread transmission. For a blanket lockdown, the maximum benefit is to ensure that no one in a city becomes infected. Because local governments are responsible for their own economic gains, the product of the gross domestic product (GDP) output per person per year and the loss adjusted per year is used as a measure of the value of disease burden reduction [21,24]:

\[
\text{Economic Loss} = \text{GDP per person per year} \times \text{Loss adjusted per year}
\]

In the above formula, refers to the productivity weight. The values of \( \omega \) are shown in Table 1.

**Estimated Losses From Lockdown**

One of the major losses due to lockdown measures is the decline in GDP resulting from the reduction in social activities and mandatory remote work. According to calculations by the Chinese University of Hong Kong, a 2-week lockdown led to a 31% reduction in GDP for the corresponding month [20]; thus, a 1-month lockdown would be expected to reduce the GDP for that month by approximately 60%. Although lockdown measures have been associated with the exacerbation of chronic diseases.
and other side effects such as mental illness, this study does not address these aspects due to the lack of sufficient data. Therefore, the loss caused by lockdown can be calculated as follows:

\[ \text{Loss} = N \times T \]

In the above equation, \( N \) refers to the total population and \( T \) refers to the number of lockdown days. The values of \( N \) and \( T \) are shown in Table 1.

**Calculation of the Balance Between Economic Losses and Health Benefits of Blanket Lockdowns**

The health benefits of blanket lockdowns can be calculated by the maximum disease burden caused by overall infection or herd immunity (equation 10). The reduction in GDP resulting from the decrease in effective production time is commensurate in value.

\[ \text{GDP} = N \times P \]

In formulas 9, 10, and 11, \( T \) refers to the number of lockdown days, \( N \) refers to the total population under lockdown, GDP exists on both sides of the equation, and \( P \) refers to the infection rate based on the entire population. Table 1 lists the parameter values.

**Calculation of the Balance Between Economic Losses and Health Benefits of Partial Lockdowns**

Similarly, the balance between economic losses and health benefits of a partial lockdown refers to the equivalence between the health benefits resulting from the lockdown and the losses in GDP caused by the lockdown. The cost of the lockdown is primarily influenced by the size of the population affected; that is, there is a certain proportion of individuals in close contact with COVID-19 cases who cease their activities.

In equations 12 and 13, \( a \) refers to the number of people quarantined around an infected person. When \( a=1 \), only the infected person is quarantined. When \( a>1 \), people who have not been infected may also be restricted from activities. The number of close contacts, subclose contacts, and other companions (not necessarily infected) is determined when a positive infection is found, and there is a redundant cost. \( n \) refers to the number of people in a community, \( P \) refers to infection rate, and GDP per capita appears on both sides of the equation. Where life expectancy (\( L \)), \( \lambda \), and \( \omega \) are constants, the benefit and loss balance boundary is inversely proportional to the lockdown time (\( T \)) and lockdown population ratio (\( a \)) under the condition that the OCFR (\( S \)) is relatively stable. Table 1 lists the parameter values.

**Results**

**Time Series Trends in OCFRs of Different Strains**

Based on global mortality data from the WHO since December 2019 and lineage data provided by Nextstrain, Figure 1 illustrates an exponential decline over time in the OCFR of SARS-CoV-2, with a half-life of approximately 8 months. The OCFR for the original SARS-CoV-2 strain was between 3.8% and 8.4%, which then decreased to 2%-3% for the Beta, Gamma, and Lambda variants, and further dropped to 1%-2% for the Delta variant. During the Omicron phase, the OCFR significantly decreased, to about 0.5%. As of January 2023, the OCFR of SARS-CoV-2 in China has been reduced to below 0.1%. According to a report in *The Lancet*, the global OCFR for seasonal influenza from 1999 to 2015 ranged between 0.04% and 0.08%, nearing 0.1% for those older than 75 years [25]. These findings suggest that the OCFR of SARS-CoV-2 is now comparable to that of seasonal influenza. Strains with higher OCFR typically depend significantly on the host’s immune response, leading to reduced reproduction and survival rates, and thus become less competitive compared to less virulent but highly transmissible strains, eventually being replaced over time [26,27].
Balance Between Economic Losses and Health Benefits With Different Strains in a Blanket Lockdown

In the absence of prevention and control measures, SARS-CoV-2 will continue to infect individuals until herd immunity is achieved. For instance, on December 7, 2022, China adjusted its dynamic zero-COVID policy and no longer implemented lockdowns, leading to concentrated outbreaks of infections. In such a scenario, more than 70% of the population would form an immune barrier after being infected, preventing the virus from spreading continuously [28,29]. Therefore, the maximum loss resulting from complete liberalization is the infection of the entire population. Conversely, the greatest benefit lockdowns can offer to self-sustaining local governments is ensuring that no one within the jurisdiction becomes infected. Different strains have different disease burdens due to their different OCFRs. As the OCFRs of the strains decrease, the disease burden of the population also gradually decreases.

Based on the OCFR of each virus strain, the maximum disease loss can be estimated. In Figure 2, the left vertical axis represents the benefits brought by lockdown and the right vertical axis represents the time of the lockdown. Combined with Figure 1, it is clear that the quantitative value of disease losses (measured in DALYs) also declines as the OCFR declines. The maximum benefits from lockdown decrease as the virus strain’s OCFR decreases over time. Using equation 10, the net gain from lockdown can be calculated when health benefits exceed economic losses. For example, during the early lockdown period in Wuhan, involving strains 19A, 19B, and 20I, the health benefits of a 76-day lockdown surpassed the cost line.

Figure 1. Trends in the observed case fatality rate of SARS-CoV-2.

Figure 2. Relationship diagram of lockdown scale and lockdown duration during partial lockdown. DALY: disability-adjusted life year; ppy: per person-year; YLD: year lived with disability; YLL: life-lost year.
The results of the benefit and loss balance calculation indicate that the benefits yielded by the city lockdown strategy are correlated only with the OCFR and the lockdown duration, exhibiting a linear relationship between these 2 factors at the balance point of benefit and loss. As shown in Figure 2 (upper graph), there are 2 distinct zones on either side of the benefit and loss balance line: the benefit zone and the loss zone. When the OCFR is low, a shorter lockdown duration can produce health benefits. Conversely, with a higher OCFR, benefit is achievable even with an extended lockdown. For instance, when the OCFR in Wuhan reached 7% in January 2020, a lockdown period of up to 76 days could still have been beneficial. For the bubonic plague, which has an OCFR of up to 25% [30], a lockdown of more than a year could still produce a strong benefit. However, benefits tend to decline with succeeding virus strains. If the OCFR of a virus strain continues to decline, accompanied by higher false-negative nucleic acid tests, the duration of a lockdown will increase. When the OCFR drops to about 1%, a blanket lockdown is more likely to produce a loss.

Optimization of Local Lockdown Strategy

Partial lockdowns have lower cost than blanket lockdowns. According to equation 13, the balance between benefits and losses in a local lockdown is mainly determined by 3 factors: the duration of the lockdown, the size of the covered population, and the OCFR. The benefit from lockdowns primarily comes from reducing the disease losses caused by the epidemic, while losses are mainly due to the reduction in social per capita output, influenced by both the scale and duration of the lockdown. The ratios for different lockdown groups can vary, such as 1:3 (a family of three), 1:5 (5 mixed with 1), 1:10 (10 mixed with 1), 1:20 (20 mixed with 1), 1:40 (a unit), 1:200 (a building), and 1:2000 (a community) for closed management. If benefit calculations only consider the avoidance of infection among isolated individuals, redundant costs arise, as a certain proportion of uninfected individuals in the locked-down area will also waste socially effective labor time. The hyperbolic relationship between population ratio and duration is depicted in Figure 3, in which the upper part of the curve corresponds to losses and the lower left indicates benefits; the darker the color, the greater the likelihood of generating a benefit. When the OCFR is high, the area for benefits (the lower left of the balance line) becomes larger, indicating more flexibility in the scale and duration of the lockdown. When the OCFR is 10% and 1 person in a family of 4 is infected, the maximum lockdown time is 26 days. At this time, in order to reduce the lockdown rate and increase the lockdown benefit, the infected people can be centrally isolated, such as in a shelter. In January 2023, the OCFR for Omicron was less than 0.5% [18], which means that if more than 3 people in close contact were quarantined (with 3 nucleic acid tests mixed with 1), then a state of loss would begin after 2 days, and the quarantine and control measures would be unproductive.

Discussion

Principal Results

The management of COVID-19 health emergencies is a huge public challenge, and as of November 28, 2022, the global COVID-19 pandemic had affected a total of 640 million individuals and resulted in 6.63 million fatalities. In China, 9.63 million cases were recorded and 30,000 deaths reported. [20] Remarkably, the lockdown measures implemented in China demonstrated a significant impact, resulting in an estimated reduction of 100 million COVID-19 cases and 1.16 million deaths compared to the global average. China’s approach has thus proved highly effective in the battle against the pandemic.
and yielded substantial gains. Thus, this paper contributes to documenting and characterizing the history of the pandemic from an epidemiological perspective, refining knowledge about pandemics, enhancing the exploration of the effectiveness of pandemic response measures, establishing the boundary conditions for exiting a strong pandemic response, shortening the emergency transition time of the early response, and providing a knowledge base for designing and framing future pandemic scenarios.

This study discusses 2 types of lockdown strategies: blanket lockdowns and partial lockdowns. In the early stages of a pandemic, when detection levels are low, a blanket lockdown can delay the spread of the epidemic, thereby easing the burden on hospitals and allowing medical institutions time for emergency preparations. When the number of infected people is small and concentrated, improved detection capabilities, such as nucleic acid detection and other methods, can accurately locate and identify infected individuals. Then, a partial lockdown can gradually replace the blanket lockdown. When a person is confirmed to be infected, close contacts (e.g., direct contacts) or secondary contacts (e.g., indirect contacts) can be identified based on epidemiological findings. The isolation ratio boundary calculated in this study determines which type of contact should be isolated. However, lockdowns based solely on mass nucleic acid screening may not be suitable for all countries. From the perspective of survival and evolution, viruses tend to evolve toward decreased virulence and increased transmissibility. Therefore, with an increasing virus transmission rate and the presence of false negatives in nucleic acid testing, lockdown measures assisted by such screening become largely ineffective, and gatherings for mass nucleic acid testing increase the risk of infection. At this stage, zero transmission is basically impossible to achieve. When herd immunity has not been achieved, long-term, large-scale lockdowns and partial lockdowns are ineffective and only slow the spread of the epidemic, rather than stop it, while imposing a significant economic burden. In this context, sewage virus detection has emerged as a crucial complement. Experience gained from sewage virus monitoring in Hong Kong shows that this method has high sensitivity, reliably detecting 1 positive patient per 20,000 to 40,000 people at a 95% confidence level. Additionally, it can forecast the number of infected individuals 1 to 4 days in advance. Sewage virus detection not only reduces the cost of nucleic acid aggregation testing and the risk of infection but also minimizes the negative impact on productivity and daily life. Each sewage monitoring point can save the effective social time of around 30,000 people who would otherwise be waiting in queues for nucleic acid testing, thus promoting production and decreasing the spread of the epidemic. With many countries worldwide working toward establishing sewage virus monitoring systems, it is foreseeable that this approach will play a critical role in global epidemic prevention and control in the future.

Limitations
This study has certain limitations. First, lockdown strategies based on large-scale nucleic acid screening, as implemented in China, may be relatively uncommon in other countries due to factors such as strong government capacity for management and implementation, public cooperation, and the economic ability to bear the associated costs. Consequently, these strategies may not be feasible for all countries and regions. While large-scale nucleic acid screening can effect and accurately identify infected individuals, false-positive results from nucleic acid tests may undermine the effectiveness of lockdown measures. Furthermore, queues and gatherings of individuals awaiting nucleic acid testing could undermine social distancing efforts and increase the risk of infection.

Second, the calculation of the half-life of the OCFR in this study is based on values derived from GISAID (Global Initiative on Sharing All Influenza Data) statistics rather than the virus’s inherent fatality rate. The OCFR is affected by various factors at different stages of human infection, such as the capacity for detection and screening, active infection (e.g., vaccination), and the enhancement of immunity due to repeated passive infections. While the WHO mandates the inclusion of asymptomatic infected individuals in the case fatality rate calculations, the deficiency in early screening capabilities across different countries and regions resulted in an overestimation of the OCFR’s actual impact on mortality. Conversely, the OCFR tends to underestimate the true mortality rate of the strain when there is an improvement in immunity during later stages. Consequently, this introduces a certain degree of uncertainty.

Last, the model is calculated on the basis of short-term economic losses. However, the long-term economic impact of COVID-19 lockdown measures is complex and profound. The economic pressures imposed on society and households by these measures often result in increased drug abuse, mental illness, domestic violence, and even suicide. Given the significant individual variations in these outcomes, our research team is currently tracking the long-term effects of COVID-19. However, we have yet to accumulate long-term cohort tracking data. In future research, we will consider this aspect using more sophisticated models.

Conclusions
This study shows that with the continuous enhancement of detection capabilities and the continuous formation of population immunity, the OCFR of COVID-19 decreases exponentially, and the half-life is about 8 months. In the infectious disease lockdown strategy model, there is a close correlation between the benefits of lockdown and OCFR. During the early transmission period of SARS-CoV-2, especially under conditions of high OCFR, the lockdown strategy in Wuhan brought huge health benefits to the population. However, as the OCFR decreases, these health benefits gradually diminish and eventually transition into losses. Although large-scale lockdown strategies may not be universally applicable for political reasons, they appear to be the least disruptive option in the early stages of an outbreak. The partial lockdown model mentioned in this study can provide a reference for various countries and regions. This study establishes a connection between saving lives and
protecting the economy, indicating that these 2 objectives are not opposites.

This study provides a critical scientific basis for policy makers, guiding the appropriate timing for the implementation or relaxation of lockdown measures. It offers essential insights for formulating public health management strategies related to COVID-19 and future pandemics, providing indispensable reference information for governments and decision-making entities. By integrating thorough risk assessments with economic impact analyses, our research aims to develop more effective and compassionate approaches to pandemic response, with dual objectives: to protect public health and safety and to minimize negative impacts on the socioeconomic landscape.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (52091545) and the National Key Research and Development Project of China (2023YFC3041300).

Authors’ Contributions

WC and WA designed the study, did the analyses, interpreted the findings, and drafted and edited the manuscript. BZ and CW collected the data, SKG modified the grammar of the article, and MY and H-kC critically revised it for intellectual content. All authors were responsible for the decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

References


Abbreviations

DALY: disability-adjusted life year
GDP: gross domestic product
GISAID: Global Initiative on Sharing All Influenza Data
MERS: Middle East respiratory syndrome
OCFR: observed case fatality rate
ppy: per person-year
SIR: susceptible-infected-recovered
WHO: World Health Organization
YLD: year lived with disability
YLL: life-lost year
Original Paper

Optimal Systolic Blood Pressure for the Prevention of All-Cause and Cardiovascular Disease Mortality in Older Adults With Hypertension: Nationwide Population-Based Cohort Study

Jae Jun Lee¹, PhD; Kyung Hee Lee¹, PhD
Mo-Im Kim Nursing Research Institute, Yonsei University College of Nursing, Seoul, Republic of Korea

Corresponding Author:
Kyung Hee Lee, PhD
Mo-Im Kim Nursing Research Institute
Yonsei University College of Nursing
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 222283321
Email: kyungheelee@yuhs.ac

Abstract

Background: Target systolic blood pressure (SBP) levels for older adults with hypertension vary across countries, leading to challenges in determining the appropriate SBP level.

Objective: This study aims to identify the optimal SBP level for minimizing all-cause and cardiovascular disease (CVD) mortality in older Korean adults with hypertension.

Methods: This retrospective cohort study used data from the National Health Insurance Service database. We included older adults aged 65 years or older who were newly diagnosed with hypertension and underwent a National Health Insurance Service health checkup in 2003-2004. We excluded patients who had a history of hypertension or CVD, were not prescribed medication for hypertension, had missing blood pressure or any other covariate values, and had fewer than 2 health checkups during the follow-up period until 2020. We categorized the average SBP levels into 6 categories in 10 mm Hg increments, from <120 mm Hg to ≥160 mm Hg; 130-139 mm Hg was the reference range. Cox proportional hazards models were used to examine the relationship between SBP and all-cause and CVD mortalities, and subgroup analysis was conducted by age group (65-74 years and 75 years or older).

Results: A total of 68,901 older adults newly diagnosed with hypertension were included in this study. During the follow-up period, 32,588 (47.3%) participants had all-cause mortality and 4273 (6.2%) had CVD mortality. Compared to older adults with SBP within the range of 130-139 mm Hg, individuals who fell into the other SBP categories, excluding those with SBP 120-129 mm Hg, showed significantly higher all-cause and CVD mortality. Subgroup analysis showed that older adults aged 65-74 years had higher all-cause and CVD mortality rates according to SBP categories than those aged 75 years or older.

Conclusions: The SBP levels within the range of 120-139 mm Hg were associated with the lowest all-cause and CVD mortality rates among older Korean adults with hypertension. It is recommended to reduce SBP to <140 mm Hg, with 120 mm Hg as the minimum value for SBP, for older Korean adults with hypertension. Additionally, stricter SBP management is required for adults aged 65-74 years.

(JMIR Public Health Surveill 2024;10:e52182) doi:10.2196/52182

KEYWORDS
aged; blood pressure; cardiovascular diseases; hypertension; mortality; older adults; geriatric; elderly; cardiovascular; Korea; Korean; insurance; cohort study; systolic; risk; aging; health outcome
Introduction

The older adult population is growing globally at a faster rate than that in previous years [1]. The World Health Organization expects the number of older adults to rise from 1 to 2 billion between 2019 and 2050, and in 2050, 22% of the world’s population is expected to comprise older adults [1]. In Korea, the aging population is progressing more rapidly than in other countries; it is expected that the proportion of the older adult population will exceed 20% in 2025, 30% in 2035, and 40% in 2050 [2]. To maintain quality of life with longevity, managing chronic disease is important in this population.

Cardiovascular disease (CVD), such as myocardial infarction, stroke, and heart failure, is the leading cause of death worldwide [3], and hypertension is one of the primary modifiable risk factors for CVD mortality [4]. Hypertension is very common in older adults because their systolic blood pressure (SBP) tends to be elevated due to increased stiffness and decreased elasticity in the arteries with aging [5]. Epidemiologic research by the United States National Health and Nutrition Examination Survey has indicated that 76.5% of older adults aged older than or equal to 65 years have hypertension [6]. Moreover, older adults with hypertension are more susceptible to CVD than younger adults with hypertension [7]; therefore, controlling the blood pressure of older adults is extremely important to prevent CVD.

However, the guidelines for target blood pressure levels in older adults with hypertension are inconsistent across countries. In 2017, the American College of Cardiology and the American Heart Association (ACC/AHA) revised the guidelines for the diagnostic criteria of hypertension from 140/90 mm Hg to 130/80 mm Hg. They recommend a target SBP level of <130 mm Hg in older adults with hypertension [8]. These guidelines are equivalent to the 2022 Taiwan Hypertension Guidelines of the Taiwan Society of Cardiology and the Taiwan Hypertension Society for the management of hypertension, recommending that blood pressure treatment in older adults should be initiated when SBP is ≥130 mm Hg; the target SBP level is <130 mm Hg [9]. On the other hand, the European Society of Cardiology and the European Society of Hypertension guidelines of 2018 maintain the previous criterion of hypertension as >140/90 mm Hg and recommend a target SBP range of 130-139 mm Hg, while advising against lowering it below 130 mm Hg due to potential risks without additional benefits [10]. This is similar to the 2019 Chinese guideline for the management of hypertension in older people, which recommends a target blood pressure level of <140/90 mm Hg [11]. The differences in target SBP levels across countries are related to variability in the prevalence of hypertension and CVD mortality rates among different racial and ethnic groups or regions [6,12,13]. Therefore, the diagnostic and treatment guidelines of hypertension are different across countries, making it difficult to achieve optimal blood pressure management in older adults and leading to challenges in determining the appropriate interventions.

In Korea, the 2022 guidelines for the management of hypertension published by the Korean Society of Hypertension (KSH) still maintain the diagnostic criterion for hypertension as >140/90 mm Hg [14]. Additionally, they recommend lowering SBP to <140 mm Hg for older adults with hypertension [14]. However, the KSH guidelines have a limitation in that they are based on research findings that may not specifically include data from older adults in Korea [15-18]. Therefore, further studies are needed to explore the optimal blood pressure levels for older Korean adults with hypertension.

In addition, previous studies conducted to determine the optimal blood pressure levels for older adults with hypertension have been limited by not investigating the impact of age differences. The 2019 Japanese Society of Hypertension guidelines recommend different target blood pressures for different age categories—<130/80 mm Hg in patients aged younger than 75 years and <140/90 mm Hg in those aged 75 years and older. This is because lowering SBP to <130 mm Hg in individuals aged 75 years or older may potentially result in adverse health outcomes, such as renal dysfunction [19]. However, to the best of our knowledge, no studies have identified age differences in optimal SBP levels in older Korean adults. Analyzing the impact of age-related differences in SBP on all-cause and CVD mortalities among older Korean adults with hypertension is an important indicator of future blood pressure management. Therefore, this study aimed to investigate (1) the effects of SBP on all-cause and CVD mortalities in older adults with hypertension and (2) the associations stratified by age (65-74 years and 75 years or older).

Methods

Study Design and Sample

This retrospective cohort study used data from the National Health Insurance Service (NHIS) database. The NHIS is a single obligatory social insurance system covering health care service costs in Korea. The entire population residing in Korea subscribes to mandatory social insurance, and all medical care institutions claim medical fees. The claims include medical information, such as diagnosis, prescription, medication, and data collected from the NHIS database. Because the NHIS also provides general health checkups biennially for all insured adults (annually for nonoffice workers) and collects health checkup data, the NHIS database is useful for continuously observing the health status of the Korean population. The NHIS database was initiated in 2002 and we followed up data until 2020.

We included older adults who (1) were aged 65 years or older, (2) were newly diagnosed with hypertension (Korean Classification of Disease [KCD] code I10), and (3) underwent the NHIS health checkup in 2003-2004. We excluded patients who (1) had a history of hypertension or CVD, (2) were not prescribed medication for hypertension, (3) had missing blood pressure or any other covariate values, and (4) had fewer than 2 health checkups during the follow-up period. Characteristics of the study population are shown in Figure 1. A total of 68,901 participants were eligible and were followed up until 2020.
Ethical Considerations
Ethics approval was obtained from the institutional review board of Yonsei University Health System (4-2021-1277), and the study complied with the Declaration of Helsinki. The NHIS database was anonymized to ensure confidentiality, and informed consent was not required because this was a retrospective study using the NHIS database.

Measurement
The independent variable in our study was the mean SBP measured during the health checkups. Blood pressure was measured using an automatic blood pressure monitor placed in front of the examiner after the participant had stabilized for at least 5 minutes. If the SBP was >120 mm Hg or the diastolic blood pressure was >80 mm Hg, the blood pressure measurement was repeated after at least 2 minutes. In this study, we categorized the SBP into 6 categories in 10 mm Hg increments from <120 mm Hg to ≥160 mm Hg, and 130-139 mm Hg was used as a reference category.

The dependent variables were all-cause and CVD mortalities. The CVD mortality was defined as death from myocardial infarction (KCD codes I21-23), heart failure (KCD code I50), hemorrhagic stroke (KCD codes I60-62), or ischemic stroke (KCD code I63). Data on the dates and causes of mortality were obtained from Statistics Korea.

Sociodemographic characteristics, health behaviors, blood test results, and hypertension medications were included as covariates. The sociodemographic characteristics included age, sex, and BMI. Age was classified as 65-74 years and 75 years or older. Health behaviors included smoking status, physical activity, and the Charlson Comorbidity Index. Blood test results included fasting glucose, total cholesterol, aspartate transaminase, and alanine transaminase. Hypertension medications included renin-angiotensin system blockers (angiotensin-converting enzyme inhibitors and angiotensin receptor blockers), beta-blockers, dihydropyridine calcium channel blockers, and diuretics.

Statistical Analysis
Descriptive analyses were used to describe sample characteristics. Continuous variables were described as means (SDs), and categorical variables were described as numbers and percentages. Thereafter, Kaplan-Meier curves and Cox proportional hazards models were used to identify the relationship between SBP and all-cause and CVD mortalities after adjusting for other covariates. A subgroup analysis was performed to identify age-related differences in the risk of all-cause and CVD mortalities according to SBP. All statistical analyses were performed using SAS (version 9.4; SAS Institute Inc), and the significance level was set at P<.05.

Results
In total, 68,901 older adults newly diagnosed with hypertension were included in this study. We divided them into the following SBP categories: 5666 (8.2%) as <120 mm Hg, 17,236 (25.0%) as 120-129 mm Hg, 25,277 (36.7%) as 130-139 mm Hg, 14,299 (20.8%) as 140-149 mm Hg, 4663 (6.8%) as 150-159 mm Hg, and 1760 (2.6%) as ≥160 mm Hg. Among the participants, 57,831 (83.9%) were aged 65-74 years and 11,070 (16.1%) were aged 75 years or older. The general characteristics of the study participants are presented in Table 1.

Figure 2 shows the Kaplan-Meier curves for the cumulative incidence of all-cause and CVD mortalities according to the
SBP category. The incidence of all-cause and CVD mortalities varied between SBP categories; older adults with SBP 160 mm Hg had the highest mortality rate, while those with SBP within the range of 130-139 mm Hg had the lowest mortality rate.

Table 1. General characteristics of the study participants (N=68,901).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), n (%)</td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>57,831 (83.9)</td>
</tr>
<tr>
<td>≥75</td>
<td>11,070 (16.1)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>30,995 (45)</td>
</tr>
<tr>
<td>Female</td>
<td>37,906 (55)</td>
</tr>
<tr>
<td>BMI, n (%)</td>
<td></td>
</tr>
<tr>
<td>Underweight (&lt;18.5 kg/m²)</td>
<td>2175 (3.2)</td>
</tr>
<tr>
<td>Normal weight (&lt;23 kg/m²)</td>
<td>23,413 (34)</td>
</tr>
<tr>
<td>Overweight (&lt;25 kg/m²)</td>
<td>18,240 (26.5)</td>
</tr>
<tr>
<td>Obese (≥25 kg/m²)</td>
<td>25,073 (36.4)</td>
</tr>
<tr>
<td>Smoking status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Non-smoking</td>
<td>53,231 (77.3)</td>
</tr>
<tr>
<td>Quit smoking</td>
<td>5672 (8.2)</td>
</tr>
<tr>
<td>Current smoking</td>
<td>9998 (14.5)</td>
</tr>
<tr>
<td>Physical activity, n (%)</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>47,038 (68.3)</td>
</tr>
<tr>
<td>1-2 times per week</td>
<td>9324 (13.5)</td>
</tr>
<tr>
<td>3 or more times per week</td>
<td>12,539 (18.2)</td>
</tr>
<tr>
<td>CCI², n (%)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>34,443 (50)</td>
</tr>
<tr>
<td>1-2</td>
<td>20,027 (29.1)</td>
</tr>
<tr>
<td>3 or more</td>
<td>14,431 (20.9)</td>
</tr>
<tr>
<td>Fasting glucose, mean (SD)</td>
<td>103.6 (38.6)</td>
</tr>
<tr>
<td>Total cholesterol, mean (SD)</td>
<td>204.9 (66.3)</td>
</tr>
<tr>
<td>AST³, mean (SD)</td>
<td>27.4 (18.2)</td>
</tr>
<tr>
<td>ALT⁴, mean (SD)</td>
<td>23.3 (19.2)</td>
</tr>
<tr>
<td>Medication, n (%)</td>
<td></td>
</tr>
<tr>
<td>RASB⁵</td>
<td>24,719 (35.9)</td>
</tr>
<tr>
<td>BB⁶</td>
<td>7116 (10.3)</td>
</tr>
<tr>
<td>CCB⁷</td>
<td>35,902 (52.1)</td>
</tr>
<tr>
<td>Diuretics</td>
<td>38,703 (56.2)</td>
</tr>
</tbody>
</table>

aCCI: Charlson Comorbidity Index.
bAST: aspartate transaminase.
cALT: alanine transaminase.
dRASB: renin-angiotensin system blocker.
eBB: β-blocker.
fCCB: dihydropyridine calcium channel blocker.
Figure 2. Kaplan-Meier curves for all-cause and CVD mortality according to SBP. CVD: cardiovascular disease; SBP: systolic blood pressure.

Table 2 shows the results of the Cox proportional hazards models for all-cause and CVD mortalities according to SBP categories. During the follow-up period, 32,588 (47.3%) patients had all-cause mortality and 4273 (6.2%) had CVD mortality. Compared to participants with SBP within the range of 130-139 mm Hg, those with SBP <120 mm Hg had a significantly higher all-cause (hazard ratio [HR] 1.40, 95% CI 1.35-1.46) and CVD mortality (HR 1.42, 95% CI 1.27-1.58) rate. These results were similar in participants with SBP within the range of 140-149 mm Hg, 150-159 mm Hg, and ≥160 mm Hg, in that the participants falling into these SBP categories had significantly higher all-cause and CVD mortality rates compared with those with SBP within the range of 130-139 mm Hg. However, when compared with participants with SBP within the range of 130-139 mm Hg, those with SBP within the range of 120-129 mm Hg did not show a significant difference in either all-cause (HR 1.02, 95% CI 0.99-1.05) or CVD mortality (HR 1.00, 95% CI 0.92-1.09).

Results of the subgroup analysis for the effect of SBP on mortality stratified by age (65-74 years and 75 years or older) are shown in Figure 3. In both age categories, there was a significantly lower mortality rate in the 130-139 mm Hg SBP category than in the other SBP categories, except for the 120-129 mm Hg category. Moreover, those aged 65-74 years showed a higher HR of all-cause and CVD mortalities according to the SBP categories than those aged 75 years or older.

Table 2. Cox proportional hazards models\textsuperscript{a} to predict all-cause and CVD\textsuperscript{b} mortalities by SBP\textsuperscript{c}.

<table>
<thead>
<tr>
<th>Events and SBP categories (mm Hg)</th>
<th>Number of events, n (%)</th>
<th>HR\textsuperscript{d} (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All-cause mortality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;120</td>
<td>3262 (57.6)</td>
<td>1.40 (1.35-1.46)</td>
</tr>
<tr>
<td>120-129</td>
<td>7641 (44.3)</td>
<td>1.02 (0.99-1.05)</td>
</tr>
<tr>
<td>130-139</td>
<td>10,833 (42.9)</td>
<td>1\textsuperscript{e}</td>
</tr>
<tr>
<td>140-149</td>
<td>6862 (48.0)</td>
<td>1.15 (1.12-1.19)</td>
</tr>
<tr>
<td>150-159</td>
<td>2745 (58.9)</td>
<td>1.51 (1.45-1.58)</td>
</tr>
<tr>
<td>≥160</td>
<td>1245 (70.7)</td>
<td>2.08 (1.96-2.21)</td>
</tr>
<tr>
<td><strong>CVD mortality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;120</td>
<td>437 (7.7)</td>
<td>1.42 (1.27-1.58)</td>
</tr>
<tr>
<td>120-129</td>
<td>984 (5.7)</td>
<td>1 (0.92-1.09)</td>
</tr>
<tr>
<td>130-139</td>
<td>1408 (5.6)</td>
<td>1\textsuperscript{e}</td>
</tr>
<tr>
<td>140-149</td>
<td>912 (6.4)</td>
<td>1.17 (1.08-1.28)</td>
</tr>
<tr>
<td>150-159</td>
<td>364 (7.8)</td>
<td>1.53 (1.36-1.71)</td>
</tr>
<tr>
<td>≥160</td>
<td>168 (9.5)</td>
<td>2.11 (1.80-2.48)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Statistical analysis adjusted for age, sex, BMI, smoking status, physical activity, Charlson Comorbidity Index, fasting glucose, total cholesterol, AST, ALT, and blood pressure medication.

\textsuperscript{b}CVD: cardiovascular disease.

\textsuperscript{c}SBP: systolic blood pressure.

\textsuperscript{d}HR: hazard ratio.

\textsuperscript{e}Reference value.
Discussions

Principal Findings

In this retrospective cohort study, the rates of all-cause and CVD mortalities were significantly lower in older adults with hypertension and SBP within the range of 130-139 mm Hg than in the other SBP categories, except for those with SBP within the range of 120-129 mm Hg. These findings suggest the importance of managing SBP within the range of 120-139 mm Hg to lower all-cause and CVD mortality risks in older adults with hypertension. The optimal target SBP level was 120-139 mm Hg in both age groups (65-74 years and 75 years of age or older), and the impact of SBP on mortality was generally higher in older adults aged 65-74 years, indicating that more stringent SBP control is required in these patients than in their older counterparts.

Our results are consistent with the SBP goal of the 2022 KSH guidelines of <140 mm Hg [14]. The 2022 KSH guidelines have limitations in that the target SBP level for older adults with hypertension was established based on the findings of studies conducted on non-Korean participants [15-18]. Additionally, 1 of the main reasons for the unchanged target SBP level of the 2022 KSH guidelines compared to the 2018 KSH guidelines is that there is still insufficient evidence regarding the optimal target SBP level [14]. Our study provides empirical evidence for the optimal SBP level of older adults with hypertension in Korea to reduce all-cause and CVD mortality rates.

This study showed the presence of a J-curved relationship between SBP and CVD risk, indicating that lowering SBP to below 120 mm Hg increased all-cause and CVD mortalities compared with SBP within the range of 130-139 mm Hg. This result was consistent with an international cohort study on 22,672 patients with coronary artery disease [20] and the data from 2 randomized controlled trials (RCTs) on 30,937 patients with high cardiovascular risk [21], which showed that SBP within the range of 120-139 mm Hg had the lowest risk of cardiovascular events, and the risk increased at SBP <120 mm Hg, with a J-curve pattern. Furthermore, 19,110 older adults aged older than or equal to 60 years in the Singapore Chinese Health Study showed that SBP <120 mm Hg tended to increase CVD mortality risk compared with SBP between 120-139 mm Hg, although the risk estimates did not reach statistical significance [22]. However, most guidelines for hypertension management, including the ACC/AHA and KSH guidelines, do not specify the minimal level of SBP. It may be appropriate to suggest the minimum SBP level when scientists and professionals update the hypertension management guidelines. Although several studies on the optimal target SBP level have been conducted, the results remain controversial across countries. The Systolic Blood Pressure Intervention Trial (SPRINT) in the United States showed that an SBP target of <120 mm Hg decreased major adverse cardiovascular events and mortality compared to an SBP target of <140 mm Hg in older adults with hypertension aged 75 years or older [15]. The Strategy of Blood Pressure Intervention in the Elderly Hypertensive Patients trial in China showed that an SBP target of 110-130 mm Hg significantly reduced major adverse cardiovascular events compared to an SBP target of 130-150 mm Hg [16]. The Berlin Initiative Study in Germany showed that SBP of <130 mm Hg is associated with increased all-cause mortality compared to SBP within the range of 140-149 mm Hg [23], whereas another cohort study in England showed that 6311 individuals aged 50-79 years and 1299 individuals aged 80-89 years had no significant associations between SBP levels and all-cause and CVD mortalities [24]. A possible explanation for the different results is ethnic disparities. The prevalence of hypertension and CVD mortality differs across countries [6,12,13]. Asian populations are known to be more vulnerable to hypertension because they tend to have an increased aortic pulse and stiffness than European populations [25], and Asians have a stronger association between blood pressure and CVD than Europeans [25]. Therefore, the target SBP level for hypertension may vary according to race and country.

The estimated risk for all-cause and CVD mortalities increased for every 10 mm Hg increase in SBP above 140 mm Hg, compared to within the 130-139 mm Hg range, regardless of
age group, although the impact of SBP on mortality was slightly higher in individuals aged 65-74 years. However, the relationship between SBP and mortality in older adults remains controversial. A total of 2 RCTs, involving patients aged 75 years and older in the SPRINT [15] and 80 years or older in the Hypertension in the Very Elderly Trial [26], showed that a lower SBP was associated with decreased cardiovascular events. However, several cohort studies suggested that there was no association in older adults 75 years or older [27,28], and some even showed a negative relationship between SBP and cardiovascular events [29,30]. The discrepancies between the RCTs and cohort studies might be that the participants of RCTs were healthier than the general population of the same age because of the restricted inclusion and exclusion criteria [31]. The reason our study revealed an association between increased SBP and higher mortality, unlike other cohort studies, could be attributed to our deliberate selection of healthy older adults who were capable of undergoing health checkups on their own and had no previous history of hypertension or CVD. Therefore, further studies targeting diverse populations (i.e., frail or multimorbid) are needed.

Limitations and Strengths
This study had several limitations. Because we included participants who underwent health checkups between 2003 and 2004, older adults who did not undergo health checkups were excluded. We adjusted for many covariates to alleviate bias, but potential age-related confounding variables, such as frailty, activities of daily living, and gait speed, remained due to data availability. In addition, adverse events of decreased blood pressure, such as renal dysfunction or orthostatic hypotension, were not considered in this study due to data limitations; therefore, we suggest that additional studies exploring optimal blood pressure should include adverse events as 1 of the outcomes. Finally, we used a conventional Cox hazard model with mean SBP measured during the health checkups as an independent variable to facilitate interpretation. However, since SBP is a time-varying variable, a time-dependent Cox model could also have been used. The strengths of this study included the use of the NHIS database, which is nationally representative and includes almost all Koreans, and a large sample size. Our results have sufficient statistical power and can be generalized to older Korean adults with hypertension. In addition, through a 17-year follow-up observation period, our study provided evidence that controlling SBP is crucial for long-term CVD-related health outcomes.

Conclusions
Controlling SBP in older adults with hypertension is important for improving overall and CVD-related health outcomes. However, hypertension guidelines regarding optimal SBP levels differ across countries; therefore, it is necessary to identify the appropriate SBP level among older Korean adults with hypertension. Our retrospective cohort study using national data showed that SBP within the range of 120-139 mm Hg was associated with the lowest risk of CVD and all-cause mortality in older Korean adults with hypertension; therefore, SBP should be controlled within this range in this cohort. Our study also suggested that lowering the SBP to below 120 mm Hg could increase all-cause and CVD mortalities, indicating that caution should be taken not to lower the SBP beyond 120 mm Hg. Additionally, SBP had a greater impact on CVD and all-cause mortality in older adults aged 65-74 years than in those aged 75 years and older; more stringent SBP management is thus required in the 65-74-year age group. This study provides critical knowledge that improves our understanding of the optimal SBP in older Korean adults with hypertension to reduce all-cause and CVD mortality rates. The KSH guidelines only recommend lowering SBP to <140 mm Hg; however, we suggest adding 120 mm Hg as the minimum value.

Acknowledgments
This study was supported by the Basic Science Research Program through the National Research Foundation of Korea, funded by the Ministry of Education (2020R1A6A1A03041989) and the Faculty–Student Research Fund from the College of Nursing, Yonsei University (6-2021-0053).

Authors' Contributions
All authors contributed to the study conceptualization, methodology, and data interpretation. JML was responsible for data analysis and writing of the original draft. KHL was responsible for the review and editing of the paper. All authors have read and approved the final paper.

Conflicts of Interest
None declared.

References


**Abbreviations**

ACC/AHA: American College of Cardiology and the American Heart Association  
CVD: cardiovascular disease  
HR: hazard ratio  
KCD: Korean Classification of Disease  
KSH: Korean Society of Hypertension  
NHIS: National Health Insurance Service  
RCT: randomized controlled trial  
SBP: systolic blood pressure  
SPRINT: Systolic Blood Pressure Intervention Trial

---

**Edited by A Mavragani; submitted 25.08.23; peer-reviewed by G Debacker, C Yang; comments to author 22.12.23; revised version received 21.01.24; accepted 05.05.24; published 11.06.24.**

Please cite as:
Lee JJ, Lee KH  
Optimal Systolic Blood Pressure for the Prevention of All-Cause and Cardiovascular Disease Mortality in Older Adults With Hypertension: Nationwide Population-Based Cohort Study  
JMIR Public Health Surveill 2024;10:e52182  
URL: https://publichealth.jmir.org/2024/1/e52182  
doi: 10.2196/52182  
PMID: 38861307

©Jae Jun Lee, Kyung Hee Lee. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 11.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic
information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Sleep Health Analysis Through Sleep Symptoms in 35,808 Individuals Across Age and Sex Differences: Comparative Symptom Network Study

Christophe Gauld¹*, MD, PhD; Sarah Hartley²,³*, MD; Jean-Arthur Micoulaud-Franchi⁴,⁵, MD, PhD, HDR; Sylvie Royant-Parola³, MD

¹Hospices Civils de Lyon, Lyon, France
²Sleep Center, APHP Hôpital Raymond Poincaré, Université de Versailles Saint-Quentin en Yvelines, Garches, France
³Réseau Morphée, Garches, France
⁴Services of Functional Exploration of the Nervous System, University Sleep Clinic, University Hospital of Bordeaux, Bordeaux, France
⁵Unité Sommeil, Addiction, Neuropsychiatrie, Centre national de la recherche scientifique Unité Mixte de Recherche–6033, Bordeaux, France

*these authors contributed equally

Corresponding Author:
Christophe Gauld, MD, PhD
Hospices Civils de Lyon
59 Bd Pinel
Lyon, 69000
France
Phone: 33 671675095
Email: gauldchristophe@gmail.com

Abstract

Background: Sleep health is a multidimensional construct that includes objective and subjective parameters and is influenced by individual sleep-related behaviors and sleep disorders. Symptom network analysis allows modeling of the interactions between variables, enabling both the visualization of relationships between different factors and the identification of the strength of those relationships. Given the known influence of sex and age on sleep health, network analysis can help explore sets of mutually interacting symptoms relative to these demographic variables.

Objective: This study aimed to study the centrality of symptoms and compare age and sex differences regarding sleep health using a symptom network approach in a large French population that feels concerned about their sleep.

Methods: Data were extracted from a questionnaire provided by the Réseau Morphée health network. A network analysis was conducted on 39 clinical variables related to sleep disorders and sleep health. After network estimation, statistical analyses consisted of calculating inferences of centrality, robustness (ie, testifying to a sufficient effect size), predictability, and network comparison. Sleep clinical variable centralities within the networks were analyzed by both sex and age using 4 age groups (18-30, 31-45, 46-55, and >55 years), and local symptom-by-symptom correlations determined.

Results: Data of 35,808 participants were obtained. The mean age was 42.7 (SD 15.7) years, and 24,964 (69.7%) were women. Overall, there were no significant differences in the structure of the symptom networks between sexes or age groups. The most central symptoms across all groups were nonrestorative sleep and excessive daytime sleepiness. In the youngest group, additional central symptoms were chronic circadian misalignment and chronic sleep deprivation (related to sleep behaviors), particularly among women. In the oldest group, leg sensory discomfort and breath abnormality complaint were among the top 4 central symptoms. Symptoms of sleep disorders thus became more central with age than sleep behaviors. The high predictability of central nodes in one of the networks underlined its importance in influencing other nodes.

Conclusions: The absence of structural difference between networks is an important finding, given the known differences in sleep between sexes and across age groups. These similarities suggest comparable interactions between clinical sleep variables across sexes and age groups and highlight the implication of common sleep and wake neural circuits and circadian rhythms in understanding sleep health. More precisely, nonrestorative sleep and excessive daytime sleepiness are central symptoms in all groups. The behavioral component is particularly central in young people and women. Sleep-related respiratory and motor...
symptoms are prominent in older people. These results underscore the importance of comprehensive sleep promotion and screening strategies tailored to sex and age to impact sleep health.

(Keywords: symptom; epidemiology; age; sex; diagnosis; network approach; sleep; sleep health)

Introduction

Sleep health is a multidimensional construct that includes objective and subjective parameters and is influenced by individual sleep-related behaviors [1] and sleep disorders, measurable by a set of techniques ranging from psychometry [2,3] to mobile devices [4,5]. Sleep health is not merely the absence of sleep disorder symptoms; it also includes factors such as sleep disturbance and satisfaction with sleep [6,7]. Moreover, sleep health considers the importance of sleep behaviors reinforcing both the homeostatic (duration of sleep) and the circadian (regularity and time of sleep) system that underpin sleep quality [8]. Good-quality sleep is important for overall health [9], but sleep is influenced both by age and by sex [10-12].

Sleep architecture evolves from birth to old age, with changes in the macro- and microstructure that are underpinned by neuroanatomical changes in both homeostatic and circadian systems [13]. More precisely, sleep evolves across the lifespan, with the development of a 24-hour circadian rhythm, a progressive reduction in daytime naps, and a gradual reduction in the quantity of slow-wave sleep (N3) in the early years. From the fifth decade onward, advanced sleep timing, longer sleep onset latency, shorter nighttime sleep potentially compensated by daytime naps, increased wake after sleep onset (WASO), increased fragmentation, a lower threshold for wake from sleep due to external stimuli, reduced N3, and consequently increased light sleep (N1 and N2) from the fifth decade are seen [11,12,14-22]. These later changes are accompanied by changes in the microstructure of sleep with a loss of the homeostatic drive expressed as reductions in slow-wave amplitude and density [23] and modifications in spindles and K-complexes [24,25]. Neuronal loss with age underpins changes in both homeostatic and circadian systems. Changes in spindles are potentially related not only to a loss of the hippocampal gray matter [26] but also to a loss of the functional integrity of corticothalamic loops [27]. A loss of suprachiasmatic nucleus (SCN) neurons, notably vasofactorial intestinal peptide (VIP)-expressing ones, may contribute to changes in the circadian rhythm [28].

Sex-determined differences in the sleep macro- and microstructure are detectable from an early age, with differences in sleep disorders increasing from early puberty onward [29]. Certain sleep disorders are more frequent in men (eg, obstructive sleep apnea syndrome) [30], while others are more frequent in women (eg, insomnia disorder) [31], and these differences modulate over the lifespan. Concerning the impact of sex according to age, biological sex markedly affects sleep, and differences are detectable in newborns, with increased divergence at puberty [29]. Changes in sleep with age are modulated by sex: the marked reduction in slow-wave sleep in men over 70 years does not affect women, although reductions in rapid eye movement (REM) sleep are seen in both [12]. Numerous sex-related differences are seen at an anatomical level: larger male-specific declines are seen in galanin-expressing neurons, orexin neurons, neuromelanin-sensitive neurons in the locus coeruleus (LC), and VIP-expressing neurons in the SCN. Women report subjective changes in sleep related to menopause, with increased insomnia and reduced sleep quality [32], although this is difficult to demonstrate objectively via polysomnography [33]. Biological sex also affects sleep disorders: obstructive sleep apnea is more common in men [30] and insomnia in women [31]. In addition, sleep behaviors may vary between men and women: as circadian timing is phase-advanced in women [34], women have earlier bedtimes [35] and an earlier chronotype, although this reverses after the age of 40 years [36].

The multifactorial components contributing to sleep health are adapted to network analysis for several reasons. First, unlike a certain number of other models limiting the number of variables analyzed (eg, due to an increase in the type 1 error), symptom network analysis offers an original way to model the interactions between numerous sleep-related clinical variables according to other sociodemographic ones. In this way, it offers a more holistic view of sleep health. Second, traditional studies often examine symptoms individually, looking at their interconnected impacts. This computational network analysis not only allows visualization of the relationships between these variables but also helps identify the strength of those mutual relationships. The analysis of these interactions seems particularly important in the context of sleep health, whose mutual influencing factors have been largely undetermined until now. Third, identifying the most influential symptoms across varying age and sex demographics, by considering a large number of variables and the interactions between them, provides a clearer roadmap for health care and public health interventions. Finally, symptom network analysis has the potential for sleep health promotion at both a local and a global scale by identifying screening strategies that are sensitive to the nuanced differences across populations.

In our previous work using a data-driven network analysis approach in a large data set including sleep disorders, sleep disturbances, sleep-related behaviors, and comorbidities as variables, the most central variables were a complaint of nonrestorative sleep, excessive daytime sleepiness, circadian irregularity, and chronic sleep deprivation [37]. The latter 2 variables are linked to the circadian system and the homeostatic system, respectively, whereas the complaint of nonrestorative sleep, or not feeling refreshed on waking, is a relatively little studied concept compared to its potential importance for patients...
Finally, the complaint of daytime sleepiness can be considered a consequence of poor sleep, and as such, its centrality is unsurprising, given its importance for sleep physiology and for its historical and methodological centrality in the construction of sleep classifications and its relevance in clinical practice [40-43].

Given the known influence of sex and age on sleep health, extending our analysis to consider these variables is highly pertinent as network analysis explores sets of mutually interacting symptoms. It can thus detect whether the differences between groups (eg, of sex or age) may or may not be explained by some specific symptoms. Thus, although group comparison based on frequency statistics only provides an index of difference (eg, a $P$ value), network analysis provides insight into which factors drive differences.

The objective of this work was to study the centrality of symptoms and to compare age and sex differences in the experience of sleep using a cross-sectional approach of comparative network analysis in a population that feels concerned by their sleep. First, we aimed to analyze the relationships between the components of these sleep networks between 2 populations, men, and women. Second, we studied these relationships according to different age classes. For these 2 analyses (sex and age), we studied the potential differences in centrality (ie, the importance), predictability, and connections between different factors related to sleep symptoms, sleep behavior, and the effects of sleep.

Methods

Data Collection

The Réseau Morphée is a nonprofit organization funded by the French government (regional health authority), aiming to improve patient care pathways in sleep health. In 2017, the new patient care pathway was launched, offering a comprehensive online questionnaire as the initial port of entry. The questionnaire looks at 5 domains: sociodemographic data (age and sex), symptoms of sleep disorders, sleep disturbances, sleep behaviors, and comorbidities [40]. In total, 39 variables were analyzed. We prioritized data extraction based on clinical relevance, guaranteeing it aligned with both clinical and public health relevance (and reinforcing by analysis of robustness). We also provided an example of the analyses by adding a variable (the BMI) in order to discuss such an addition and its implications. Data from 3 validated scales were included: the Epworth Sleepiness Scale (ESS; excessive daytime sleepiness score<10) [44], the Insomnia Severity Index (ISI; clinical insomnia score>14) [45], and the Hospital Anxiety and Depression Scale for anxiety (HADS-A) and depression (HADS-D) (scores>10 for each) [46].

Ethical Considerations

Patients were informed about the use of their data for epidemiological research when data were originally collected. Informed consent was obtained from all participants included in the study. The study was approved by the scientific committee of the Réseau Morphée and by 2 patient associations, the Sommeil et Santé and the France Insomnie. As a noninterventional study (MR004), this research was approved by the Commission Nationale Informatique et Liberté (CNIL; approval number: 8013081; date of approval: December 19, 2016). Only data from participants over the age of 18 years were included in the analysis. The anonymization of data provides confidentiality and total respect for the privacy of participants. No publicity was used to attract participation. No financial compensation was provided for participating in the study.

Participants

Participants were French-speaking members of the general population who filled in an online questionnaire aimed at people concerned about their sleep. This population does not present people who already experience sleep problems but rather people who are sufficiently concerned about their sleep to use the Réseau Morphée online questionnaire. They were not in contact with sleep services and thus did not have a formal diagnosis of their sleep problems. We included all respondents over 18 years old from March 1, 2017, to September 9, 2020. All participants in our study speak French, but all may not be currently residing in France. Regarding demographic data, we collected the sex and age of each participant. As specified later in the Limitations section, we considered the person’s sex as assigned at birth, not gender as defined by the person, and we chose to use the terms “men” and “women.” The entire sample was divided into 4 age groups: 18-30, 31-45 years, 46-55, and >55 years. Age group stratification was determined by known sociodemographic boundaries in France; the 18-30-year group is either still in education or in the early stages of their careers and less likely to have children (the mean childbearing age in France was 30.9 years in 2021 according to the [47]). Participants aged 31-45 years are settled in their careers and may have children. The 46-55-year group was chosen specifically to isolate the period of the menopause, during which sleep complaints in women markedly increase [32]. We provided the number of men and women (sex), the mean (SD) of age, and the BMI for the ESS, the ISI, the HADS-A, and the HADS-D, and we detailed these measurements by age group.

Network Estimation, Inferences, Robustness, and Comparison

The main aim of this network analysis was to analyze a set of 39 variables described by a large population divided into several subpopulations by age and sex to compare the subpopulations. To this end, 6 networks were constructed for comparison: a men’s network, a women’s network, and a network for each age group (18-30, 31-45, 46-55, and >55 years).

In this methodology, the correlations between items were understood as a network in which each item represents a node and where correlations between symptoms are connections (or edges) between nodes. A connection between any 2 symptoms was said to exist if those symptoms were present in the same individual. Symptoms of 2 different disorders were connected if the symptoms were present in different individuals. We followed the network guidelines for computational analysis of network properties [48].
Network Estimation

Network estimation was conducted with the Ising Model, which is, to date, the state-of-the-art network model used in psychopathology research for binary data (see Multimedia Appendix 1 [49-69] for details on symptom networks) associated with a regularization technique called graphical least absolute shrinkage and selection operator (LASSO) [59,60]. The final model was chosen by using the extended Bayesian information criterion (EBIC) [61].

The network was graphically represented according to the Fruchterman-Reingold algorithm, where variables with stronger or more connections are placed closer to each other [62]. Nodal predictability was calculated based on models derived from mixed graphical models (MGMs) [63,70] and graphically represented as a pie chart in the ring around each variable. Predictability has a value between 0 and 1, provided on the basis of the variance of the prediction error calculated according to the $R^2$ of the MGM. It refers to how well a given node in the network can be predicted by all remaining nodes. It thus shows how relevant edges are (eg, a node may be connected to many other nodes, but if these only explain only 1% of its variance, it is unlikely that this node is relevant). This has further implications: for example, designing an intervention to affect certain nodes or detecting where data are lacking (eg, when parts of the network are little influenced by related nodes and thus must depend on external factors). In clinical practice, the predictability of a symptom indicates “whether an intervention on that symptom through the symptom network is promising” [70]. The predictability of sexes and age groups is given in Multimedia Appendix 2.

In agreement with previous work [37] and the main international sleep-wake classifications [71,72], as well as to facilitate the reading of the graphic representations, we classified elements from the questionnaire into 8 groups: hypersomnia symptoms, insomnia symptoms, respiratory symptoms, motor symptoms, psychiatric symptoms, sleep disturbance, sleep behaviors, and comorbidities. These groupings had no impact on the main results of the study and were only given for data visualization purposes.

Network Inferences

The most clinically relevant network inferences correspond to centrality measures [64]. Centrality measures are important for identifying symptoms that play a crucial role in connecting all the symptoms of a network [65]. Nodes with high centrality index measures represent variables that are highly connected to other variables. Precisely, 4 measures are classically described: strength, closeness, betweenness, and expected influence [66]. The strength of a node computes the degree to which it is connected with all the other nodes of the network [67]. The 3 other centrality measures are described in Multimedia Appendix 1. The centrality strengths of age groups and sexes were plotted and are detailed in Multimedia Appendix 2. The color (red=0, green=1) and bar plot of each symptom provide visual insight into the $P$ value. Symptoms were sorted in descending order of strength and predictability.

Network Comparisons

The networks were systematically compared between men and women and between the 4 age groups (18-30, 31-45, 46-55, and >55 years). Cross-age and cross-sex networks were compared between each other with a network comparison tool, providing significance indices for the overall structures of the networks and allowing us to evaluate the differences between each network.

We used the network comparison test (NCT) [73], a 2-tailed permutation test that examines differences between 2 networks. This test analyzes the difference in terms of relationships between (1) variables connections (“Are the connections between variables different between the 2 networks?”) and (2) centrality measures, named global strength (“Is the centrality of the variables different between the 2 networks?”). The NCT therefore presented 2 results: a comparison of connections between variables and a comparison in terms of centrality. $P<.05$ was considered a statistically significant difference.

There is a potential risk of loss of power when the samples are not of equal size [61]. To deal with our unequal sample, we subsampled the larger data set to the size of the smaller data set 5 times. This meant we ran the NCT 5 times, with 2000 replications for each comparison [74]. We presented the average of the results without adjusting the $P$ values for multiple testing.

Moreover, to offer inferences on the comparison between networks at the level of each symptom, we computed the local symptom-by-symptom correlations between the nodes of the 2 networks, with a $P$ value resulting from the permutation test concerning differences both in global strength and in edges weights. Local symptom-by-symptom correlations ($P$ value based on permutations) between the strength by sex and by age are given in Multimedia Appendix 3. Local symptom-by-symptom correlations ($P$ value based on permutations) between connections by sex and by age are also given in Multimedia Appendix 3. The color (red=0, green=1) and bar plot of each symptom provide visual insight into the $P$ value. The symptoms were sorted according to the order of the figures in the paper. However, we only discussed the most relevant relationships rather than variables systematically significantly associated, which should only be interpreted according to all the other relationships of the network.

Network Robustness

To verify that the number of participants was adequate to perform such a network analysis, we studied the robustness of the network using bootstrap analysis (N=2000 iterations) [68,69]. The robustness of the results was tested on a centrality measure, the strength (ie, we checked the stability of edges and centrality measures). We used a case-dropping subset to assess the stability of this centrality measure, providing a centrality stability correlation coefficient (ie, correlation stability [CS] coefficient; we computed how well the order of centralities was retained after observing only a subset of the data): the CS coefficient (ie, the maximum proportion of participants that could be dropped, while maintaining 95% probability that the correlation between centrality metrics from the full data set and...
the subset data was at least 0.70). Based on a simulation study [68,69], a minimum CS coefficient of 0.25 was recommended.

All analyses and graphical visualizations were performed in R version 4.2.2 (R Foundation for Statistical Computing). Generative artificial intelligence (AI) was not used at any point.

**Results**

**Participant Details**

Table 1 provides details of the study sample. There were twice as many women (n=24,964, 69.7%) as men (n=10,844, 30.3%).

Table 1. Description of the French-speaking adult population (N=35,808) participating in this study.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All participants</th>
<th>Age groups (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18-30 (n=9051)</td>
<td>31-45 (n=12,129)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>10,844 (30.3)</td>
<td>2259 (25.0)</td>
</tr>
<tr>
<td>Women</td>
<td>24,964 (69.7)</td>
<td>6792 (75.0)</td>
</tr>
<tr>
<td>Age (years), mean (SD, 95% CI)</td>
<td>42.7 (15.7, 42.75-42.92)</td>
<td>-</td>
</tr>
<tr>
<td>BMI, mean (SD, 95% CI)</td>
<td>24.7 (5.12, 19.43-27.21)</td>
<td>24.44</td>
</tr>
<tr>
<td>ESS&lt;sup&gt;b&lt;/sup&gt;, mean (SD, 95% CI)</td>
<td>9.47 (4.95, 9.47-9.53)</td>
<td>9.81</td>
</tr>
<tr>
<td>ISI, mean (SD, 95% CI)</td>
<td>16.34 (5.24, 16.34-16.39)</td>
<td>15.69</td>
</tr>
<tr>
<td>HADS-A&lt;sup&gt;d&lt;/sup&gt;, mean (SD, 95% CI)</td>
<td>6.56 (3.97, 3.95-9.62)</td>
<td>10.01</td>
</tr>
<tr>
<td>HADS-D&lt;sup&gt;e&lt;/sup&gt;, mean (SD, 95% CI)</td>
<td>9.58 (3.97, 9.58-9.62)</td>
<td>6.37</td>
</tr>
</tbody>
</table>

<sup>a</sup>Not applicable.
<sup>b</sup>ESS: Epworth Sleepiness Scale.
<sup>c</sup>ISI: Insomnia Severity Index.
<sup>d</sup>HADS-A: Hospital Anxiety and Depression Scale for anxiety.
<sup>e</sup>HADS-D: Hospital Anxiety and Depression Scale for depression.

Sleep Health Network Analysis

**Network Estimation**

Each network contained 39 nodes (items) and 741 edges (connections), with a mean weight of 0.011 (SD 0.005). Figure 1 presents results of the sleep health network analysis related to sex groups, with negative edges. The predictability of sexes and age groups is given in Multimedia Appendix 2.
Figure 1. Sleep networks related to sex groups of a French-speaking adult population concerned about their sleep (N=35,808). (A) Women’s network. (B) Men’s network. Blue edges (connections) represent the positive associations between variables and orange edges the negative associations. The thickness of the line represents the level of correlation between 2 variables. The predictability of the nodes is depicted as a pie chart in the rings around the nodes: the area in the outer ring of a node represents the percentage of variance of the node that is explained by all neighboring nodes. Color groupings are only given for data visualization purposes. A higher-resolution version of this image is available in Multimedia Appendix 4.

Results of the sleep health network analysis related to age groups are shown in Figure 2. In the figure, the 2-to-2 relationships, conditional on all other relationships between nodes in the network, can be visualized.
Figure 2. Sleep networks related to the 4 age groups of a French-speaking adult population concerned about their sleep (N=35,808). (A) 18-30 years, (B) 31-45 years, (C) 46-55 years, and (D) >55 years. Blue edges (connections) represent the positive associations between variables and orange edges the negative associations. The thickness of the edges represents the level of correlation between 2 variables. The predictability of the nodes is depicted as a pie chart in the rings around the nodes: the area in the outer ring of a node represents the percentage of variance of the node that is explained by all neighboring nodes. Color groupings are only given for data visualization purposes. A higher-resolution version of this image is available in Multimedia Appendix 4.

We first deduced from the graphical presentation which node-by-node correlations were the most relevant. In the women’s network, it was interesting to see that nonrestorative sleep was strongly negatively associated with early insomnia ($r=-0.2$). Similarly, in the men’s network, excessive daytime sleepiness was strongly negatively correlated with wakefulness satisfaction ($r=-0.26$) but positively correlated with falling asleep while driving ($r=0.23$) or with naps ($r=0.2$). Across the different age group networks, a number of relationships were consistently found, for example, between the symptoms nonrestorative sleep and awakening difficulties ($r=0.39$, $0.39$, $0.36$, and $0.38$ in the 18-30–, 31-45–, 46-55–, and >55-year age groups, respectively, in ascending order), as well as weak relationships between breathing abnormalities observation and nighttime leg pain ($r=0.12$, $0.08$, $0.01$, and $0.1$ in the 18-30–, 31-45–, 46-55–, and >55-year age groups, respectively, in ascending order).

Predictability results (ie, if a given node could be predicted by adjacent nodes) were interesting when identifying relevant relationships. For instance, in the women’s network, the 6 most central nodes had the following centrality values: nonrestorative sleep, 2.50; excessive daytime sleepiness, 1.70; chronic sleep deprivation, 1.53; chronic circadian misalignment, 1.40; leg sensory discomfort, 1.08; and wakefulness satisfaction, 0.87, as shown in Figure 3. These symptoms were also among the most predictable: predictability values were 0.54, 0.40, 0.68, 0.36, 0.33, and 0.39, respectively (mean 0.26, SD 0.15). All the predictability measures (variance) are given in Multimedia Appendix 2. However, in the women’s network, for nearly identical centrality, chronic sleep deprivation had greater predictability (0.68) than chronic circadian misalignment (0.36), roughly two-thirds to one-third. This result illustrates the need to consider the first of these nodes as being highly predictable but also central to modifying other nodes.
Network Centrality

In terms of the local structure, 2 symptoms were found among the 4 most central ones, regarding all the centrality measures of all the networks (for all age and sex groups, see Figure S1 in Multimedia Appendix 1): nonrestorative sleep (belonging to the “sleep disturbances” type) and excessive daytime sleepiness (belonging to the “symptoms of sleep disorders” type). The value of the strength for each measure is given in Multimedia Appendix 2.

This means that these variables exhibited a high degree of connection in the entire sleep network (strength), a high degree of connection with proximal variables of the sleep network after considering negative correlation values (expected influence), the shortest path between 2 other variables (betweenness), and the shortest mean distance from other variables (closeness).

The differences in centrality for the 2 sexes are given in Figure 3. The 3 most important centrality measures for the women’s network were, in descending order, nonrestorative sleep, excessive daytime sleepiness, and chronic sleep deprivation; for the men’s network, these were nonrestorative sleep, chronic circadian misalignment, and excessive daytime sleepiness.

Figure 4 shows the strength (centrality measure) of the 4 age groups in the sleep health network analysis. The 3 most important centrality measures for the 18-30–year network were, in descending order, nonrestorative sleep, chronic circadian misalignment, and excessive daytime sleepiness; for the 31-45–year network, these were nonrestorative sleep, excessive daytime sleepiness, and leg sensory discomfort; for the 46-55–year network, these were nonrestorative sleep, leg sensory discomfort, and excessive daytime sleepiness; and for the >55-year network, these were excessive daytime sleepiness, nonrestorative sleep, and leg sensory discomfort.
Nonrestorative sleep and excessive daytime sleepiness were among the 3 most central symptoms in all 4 groups (strengths were 1.73-2.21 and 1.35-2.19, respectively). However, in the 18-30-year group, the 2 other most central symptoms were chronic circadian misalignment (strength=1.42) and chronic sleep deprivation (strength=1.27), particularly in the women’s network, belonging to the “sleep behaviors” type. Between the ages of 31 and 55 years, the 2 other most central symptoms were chronic circadian misalignment (strength=1.30 for the 31-45-year group and strength=1.32 for the 46-55-year group) and leg sensory discomfort (strength=1.54 for the 31-45-year group and strength=1.78 for the 46-55-year group, belonging to the “motor symptoms” type. In the older group of participants (>55 years old), leg sensory discomfort (strength=1.48) remained 1 of the most central symptoms, while breath abnormality complaint became among the 4 most central symptoms (strength=1.13, belonging to the “respiratory symptoms” type), becoming more central than sleep behaviors found in the youngest. In Multimedia Appendix 4, we provide and discuss the example of a network with its centrality, including the BMI. Finally, Figure 5 shows the differences in centrality between the ages according to sexes.
Figure 5. Centrality measure (strength) of the 4 age groups relative to the women’s and men’s networks, distinguished by sleep health network analysis: the 4 age groups for (A) women and (B) men. At the top of the 4 tables, nonrestorative sleep has the highest centrality for the 31-45–year group in men. At the bottom, diabetes disease has the lowest centrality for the 18-31–year group in both women and men. A higher-resolution version of this image is available in Multimedia Appendix 4.

Network Comparisons

Here, we present the 2 results of the NCT, comparing (1) the connections between variables and (2) the results concerning centrality (ie, global strength). All the comparisons between variable connections for the 2 networks (men and women) were significantly different ($M$ test: 0.100, $P<.001$). However, there were no differences in terms of centrality (global strength of the women’s network=15.27, global strength of the men’s network=13.34; $t$ test: 1.92, $P=.68$).

Table 2. Difference in connections between age groups ($M$ test and $P$ value) of the French-speaking adult population concerned about their sleep (N=35,808).

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>18-30</th>
<th>31-45</th>
<th>46-55</th>
<th>&gt;55</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$ test</td>
<td>0.057</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46-55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$ test</td>
<td>0.087</td>
<td>0.061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$ test</td>
<td>0.159</td>
<td>0.122</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>$P$ value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>

$a$Not applicable.

In the same way, as shown in Table 2, all the comparisons between variable connections for the 4 age networks were significantly different. However, as shown in Table 3, no significant difference in terms of centrality measures was found. Multimedia Appendix 3 presents the local symptom-by-symptom correlations ($P$ value based on permutations) between connections and strength on sexes and all age groups.
Table 3. Difference in centrality (global strength) between age groups of the French-speaking adult population concerned about their sleep (N=35,808).

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>18-30</th>
<th>31-45</th>
<th>46-55</th>
<th>&gt;55</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>18-30</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strength</td>
<td>12.78 vs 12.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S test</td>
<td>0.569</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>31-45</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strength</td>
<td>12.76 vs 11.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S test</td>
<td>1.213</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>46-55</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strength</td>
<td>12.76 vs 11.28</td>
<td>12.19 vs 11.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S test</td>
<td>0.477</td>
<td>0.643</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>.89</td>
<td>.92</td>
<td></td>
<td>.41</td>
</tr>
<tr>
<td><strong>&gt;55</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strength</td>
<td>12.28 vs 12.28</td>
<td>12.28 vs 11.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S test</td>
<td>0.735</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>.99</td>
<td></td>
<td></td>
<td>.41</td>
</tr>
</tbody>
</table>

aNot applicable.
bStrength in the networks in columns versus strength in the networks in rows.

Network Robustness

The robustness of strength is shown in Figure S2 in Multimedia Appendix 1. Robustness was judged acceptable, with a CS coefficient for strength at 0.91 for the women’s network; at 0.75 for the men’s network, the 18-30- and 31-45-year networks; at 0.672 for the 46-55-year network; and at 0.75 for the >55-year network.

Discussion

Principal Findings

Our study built on initial work describing key variables in people’s experience of their sleep complaints, in a population that was sufficiently concerned about their sleep to use the Réseau Morphée online sleep questionnaire. We showed that in network analysis on clusters defined by sex or age, the symptoms nonrestorative sleep and excessive daytime sleepiness and an element of sleep behavior (circadian irregularity) remained central to all groups. There were no statistical differences in these elements between the groups, implying that the underlying processes that drive the network and the underlying symptom structure are the same in men and women and the same across different age groups. However, differences in internodal connections between the clusters mean that sleep complaints are not connected in the same way in each of them. This difference should be interpreted as a difference in terms of the structure and the hierarchy in interactions between symptoms. For instance, in the women’s network, the symptoms breath abnormality observation, breath abnormality complaint, and snoring are much less related than they are in the men’s network (see Figure 1).

Comparison With Prior Work

The elements identified as central to the network for both men and women across all ages are best understood as the impact of underlying processes on the experience of sleep. As discussed in the Introduction section, the basic neuroanatomy underpinning sleep is identical in men and women, and although sex-related differences exist and functions may diminish with age, the structure remains. This leads to an important conclusion about improving sleep at a population level. For interventions to be effective in men and women and across all age groups, it is necessary to target central elements with high predictability: high predictability means that interventions aimed at these variables have a high chance of influencing the variables around them [70]. Fortunately, all 3 central variables showed relatively high predictability: interventions aimed at the key sleep-related behaviors have thus a high chance of working in men and women of all ages and of positively influencing other elements of the network.

Excessive daytime sleepiness is due to hypofunction of the wake system. This may be due to neuroanatomical modifications (eg, neuronal loss in the elderly or in patients with narcolepsy) or due to the effect of sleep fragmentation (eg, in obstructive sleep apnea) with incomplete recovery of the homeostatic system. Excessive daytime sleepiness has been shown to be central in both the International Classification of Sleep Disorders, Third Edition (ICSD-3) symptom network [40] and in the sleep-wake disorders symptom network of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [41]. Although sensitivity to sleep need may decrease with age and excessive daytime sleepiness is more common in women [75], the centrality of excessive daytime sleepiness as a consequence of impaired sleep or the wake system or both remains across all clusters.

Nonrestorative sleep, defined as a “subjective feeling of being unrefreshed upon awakening,” is considered a result of poor-quality or unrestful sleep. This symptom is poorly understood and has been little researched [6]. Ohayon [38]
reported that nonrestorative sleep was present in 10.8% of a European sample and 11.4% of French-speaking respondents. Nonrestorative sleep is found to be more common in women and younger subjects, increased in patients with psychiatric disorders, and much more frequent in people dissatisfied by their sleep and those with difficulty getting started in the morning. It is also associated with excessive daytime sleepiness, which would confirm our findings related to the centrality of nonrestorative sleep and excessive daytime sleepiness. However, the relationship between these 2 elements differs with sex: there is a direct and positive link between the 2 elements in women, whereas in men, the 2 nodes are more distant from each other and less strongly linked. The link also varies with age, with excessive daytime sleepiness and nonrestorative sleep being closer together and more strongly linked in middle age compared to young age (18-30 years) and older persons (>55 years). It is tempting to see nonrestorative sleep as an element related to the homeostatic system, with a residual sleep need at wakening due to either sleep disorders or circadian misalignment. The maintenance of its centrality across the different clusters is interesting. It clearly underpins the experience of sleep, and further research on its nature and correlates is overdue.

Finally, a behavior-related variable, circadian irregularity, retained its centrality across all clusters. Circadian irregularity relates to irregular sleep timing, with the risk that the actual timing of sleep is out of phase with the circadian rhythm. This is termed “circadian misalignment” and has been recently identified as 1 of the early indicators of cardiometabolic and neurological disorders. It is interesting to note that the central common features reunite physiological elements that underpin sleep: the homeostatic system, behavior affecting the circadian system, and the consequences of poor sleep. However, the centrality of circadian irregularity is particularly important as it is a clear target for behavioral interventions involving sleep: programs encouraging regular sleep are likely to positively influence the experience of sleep across the population.

After the homogeneity of the 3 central variables, the clusters differed in the subsequent variables: sleep-related respiratory and motor symptoms are more frequent in the older age groups. Sleep disorders, notably sleep apnea, are known to increase with age [30]. In the younger age group, sleep-related behaviors are more central, with chronic sleep deprivation being particularly important in young women.

Although no overall differences were detected for centrality, there were clear differences in the connectedness of the networks, which implies that the links between the different nodes are different: that is, although the structure of the network may be the same, the strength of the links between the different nodes is influenced by age and sex. This is seen in the relationship between excessive daytime sleepiness and nonrestorative sleep, where the relationship is closer and more direct in middle age compared to young age or those >55 years old (eg, the $p$ value was highly significant between the 18-30– and >55-year groups; $P<.001$).

Global strength increased in women and also in the 45-55-year group. The increase in global strength relates to the fact that all the variables are more closely related. This indirectly implies increased connectedness in the network: sleep symptoms, sleep behaviors, and sleep disorders are all more closely interlinked in women than in men or in older or younger age groups.

**Limitations**

Our study has clear limitations. First, we did not consider gender but rather considered sex as assigned at birth, which, despite the lack of consideration for the choice of the person involved, can improve epidemiological comparability. Second, our population comprises French-speaking participants who were concerned about their sleep and is thus not representative of the general population. Despite the large size of our sample, we should therefore be cautious about the generalizability of these results to the general population. Due to anonymity regarding certain sociodemographic data, we cannot provide the location of the participants’ primary residence. The fact that the average ISI, HADS-A, and HADS-D scores exceeded their respective thresholds indicates that on average, participants in our study displayed elevated scores for insomnia, anxiety, and depression. These characteristics may limit the representativeness (eg, whether participants were selected or self-selected based on certain criteria) or overall interpretation of the results (eg, the need to create subgroups). Moreover, an increase in nonrestorative sleep was noted in patients dissatisfied with their sleep in a study by Ohayon [76], and the interpretation of our results should bear this in mind. Third, participants responded to an online questionnaire with a structured symptom and sleep behavior list rather than a validated questionnaire: although this limited comparisons with other studies, it did, however, permit the inclusion of items such as screen use, which are often lacking from older questionnaires [41]. Finally, network analysis is a new and powerful way of analyzing large data sets, but interpretation of results and, notably, centrality should be performed with care. Network analysis highlights the dense interaction between variables and their interdependence at a population level but not at an individual level.

**Future Research Direction**

The results of this study provided 3 main perspectives. First, at the clinical level, this research advances our understanding of sleep variations by age and sex, centering on symptoms rather than diagnoses. Second, at the methodological level, it explores an area often underrepresented in the literature but essential for clinical practitioners, providing a weighted visualization of symptom relationships of sleep health, which is important when considering treatments such as cognitive-behavioral therapies. Finally, the study helps promote public health strategies, emphasizing the importance of the underlying unity of sleep health, while promoting a tailored and precise strategy for specific subgroups.

**Conclusion**

To conclude, this symptom network analysis on sleep health across various age and sex demographics demonstrates the stability of sleep health across sex and age groups, while highlighting the centrality of nonrestorative sleep and excessive daytime sleepiness. More precisely, younger individuals, especially women, show a higher centrality of behavioral-related sleep issues, whereas the older group has more sleep-related
respiratory and motor symptoms. These findings emphasize the need for both general and tailored sleep promotion and screening strategies based on age and sex to effectively address and enhance sleep health in the broader population. Furthermore, it paves the way for health care professionals to adopt a more individualized approach in assessing and treating sleep disorders, thereby enriching the overall quality of patient care.

Data Availability
Data can be provided upon request from the corresponding author.

Authors' Contributions
All authors contributed to the study conception and design. Material preparation and data collection were performed by SH and SRP. Methodology, conceptualization, editing, supervision, and reviewing were driven by JAMF, SH, and SRP. The first draft of the manuscript was written by SH and CG. Analyses were performed by CG. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary material 1.
[DOCX File , 264 KB - publichealth_v10i1e51585_app1.docx ]

Multimedia Appendix 2
Supplementary material 2.
[DOCX File , 1849 KB - publichealth_v10i1e51585_app2.docx ]

Multimedia Appendix 3
Supplementary material 3.
[DOCX File , 713 KB - publichealth_v10i1e51585_app3.docx ]

Multimedia Appendix 4
Higher-resolution images of Figures 1-5.
[DOCX File , 4653 KB - publichealth_v10i1e51585_app4.docx ]

Multimedia Appendix 5
Supplementary material 4.
[DOCX File , 612 KB - publichealth_v10i1e51585_app5.docx ]

References


**Abbreviations**

CS: correlation stability  
ESS: Epworth Sleepiness Scale  
HADS-A: Hospital Anxiety and Depression Scale for anxiety  
HADS-D: Hospital Anxiety and Depression Scale for depression  
ISI: Insomnia Severity Index  
MGM: mixed graphical model  
NCT: network comparison test  
SCN: suprachiasmatic nucleus  
VIP: vasoactive intestinal peptide
©Christophe Gauld, Sarah Hartley, Jean-Arthur Micoulaud-Franchi, Sylvie Royant-Parola. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 11.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Early Detection of 5 Neurodevelopmental Disorders of Children and Prevention of Postnatal Depression With a Mobile Health App: Observational Cross-Sectional Study

Fabrice Denis¹, MD, PhD; Florian Le Goff², ING; Madhu Desbois², MSc; Agnes Gepner², MD; Guillaume Feliciano³, MSc; Denise Silber³, MBA; Jean-David Zeitoun⁴, MD, PhD; Guedalia Peretz Assuied⁵, MD

¹Institute for Smarthealth INeS, Le Mans, France
²Kelindi, Lille, France
³Basil Strategies, Paris, France
⁴Centre d’Épidémiologie Clinique, Hôtel Dieu Hospital, Assistance Publique Hopitaux de Paris, Paris, France
⁵Service de Pédiatricie, Centre Hospitalier des Pyrénées, Pau, France

Corresponding Author:
Fabrice Denis, MD, PhD
Institute for Smarthealth INeS
64 rue de Degré
Le Mans, 72000
France
Phone: 33 0684190480
Email: fabrice.denis.home@gmail.com

Abstract

Background: Delay in the diagnosis of neurodevelopmental disorders (NDDs) in toddlers and postnatal depression (PND) is a major public health issue. In both cases, early intervention is crucial but too rarely implemented in practice.

Objective: Our goal was to determine if a dedicated mobile app can improve screening of 5 NDDs (autism spectrum disorder [ASD], language delay, dyspraxia, dyslexia, and attention-deficit/hyperactivity disorder [ADHD]) and reduce PND incidence.

Methods: We performed an observational, cross-sectional, data-based study in a population of young parents in France with at least 1 child aged <10 years at the time of inclusion and regularly using Malo, an “all-in-one” multidomain digital health record electronic patient-reported outcome (PRO) app for smartphones. We included the first 50,000 users matching the criteria and agreeing to participate between May 1, 2022, and February 8, 2024. Parents received periodic questionnaires assessing skills in neurodevelopment domains via the app. Mothers accessed a support program to prevent PND and were requested to answer regular PND questionnaires. When any PROs matched predefined criteria, an in-app recommendation was sent to book an appointment with a family physician or pediatrician. The main outcomes were the median age of the infant at the time of notification for possible NDD and the incidence of PND detection after childbirth. One secondary outcome was the relevance of the NDD notification by consultation as assessed by health professionals.

Results: Among 55,618 children median age 4 months (IQR 9), 439 (0.8%) had at least 1 disorder for which consultation was critically necessary. The median ages of notification for probable ASD, language delay, dyspraxia, dyslexia, and ADHD were 32.5 (IQR 12.8), 16 (IQR 13), 36 (IQR 22.5), 80 (IQR 5), and 61 (IQR 15.5) months, respectively. The rate of probable ADHD, ASD, dyslexia, language delay, and dyspraxia in the population of children of the age included between the detection limits of each alert was 1.48%, 0.21%, 1.52%, 0.91%, and 0.37%, respectively. Sensitivity of alert notifications for suspected NDDs as assessed by the physicians was 78.6% and specificity was 98.2%. Among 8243 mothers who completed a PND questionnaire, highly probable PND was detected in 938 (11.4%), corresponding to a reduction of –31% versus our previous study without a support program. Suspected PND was detected a median 96 days (IQR 86) after childbirth. Among 130 users who filled in the satisfaction survey, 99.2% (129/130) found the app easy to use and 70% (91/130) reported that the app improved follow-up of their child. The app was rated 4.8/5 on Apple’s App Store.

Conclusions: Algorithm-based early alerts suggesting NDDs were highly specific with good sensitivity as assessed by real-life practitioners. Early detection of 5 NDDs and PNDs was efficient and led to a possible 31% reduction in PND incidence.

Trial Registration: ClinicalTrials.gov NCT06301087; https://www.clinicaltrials.gov/study/NCT06301087

https://publichealth.jmir.org/2024/1/e58565
Introduction

After birth, the mother-child dyad can be impacted by impairments that remain undetected or are detected too late. Among these impairments, a neurodevelopmental disorder (NDD) such as autism spectrum disorder (ASD) affects 1 in 166 children [1]. The average time to diagnosis is approximately 4 to 6 years, whereas consensus statements indicate that a diagnosis could be made as early as 12 or 18 months of age [2-6]. Interestingly, parents are the main contributors to the NDD screening of their children [7]. Other disorders that deserve early screening are dyspraxia, language delay, dyslexia, and attention-deficit/hyperactivity disorder (ADHD) [8-11]. It is crucial to provide parents with screening tools and to recommend that they consult the physician at the first symptoms to treat them as early as possible.

Postnatal depression (PND) in mothers is another example of an undiagnosed disorder with severe consequences. PND—an episode of depression occurring during the first year after childbirth—has a prevalence of 17% and may have a negative impact on the synchrony or receptivity loop that is crucial to the proper neurodevelopment of the baby [12,13]. All these disorders can benefit tremendously from early detection by electronic patient-reported outcome questionnaires for parents and their children, which would enable early intervention.

We thus developed Malo, an “all-in-one” multidomain digital health record electronic patient-reported outcome app for smartphones, aiming to facilitate early screening of NDDs in children from birth to the age of 10 years and PND in mothers. We previously reported the results of a first observational, cross-sectional, data-based study in a population of 4242 children in 2022, showing a sensitivity of the alert notifications of suspected NDDs (possible ASD, vision, audition, socialization, language, or motor disorders) as assessed by the physicians of 100%, and a specificity of 73.5%. We also reported an earlier detection of PND in 907 mothers showing an incidence rate of 16.6% [14]. Since these results, changes in algorithms were performed to improve the specificity of NDD screening to avoid false positive results and we added early and continuous advice and support programs after childbirth to mothers to reduce the incidence of PND.

We thus report here the results of the revised algorithm aiming to be more specific for the screening of 5 NDDs (ASD, language delay, dyspraxia, dyslexia, and ADHD) and to assess the impact of the app and support program on the reduction in PND incidence.

Methods

Ethical Considerations

We ran an ecological, observational, cross-sectional, data-based study. Our study was approved by the French National Health Data Institute (Health Data Hub approval 16562971), which ensures ethical conduct in human participant research regarding data confidentiality and safety. The approval number for our human participant’s review was F20210420115840. The study complied with good epidemiology practices defined by the Association for French-speaking Epidemiologists. All the users received written information on our primary data collection purpose and the ability to be included in a secondary analysis in a deidentified manner—as specified in the app’s general conditions and privacy policy. However, users did opt in explicitly to be included in this specific study. Data were deidentified and aggregated at the server level before being passed to the researchers.

Data collection was embedded in the app. Data were collected in a French-certified health data cloud, as requested by local laws. Respondents self-entered the age and gender of their infants. The app also allowed for the entry of the children’s height, weight, vaccination status, medical background, and ongoing or previous treatments. No identifying data were used to produce our analysis, with emails, names, and exact dates of birth being segregated in other databases. For instance, the date of birth was converted to the child’s age in days at the time of answer in the data set. No compensation was provided to participants in this study.

Population

Malo is a mobile health (mHealth) app available on iOS and Android—only available in France at the moment. The kick-off of the first version of the app was historically initiated by a French national media campaign that was disseminated through social media between November 11 and 18, 2021. Details of kick-off modalities are provided by Denis et al [14]. The places of recruitment were multiple—3 maternity units, a dozen daycare centers, 15,000 followers’ social media accounts, and some insurers’ marketing efforts.

For this study, we included 55,618 users matching the criteria and agreeing to participate between May 1, 2022, and February 8, 2024. Enrollment in the study was strictly optional. Recruitment was open with no exclusion criteria. The inclusion criteria were to download the app, to have at least 1 child at least 10 years of age, and to provide informed consent (in-app). We extended the inclusion age of children up to 10 years to allow assessment of the incidence of dyspraxia/dyslexia disorders that can occur during this period.
Data Collection

Questionnaires and scales, each containing 25-50 questions assessing neurodevelopmental skills, were automatically submitted every month from birth to 9 months, then at 11, 12, 16, 18, 21, 24, 30, and 36 months and every 6 months until 10 years of age. Questionnaires were focused on sociability, attention or activity, motricity, language of their infants, and possible ASD to screen for ASD, language delay, dyspraxia, dyslexia, and ADHD.

Questionnaires and notifications were based on French health authorities’ reports, international recommendations, and experts’ agreements [8,15,16]. The questionnaire for the screening of PND was submitted to mothers every 2 to 4 weeks for 7 months after childbirth, using a modified questionnaire of the Edinburgh Postnatal Depression Scale adapted for self-assessment.

Threshold-Based In-App Notification and Outcome of NDD Module of the App

Notifications were sent automatically to the user if some symptoms matched predefined criteria and a physician consultation was recommended. Regarding NDDs, once a threshold of concern was reached, a notification was sent recommending that the mothers discuss their symptoms with their general practitioner or pediatrician.

The main outcome of the study was the median age of possible NDD notification of infants. The secondary outcomes were, first, user satisfaction regarding app experience and the level of support in child follow-up and second, relevance of the NDD notifications assessed by physicians, using a specific optional survey asking parents the following questions: (1) “In the last month, did your doctor detect a developmental disorder in your child during a follow-up consultation? YES or NO”; (2) “If you had a notification from Malo, did you follow the recommendation of the app to visit a physician? YES or NO”; and (3) “Which of the following reflects the physicians’ reply? (A) The notification is not relevant, (B) the notification is relevant and a medical surveillance of the evolution of the symptom is needed, (C) the advice of an expert is needed, or (D) a treatment is indicated.”

Threshold-Based In-App Notification and Outcome of PND Module of the App

Regarding maternal PND, there were 4 grades of notifications sent to the mother—grade 0 (score lower than 25) was associated with a message indicating that everything is ok, grade 1 (score between 26 and 50) was associated with a recommendation to talk about symptoms with a close relative, grade 2 (score between 51 and 65) recommended that they quickly discuss their symptoms with a family doctor, and grade 3 (score higher than 65) recommended that they meet a family doctor as soon as possible. Grades 2 and 3 were considered a high probability for PND. This algorithm and questionnaire were the same as those used in our previous study [14].

To reduce PND incidence, we added, since 2022, a support program in the current version of the app with early and continuous advice after childbirth to lead parents to take care of their mental burden and to be aware of burnout and PND. We also provided them with the option of joining speaking groups and accessing testimonials from other mothers.

The last secondary outcome was the rate and the median time of the mothers’ PND notifications after childbirth and subsequently to the support and prevention program.

Analysis

The analysis was performed on the 55,618 users, matching criteria were assessable for analysis and at least 1 neurodevelopment disorder strongly requiring a consultation was observed in 439 children.

Sensitivity, specificity, predictive positive and negative values, and the Youden index of algorithms triggering notifications of suspected NDDs were calculated according to the physician’s feedback. A notification was considered relevant if a physician suggested specific medical surveillance of the disorder or the consultation of an expert or a therapist or initiated therapy.

The χ² test was used in 2×2 tables to assess the statistical association between the medical relevance of the notification (relevant or not) and the notification results (notification or no notification of a possible NDD). We also assessed the rate of probable PND of mothers having a score >50 in the survey and the changes in the rate of PND during the time after childbirth. The level of statistical significance was 5% for all statistical tests.

Results

Overview

Among 99,916 nationwide users of the app between May 1, 2022, and February 8, 2024, a total of 55,618 children met the inclusion criteria (Figure 1). The median age of assessable children was 4 months (IQR 9), and 344,640 questionnaires were analyzed. Data analysis was performed at the end of February 2024. Among children, 0.8% (439/55,618) had at least 1 possible NDD strongly requiring a consultation. The median age of notification for probable ASD, language delay, dyspraxia, dyslexia, and ADHD was 32 (IQR 24-42), 16 (IQR 5-43), 36 (IQR 12-75), 80 (IQR 77-86) and 61 (IQR 52-85) months, respectively (Figure 2). The rate of probable ADHD, ASD, dyslexia, language delay, and dyspraxia in the population of children of the age included between the detection limits of each alert was 1.48%, 0.21%, 1.52%, 0.91%, and 0.37%, respectively.
Figure 1. Flowchart of users of the Malo app.

Nationwide number of users of the Malo app for a toddler (0-10 years of age) n=99,916

Nationwide number of children with at least one questionnaire filled for neurodevelopment screening n=55,618

Users who also filled in the questionnaire for screening of maternal postnatal depression n=8243

Users who filled in the survey about the relevance of neurodevelopment notifications and their satisfaction with them n=130

Questionnaires assessable for relevance of neurodevelopment notifications in users who met physician. n=125
Figure 2. Distribution of the notifications of possible neurodevelopmental disorders and their type according to children’s age.

Analysis of the Assessment of the Relevance of the Alerts by Physicians

A 1-week survey was done at the end of January 2024 concerning the physician consultation feedback to parents following notification of possible NDD and to assess the satisfaction of users. Among the 130 parents who agreed to answer the survey, 113 had no alert, and 17 (13.1%) had received an alert of a possible NDD, which suggested a visit to their physician. Among users who received a notification suggesting a visit to their physician for a neurodevelopmental issue, 70.6% (12/17) users answered “YES” to the question “If you had a notification, did you follow the recommendation of the app to visit a physician?”

As 5 users with notification did not meet their physician, they were not included in the analysis of the clinical relevance of the alerts assessed by the physician. Data of the 125 assessable users who met the physician after notification showed a sensitivity of notification of 78.6%, a specificity of 98.2%, a positive predictive value of 84.6%, a negative predictive value of 97.3%, and a Youden index of 0.77 (P<.001).

Among the 11 children with true positive notifications of a possible NDD suggested by the app, medical surveillance of the evolution of the symptoms was proposed in 6 (54.5% of relevant notifications) cases, the advice of an expert was needed in 2 (18.2%) cases, and treatment was immediately initiated in 3 (27.3%) cases.

Satisfaction Analysis

Among 130 users who filled in the satisfaction survey, 99.2% (129/130) found the app easy to use and 70% (91/130) reported that the app improved the follow-up of their child. Moreover, the app was rated 4.8/5 on Apple’s App Store with 990 votes and 4.7/5 on Android Stores with 1100 votes.

Screening of PND

Among 8243 mothers who completed a PND questionnaire, highly probable PND (grades 2 or 3) was suspected in 938 (11.4%). Grades 1, 2, or 3 were reported in 54.2% (n=4468) of mothers. The a median 96 days (IQR 86) after childbirth and the incidence of supposed PND was equivalent during the 7-month follow-up (P<.001).

Discussion

Principal Findings

Our study prospectively assessed, in a “real-world” manner, in over 55,000 users, the benefit of mother-child dyad follow-up via a dedicated multidomain familial mHealth smartphone app providing early detection of 5 NDDs and leading to reduced incidence of maternal PND.

The main result is that the median age of alert for 5 NDDs (ASD, language delay, dyspraxia, dyslexia, and ADHD) by this smartphone screening app containing a dedicated questionnaire for parents seems to allow earlier assessment than in historical data. It also appears that the relevance of notifications was confirmed by physicians consulted following notification with high sensitivity (78.6%) and high specificity (98.2%).

Optimization of the neurodevelopment follow-up of children is very important as the identification of the first symptoms of NDDs is usually done by parents (without a dedicated digital device) in 61% of cases and by a health professional in only 14% as reported in a recent French study [7]. In our study, the median age of notification for probable ASD, ADHD, dyspraxia, dyslexia, and language delay was 32.5 (IQR 12.8), 61 (IQR 15.5), 36 (IQR 22.5), 80 (IQR 5), and 16 months (IQR 13), respectively, while time to diagnosis in France is usually 78 and 120 months for ASD and ADHD, respectively, that is, 4-5
The incidence of each disorder suspected in our study was similar to those historically reported, strongly suggesting that our cohort was representative of the general population—0.21% for ASD in our study versus 0.6% in the literature [1], 1.48% for ADHD 0.4% versus 5% in the literature [19], and 0.91% for language delay versus 2% in France in the literature [20]. The rate of dyslexia was lower in our study (1.52%) than in the literature (3% to 6% [10,20]) as was the rate of dyspraxia: 0.37% in our study versus 3% in literature [20]. The lower rates of dyslexia and dyspraxia screening in our study are probably caused by a reduced sensibility of the algorithm.

We also performed an analysis of physician feedback after an alert about a possible NDD. Most users (13/18, 72.2%) followed the recommendation of the app to visit their family doctor or pediatrician after an alert. In our previous study published in 2022 with the older version of the NDD trigger algorithm, the rate of parents following the recommendation was 84.4% (54/64). Both results suggest a high level of confidence of parents in the notification [14].

The main modification of the algorithm and forms between the previous and current versions of the app consisted of reducing sensitivity to optimize specificity by displaying several sub-questionnaires over a more extended period (at least 1 month between early warnings and confirmation) for confirmation and investigation of suspicious symptoms on the main questionnaire. In this study, the sensibility of notification was reduced to 78.6% while it was 100% in the 2022 study, but the specificity was higher with new algorithms (98.2% vs 73.5%). This was a deliberate choice to improve specificity by reducing false positive alerts, as observed in the 2022 version of the app to reduce inopportune and anxiogenic alerts to parents. This is also associated with a high negative predictive value of 97.3% which is interesting to reassure parents in the absence of alerts. The high positive predictive value is also an important factor for the confidence of doctors confronted with an alert, delivered by the app. It suggested that 84.6% of notifications were considered relevant by a physician. Among these alerts, the physician triggered a specific medical surveillance in 54.5% of notifications initiated a treatment or recommended parents to an expert in 45.5%.

Although these data were declarative by users and were not directly confirmed by physicians, we assume that the specificity of the ASD notifications is close to the result of Pierce et al [6], showing overall stability or specificity of an autism spectrum diagnosis of 84% at earlier than 18 months of age through a universal screening program in primary care. In a recent diagnostic accuracy study including 13,511 children aged 11-42 months, Barbaro et al [21] showed 83% positive predictive value and 99% estimated negative predictive value of the Social Attention and Communication Surveillance-Revised tool for autism identification when it was used by nurses for children aged 12 months. Our results seem to be similar when parents perform a screening using our app.

The inclusion of efficient digital tools is important in the logic of care pathways because it promotes acceptability and relevance by families and professionals. Early screening allows for early diagnosis and interventions as reported by works on the efficacy of early treatments of cases among young children and recent promising studies on early interventions [9-11,22-25].

We also reported a lower incidence of maternal PND (11.4%, 938/8243) assessed in 8243 mothers than in our previous study performed in 2022 (16.6%, 157/907 assessed mothers). The form and algorithms were not modified for the part relating to maternal postpartum depression between the previous and current versions of the app. To prevent the PND rate, we added to the early PND screening a prevention program to be initiated by mothers after childbirth. PND is well known to disrupt the crucial mother-infant relationship on which optimal child development depends. It is the most common complication associated with childbirth, and it may exert harmful effects on children such as increased risk of ASD [26]. It is usually underdetected or detected after many months. The early treatment of PND is effective, avoids negative impact on child development, and does not necessarily require drugs to improve symptoms in the earliest stages [27]. Its prevalence in France is 18%.

In our previous study performed in 2022 with the old version of the app, we reported that 16.6% of mother users had probable PND and a median time of detection between 8 and 12 weeks after childbirth. We thus added a support program since 2022 in the new version of the app, consisting of advice, information, sensitization to PND, as well as access to speaking groups and testimonials of mothers. This study reported a lower rate of suspected PND of 11.4% (151/907), that is, a 31% reduction in PND incidence compared to the results of the 2022 study. There were no changes in the questionnaire; the same algorithm for PND detection was active in both versions of the app, but only the prevention program was added. Interestingly, the incidence of all grades of assessment (1, 2, or 3) was similar between both studies (56.7%, 515/907 mothers in the 2022 study vs 54.2%, 4465/8243), suggesting that a switch to lower grade symptoms was associated with the new support program which in turn was associated with a diminished incidence of PND. This is, as far as we know, the first time that a reduction in PND incidence can be observed through the use of a smartphone app.

The level of satisfaction was also high (between 70%, 91/130 and 99.2%, 129/130 according to the assessed domains) and contributed to the high rate of adoption.

Limitations
There are limitations to our study. First, it was an observational study without a control group. Therefore, even though our sample was very significant, we could only proceed to an indirect (historical) comparison when intending to assess the efficacy of the tool regarding the detection of mental problems. Sample selection bias is always possible in the absence of randomization, due to social media recruitment modalities and because using the mobile app requires possession of a smartphone. We could have asked users questions about their educational level, practice classification (rural or urban), technical experience, and marital status, but we designed the app to collect as little personal data as possible. However, the very high rate of smartphone penetration in France (92% in a...
2018 survey) in people aged 25-39 years led us to believe that the risk of a selection bias associated with smartphone use was low. Nonetheless, we do note that parents without smartphones cannot benefit from the app [28].

The second limitation is that the data were declarative by users without a comparative arm, but we found similar results to our previous study in terms of NDD incidence and time to the detection of benefit.

The third limitation is that NDD suspicions were not directly transmitted by physicians. As diagnostic confirmation takes time, prospective follow-up of patients can be interesting to assess if suspicion is confirmed and makes it possible to study the confirmation rate of suspicions.

Fourth, the attrition rate (ie, the discontinuation of eHealth app use) was not assessed, but it could be interesting to study whether the benefit of early detection of NDD is maintained over time, thanks to prolonged use [29]. We need further studies to improve the lack of follow-up rate which is usually high in real-life studies of eHealth instruments.

Conclusions
To our knowledge, this multidomain mHealth app dedicated to both the early detection of 5 NDDs in children and the early detection and prevention of maternal PND is the first app with real-life data of clinical relevance. Results based on a large population of more than 55,000 users confirmed previous results and suggested that a multidomain familial mHealth app is suitable and effective for regular use in the mother-child dyad follow-up.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to the sensitive nature of the patient data and the risk for reidentification if analyzed with this study are not publicly available due to the corresponding author on reasonable request.

Authors' Contributions
All authors had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. FD conceptualized and designed the study. All authors contributed to the acquisition, analysis, interpretation of the data, and drafting the paper. Critical revision of the paper for important intellectual content was done by FD, FLG, MD, and AG. FD, FLG, GF, and MD oversaw statistical analysis as well as administrative, technical, and material support.

Conflicts of Interest
FD reports receiving personal fees from AstraZeneca, Ipsen, Kelindi, Pfizer, Chugai, and Roche and has stocks in Kelindi and Institute for Smarthealth. FLG, AG, and MD have stocks in Kelindi. The other authors declare no conflicts of interest. JDZ is an investor in Wandercraft, Kiro, Sim & Cure, Posos, Tilak, Orakl, Kelindi, Intelligent Location, Healthcube, Medicaim, Savana medical, Brightheart, and Deeplife. Kelindi developed the tool around which this study has been conducted.

References
8. Attention deficit disorder with or without hyperactivity: identifying suffering, supporting the child and the family - questions/answers. Haute Autorité de Santé. URL: https://www.has-sante.fr/jcms/c_2025618/fr/


Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ASD: autism spectrum disorder
mHealth: mobile health
NDD: neurodevelopmental disorder
PND: postnatal depression

Please cite as:
Denis F, Le Goff F, Desbois M, Gepner A, Feliciano G, Silber D, Zeitoun JD, Assuied GP
Early Detection of 5 Neurodevelopmental Disorders of Children and Prevention of Postnatal Depression With a Mobile Health App: Observational Cross-Sectional Study
JMIR Public Health Surveill 2024;10:e58565
URL: https://publichealth.jmir.org/2024/1/e58565
doi: 10.2196/58565
PMID: 38888952

© Fabrice Denis, Florian Le Goff, Madhu Desbois, Agnes Gepner, Guillaume Feliciano, Denise Silber, Jean-David Zeitoun, Guedalia Peretz Assuied. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 18.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Predicting the Population Risk of Suicide Using Routinely Collected Health Administrative Data in Quebec, Canada: Model-Based Synthetic Estimation Study

JianLi Wang1, PhD; Fatemeh Gholi Zadeh Kharrat2, PhD; Geneviève Gariépy3, PhD; Christian Gagné2, PhD; Jean-François Pelletier4, PhD; Victoria Kubuta Massamba5, PhD; Pascale Lévesque5, MSci; Mada Mohammed1, BSc; Alain Lesage4, MD

1Department of Community Health and Epidemiology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada
2Institut intelligence et données, Université Laval, Quebec City, QC, Canada
3Centre for Surveillance and Applied Research, Health Promotion and Chronic Disease Prevention Branch, Public Health Agency of Canada, Ottawa, ON, Canada
4Department of Psychiatry, University of Montreal, Montreal, QC, Canada
5Institut national de santé publique du Québec, Quebec City, QC, Canada

Corresponding Author:
JianLi Wang, PhD
Department of Community Health and Epidemiology
Dalhousie University
5790 University Ave
Halifax, NS, B3H 1V7
Canada
Phone: 1 9024736684
Email: JianLi.Wang@dal.ca

Abstract

Background: Suicide is a significant public health issue. Many risk prediction tools have been developed to estimate an individual’s risk of suicide. Risk prediction models can go beyond individual risk assessment; one important application of risk prediction models is population health planning. Suicide is a result of the interaction among the risk and protective factors at the individual, health care system, and community levels. Thus, policy and decision makers can play an important role in suicide prevention. However, few prediction models for the population risk of suicide have been developed.

Objective: This study aims to develop and validate prediction models for the population risk of suicide using health administrative data, considering individual-, health system–, and community-level predictors.

Methods: We used a case-control study design to develop sex-specific risk prediction models for suicide, using the health administrative data in Quebec, Canada. The training data included all suicide cases (n=8899) that occurred from January 1, 2002, to December 31, 2010. The control group was a 1% random sample of living individuals in each year between January 1, 2002, and December 31, 2010 (n=645,590). Logistic regression was used to develop the prediction models based on individual-, health care system–, and community-level predictors. The developed model was converted into synthetic estimation models, which concorded the individual-level predictors into community-level predictors. The synthetic estimation models were directly applied to the validation data from January 1, 2011, to December 31, 2019. We assessed the performance of the synthetic estimation models with four indicators: the agreement between predicted and observed proportions of suicide, mean average error, root mean square error, and the proportion of correctly identified high-risk regions.

Results: The sex-specific models based on individual data had good discrimination (male model: C=0.79; female model: C=0.85) and calibration (Brier score for male model 0.01; Brier score for female model 0.005). With the regression-based synthetic models applied in the validation data, the absolute differences between the synthetic risk estimates and observed suicide risk ranged from 0% to 0.001%. The root mean square errors were under 0.2. The synthetic estimation model for males correctly predicted 4 of 5 high-risk regions in 8 years, and the model for females correctly predicted 4 of 5 high-risk regions in 5 years.
Conclusions: Using linked health administrative databases, this study demonstrated the feasibility and the validity of developing prediction models for the population risk of suicide, incorporating individual-, health system–, and community-level variables. Synthetic estimation models built on routinely collected health administrative data can accurately predict the population risk of suicide. This effort can be enhanced by timely access to other critical information at the population level.

(JMIR Public Health Survell 2024;10:e52773) doi: 10.2196/52773

KEYWORDS
population risk prediction; case-control; development; validation; health administrative data; suicide; depression; anxiety; Quebec; Canada; mental health; suicide prevention; prevention; adolescent; adolescents; teen; teens; teenager; teenagers; male; female

Introduction
Each year, over 4500 Canadians take their own life [1], and more than 700,000 people die because of suicide worldwide [2]. As such, suicide has become a major international public health challenge. To facilitate suicide prevention, mechanisms should be in place that enable policy and decision makers to make informed decisions and mobilize resources to high-risk populations at the right places before tragic events occur. To achieve this goal, methods of predicting the population risk of suicide are critical.

Many suicide risk assessment tools have been developed in clinical settings with the hope that identifying patients at high risk and providing needed mental health services would reduce an individual’s risk of suicide [3,4]. Going beyond individual risk assessment, one important application of risk prediction models is population health planning [4]. In the realm of population health, population refers to “a group of individuals, in contrast to the individuals themselves, organized into many different units of analysis, depending on the research or policy purpose” [5]. These different units may be geographic regions (eg, states, provinces, and cities), groups (eg, workplaces and schools), and policy-relevant subpopulations (eg, by sex/gender and ethnicity). The aggregate health of the populations in these units is the focus of policy and decision makers through population health planning and policy [5]. Suicide has a complex etiology and is a result of the interaction of the risk and protective factors at the individual, health care system, and community levels [6-13]. Therefore, policy and decision makers and mental health service planners can play an important role in suicide prevention. However, policy and decision makers need tools that allow them to identify communities/regions that are at high risk so that they can mobilize resources and complex population-based interventions to these high-risk regions in advance. Ideally, such tools are developed based on readily accessible and real-time data so that they can closely monitor the population risk and the effects of the interventions.

There is a paucity of prediction models for the population risk of suicide. Gradus and colleagues [14] developed sex-specific machine learning algorithms for suicide using data from eight Danish national health and social registries. Kessler et al’s [15] machine learning algorithms targeted US Army soldiers who were hospitalized. Predictive models for population risk may use not only individual data but also health system–level (eg, quality of mental health care and mental health budget) and community-level data (eg, unemployment rate and social deprivation levels in the community). Therefore, a limitation of these individual-oriented models is the lack of consideration of health care system– and community-level factors as well as the potential changes in the distributions of the predictors over time. Research [16-18] has shown that community-level social vulnerability and antidepressant use/prescription are significantly associated with suicide at the population level. Kandula et al [19] went further by building a model to predict county-level suicide mortality in the United States using county-level annual measures of socioeconomic predictors of suicide risk and state-level prevalence of major depressive episodes and firearm ownership. Kandula et al’s [19] model used data from a variety of US public data sources, which makes the application of the model highly feasible. However, health administrative data was not one of the data sources. Studies from different regions have shown that a large proportion of people contacted health services in the year before their death [20-23]. For example, Canadian studies found that 82% of suicide decedents contacted health services in the year before their death [24] and that over 50% specifically contacted mental health services [25]. Therefore, the role of health service use data in predicting suicide risk at the population level cannot be ignored. The objective of this study was to develop and validate sex-specific predictive models for the population risk of suicide based on individual-, health system–, and community-level indicators in Quebec, Canada.

Methods
Population and Setting
The target population is the general population aged ≥15 years residing in the province of Quebec, Canada. Health care in Canada is provided through provincial and territorial systems of publicly funded health care that are universally accessible. In Quebec, health services are planned and delivered through 18 health regions, 22 integrated health and social services centers, and 166 local community services centers. Budgetary decisions are made at the levels of province and health regions.

Data Sources
We linked 5 health administrative databases using residents’ health care insurance numbers and the Canadian Urban Environmental Health Research (CANUE) data by postal codes, including the vital statistics death database, the physician claims database, the hospital discharge database, the Insured Person Registration File, and the public drug plan. The data of these databases (eg, billing and service procedures codes and service dates) are routinely submitted by clinics and hospitals for billing and administration purposes. No self-reported data were collected from residents. These databases cover up to 98% of
the population in Quebec. Together, the 5 health administrative databases can provide information on individual- (eg, sex and age), program- (eg, hospitalization and emergency department visits), and system-level (eg, mental health and addiction budgets) indicators [8]. CANUE is a Canadian consortium database [26], which contains indicators for social deprivation, material deprivation, and built environment at the community level. The data linkage was performed at the Quebec Institute of Public Health (INSQP) where the health administrative databases and CANUE data are kept. This study was approved by the Health Sciences Research Ethics Board of Dalhousie University (2021-5913). The methodological details of this study can be found in a previous publication [27].

Study Design

Because the base rate of suicide in the population is low, we used a case-control study design to develop sex-specific suicide risk predictive algorithms. For the training data set, we included all death by suicide cases that occurred from January 1, 2002, to December 31, 2010. The control group was a 1% random sample of living individuals in each year between January 1, 2002, and December 31, 2010, from the Quebec physician claim database. Individuals in the control group were only allowed to be selected once. The cases and controls were not matched to allow for maximum variability in predictors. The health administrative data from January 1, 2011, to December 31, 2019, were used for validation.

Suicide

Death by suicide cases were ascertained by Quebec’s Coroner’s Office after investigation. The decision is registered in the Quebec vital statistics database.

Predictors

Individual, programmatic, systemic, and community factors (Multimedia Appendix 1) 5 years before the suicide event, or the index date for controls, were used as candidate predictors to develop the risk predictive algorithms. For example, we extracted the data about the diagnosis of major depression (an individual-level factor) in the past 6, 12, 24, 48, and 60 months as 5 separate candidate predictors. Similarly, we extracted the annual mental health and addiction budget of each health region (a systemic-level factor) in the past 5 years as candidate predictors. The percentages of missing values associated with the variables in the databases ranged from 0.87% to 4.12%. The initial selection of candidate predictors was determined by content knowledge (ie, known relationships between suicide or suicide behaviors and individual- and local area–level variables), clinical utility, and policy relevance through team meetings (Multimedia Appendix 1).

Model Development

We included all preselected variables in penalized least absolute shrinkage and selection operator (LASSO) regression. The LASSO penalization factor selects important predictors by shrinking coefficients for weaker predictors toward zero, excluding predictors with estimated zero coefficients from the final sparse prediction model. We performed a correlation analysis among variables selected by LASSO regression and identified variables that were strongly correlated (γ≥0.60). Correlated variables were discussed by team members, and the variables that have better policy implication and clinical utility were kept and became the candidate predictors for model development.

We used logistic regression to develop the sex-specific statistical models. The backward selection method was used to eliminate unpredictive variables and to identify the model with the best calibration and discrimination. The decisions of model selection were initially based on the changes in the values of the Akaike information criterion and Bayesian information criterion [28]. Prediction accuracy was assessed by the discrimination and calibration of the model. Discrimination is the ability of a prediction model to separate those who experienced the outcome events from those who did not. We quantified this by calculating the C statistic, analogous to the area under a receiver operating characteristic curve. C statistic ranges from 0.5 to 1, with a higher value indicating better discrimination. A C statistic of 0.7, 0.8, and 0.9 may be considered the threshold for acceptable, good, and excellent discrimination, respectively. Calibration measures how closely predicted outcomes agree with actual outcomes. We used the Brier score to measure calibration. Brier score is the mean squared difference between the predicted probability and the actual outcome. The lower the Brier score is for a set of predictions, the better the predictions are calibrated. Given that the program will be used to forecast population risk by policy and decision makers, we prioritized calibration over discrimination in model development.

The second step of model development was to estimate the synthetic proportions of suicide. A synthetic estimate is a prevalence estimate for a local area that is calculated by using descriptive or demographic data at the community level [29]. The model-based synthetic estimation consists of two stages. First, for each predictor, the proportion of individuals within each category of that predictor in the initial modeling was computed separately by region. For instance, if hospitalization due to a suicide attempt in the past 5 years is a predictor in the model, the proportion of individuals with this attribute in a specific health region is calculated. If age is a continuous variable in the model, the mean age of the population in a health region is calculated. As such, the synthetic model contained community characteristics as predictors. A syntax program was then developed to apply the regression coefficients to the corresponding proportions and means in the data set, and to calculate the logit estimates for each of the health regions. The resulting logit values for each of the health regions were then converted into probabilities (ie, the synthetic estimate of the risk of suicide in the health region).

Validation

The validation data set included all suicide cases and 1% of controls from January 1, 2011, to December 31, 2019. We first calculated the yearly proportion of suicide at the provincial and health regional levels for males and females (ie, observed proportion). We then applied the developed synthetic models to the validation data to estimate the yearly prevalence of suicide death at the provincial and health regional levels in males and females (ie, predicted proportion). We visually compared and calculated the absolute differences between the predicted and
observed risks; smaller differences indicate better calibration with the data and model accuracy. Additionally, we assessed the synthetic model performance using three indicators: mean average error (MAE), root mean square error (RMSE), and the proportion of correctly identified high-risk regions. The MAE is the average magnitude of the difference between the predicted and observed suicide death rate for each health region. The RMSE is the square root of the average magnitude of the difference squared and is, therefore, similar to MAE but penalizes prediction errors with greater magnitude. More accurate predictions will result in smaller MAE and RMSE. To assess the extent to which high-risk regions are correctly identified, the top 5 health regions with the highest predicted and observed suicide risks were identified. The proportion of health regions observed in the top quartile of observed suicide death risks that were rightly predicted to be in the top 5 was calculated.

Ethical Considerations

This study was approved by the Health Sciences Research Ethics Board of Dalhousie University (REB number: 2021-5913), which waived the need for informed consent. This study used existing data held by the INSPQ, which are routinely collected by the provincial government. Under provincial health information regulations, deidentified INSPQ data were used, as requested, in the context of epidemiological surveillance of suicide, with ethical approval. With the deidentified data, individual patients cannot be identified, and the results of this study were reviewed and vetted by the INSPQ before publication. No compensation was provided to individuals in the databases.

Results

The demographic and socioeconomic characteristics of participants in the training data are in Table 1. Between January 1, 2002, and December 31, 2010, there were 8899 suicide cases (6713 males and 2186 females). We included a 1% random sample of the Quebec general population as controls (316,574 males and 329,016 females). Most of the participants lived in urban areas. Participants were grouped by quartile values of social deprivation and material deprivation scores based on population norms.

The final models for males and females are in Tables 2 and 3, respectively. The model for males included 20 predictors, and the model for females had 22 predictors. The predictors in the models covered the levels of individual (eg, age and mental health physical diagnoses), health system (eg, hospitalization and mental health budget), and community (eg, material deprivation). The male and female models had common predictors (eg, age; living in rural area; hospitalization for suicide attempt; outpatient psychiatrist visits for mental health reasons; and presence of mood, anxiety, substance use, and personality disorders); some predictors are sex specific (in males: Charlson score, emergency and general physician [GP] visits for physical health reasons, and regional mental health budget; in females: material deprivation score and emergency and GP visits for mental health reasons). The sex-specific models had good discrimination (male model: C=0.79, 95% CI 0.78-0.79; female model: C=0.85, 95% CI 0.84-0.86) and calibration (Brier score male model 0.01; Brier score for female model 0.005). Figure 1 shows the visual comparison between the predicted and observed risk of suicide. The models calibrated well with the data, especially the model for females.

We converted the developed models into synthetic estimation models as described in the Methods section and directly applied the models in the development (from 2002 to 2010) and validation (from 2011 to 2019) data. We estimated the annual prevalence of suicide in Quebec from 2002 to 2019 and compared it with the observed risk in each year (Table S1 in Multimedia Appendix 1). During this period, the annual prevalence of suicide in males steadily decreased from 27 per 100,000 in 2002 to 20 per 100,000 in 2019, while the annual prevalence of suicide in females remained stable at around 6 per 100,000 to 7 per 100,000. The predicted annual prevalence of suicide in males and females based on the synthetic estimation models was very close to the observed proportions. Over 18 years, the synthetic estimation models had 1 per 100,000 over- or underestimation in 8 years for males and only in 3 years for females. For the rest of the years, the synthetic estimations were exactly the same as the observed proportions of suicide in the population. The good performance of the synthetic estimation models was also reflected by the small MAE and RMSE, with most of the RMSEs around 0.1. With the validation data from 2011 to 2019, the synthetic estimation model for males correctly predicted 4 of 5 high-risk regions in 8 years, and the model for females correctly predicted 4 of 5 high-risk regions in 5 years (Table 4).

To examine the accuracy and fairness of model prediction, we validated the synthetic estimation models by age groups and health regions using the 2019 data. As seen in Table S2 in Multimedia Appendix 1, the models performed well in different age groups. In 2019, the prevalence of suicide varied by health region, ranging from 14 per 100,000 (Montreal) to 275 per 100,000 (Nunavik) in males and from 4 per 100,000 (Laval) to 92 per 100,000 (Nunavik) in females. The models predicted the same regional variations in males and females with small absolute differences (Table S3 in Multimedia Appendix 1).
Table 1. The sociodemographic characteristics of the participants in the training data (2002-2010).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Men</th>
<th></th>
<th>Women</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (n=316,574), n (%)</td>
<td>Suicide (n=6713), n (%)</td>
<td>Control (n=329,016), n (%)</td>
<td>Suicide (n=2186), n (%)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-39</td>
<td>122,712 (38.76)</td>
<td>1948 (29.01)</td>
<td>120,116 (36.50)</td>
<td>612 (27.99)</td>
</tr>
<tr>
<td>40-59</td>
<td>108,687 (34.33)</td>
<td>2998 (44.65)</td>
<td>108,740 (33.05)</td>
<td>1043 (47.71)</td>
</tr>
<tr>
<td>≥60</td>
<td>85,175 (26.90)</td>
<td>1767 (26.32)</td>
<td>100,160 (30.44)</td>
<td>531 (24.29)</td>
</tr>
<tr>
<td>Urbanicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural town</td>
<td>62,682 (19.80)</td>
<td>1962 (29.22)</td>
<td>60,449 (18.37)</td>
<td>478 (21.86)</td>
</tr>
<tr>
<td>Social deprivation score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (most privileged)</td>
<td>60,096 (18.98)</td>
<td>1167 (17.38)</td>
<td>59,127 (17.97)</td>
<td>310 (14.18)</td>
</tr>
<tr>
<td>2</td>
<td>60,283 (19.04)</td>
<td>1323 (19.70)</td>
<td>60,002 (18.23)</td>
<td>339 (15.50)</td>
</tr>
<tr>
<td>3</td>
<td>59,937 (18.93)</td>
<td>1223 (18.21)</td>
<td>61,491 (18.68)</td>
<td>358 (16.37)</td>
</tr>
<tr>
<td>4</td>
<td>58,681 (18.53)</td>
<td>1162 (17.30)</td>
<td>62,807 (19.08)</td>
<td>427 (19.53)</td>
</tr>
<tr>
<td>5 (most deprived)</td>
<td>58,620 (18.51)</td>
<td>1427 (21.25)</td>
<td>63,683 (19.35)</td>
<td>592 (27.08)</td>
</tr>
<tr>
<td>Material deprivation score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (most privileged)</td>
<td>58,609 (18.51)</td>
<td>875 (13.03)</td>
<td>62,815 (19.09)</td>
<td>323 (14.77)</td>
</tr>
<tr>
<td>2</td>
<td>58,650 (18.52)</td>
<td>1034 (15.40)</td>
<td>61,194 (18.59)</td>
<td>365 (16.69)</td>
</tr>
<tr>
<td>3</td>
<td>60,167 (19.00)</td>
<td>1274 (18.97)</td>
<td>61,428 (18.67)</td>
<td>378 (17.29)</td>
</tr>
<tr>
<td>4</td>
<td>60,102 (18.98)</td>
<td>1450 (21.59)</td>
<td>61,349 (18.64)</td>
<td>449 (20.53)</td>
</tr>
<tr>
<td>5 (most deprived)</td>
<td>60,089 (18.98)</td>
<td>1669 (24.86)</td>
<td>60,324 (18.33)</td>
<td>511 (23.37)</td>
</tr>
</tbody>
</table>

*a*Missing: men in control group: n=4611, 1.45%; men in suicide group: n=30, 0.44%; women in control group: n=3196, 0.97%; women in suicide group: n=6, 0.22%.

*b*Missing: men in control group: n=18,957, 5.98%; men in suicide group: n=411, 6.12%; women in control group: n=21,906, 6.65%; women in suicide group: n=160, 7.31%.
Table 2. The predictive model for suicide in males based on individual data from 2002 to 2010.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Odds ratio</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.01</td>
<td>1.01</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Charlson score</td>
<td>0.11</td>
<td>1.12</td>
<td>.01</td>
</tr>
<tr>
<td>Rural town_60m</td>
<td>0.44</td>
<td>1.55</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hospitalisation for Suicide attempt_60m</td>
<td>1.88</td>
<td>6.55</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Emergency room visits for Physical health reasons_3m</td>
<td>1.00</td>
<td>2.72</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Outpatient psychiatrist visits for mental health reasons_60m</td>
<td>0.67</td>
<td>1.95</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Outpatient GP visits for Physical health reasons_60m</td>
<td>–0.15</td>
<td>0.86</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Psychotherapy visits with a GP_3m</td>
<td>0.24</td>
<td>1.27</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mood and anxiety disorders_60m</td>
<td>0.99</td>
<td>2.69</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Substance use disorders_60m</td>
<td>0.89</td>
<td>2.44</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Personality disorders_60m</td>
<td>0.41</td>
<td>1.51</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Respiratory disorders_60m</td>
<td>–0.28</td>
<td>0.76</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Other mental disorders_60m</td>
<td>0.26</td>
<td>1.30</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Symptoms, Signs and Ill-defined Conditions_3m</td>
<td>0.13</td>
<td>1.14</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Non-intentional trauma_48m</td>
<td>0.21</td>
<td>1.23</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Infectious disease_6m</td>
<td>0.30</td>
<td>1.35</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Endocrine system disorder_48m</td>
<td>–0.15</td>
<td>0.86</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Genito-urinary disorders_24m</td>
<td>–0.08</td>
<td>0.92</td>
<td>.01</td>
</tr>
<tr>
<td>Cancer_60m</td>
<td>–0.15</td>
<td>0.86</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Regional mental budget</td>
<td>–0.001</td>
<td>1.00</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Constant</td>
<td>–4.85</td>
<td>__c</td>
<td>—</td>
</tr>
</tbody>
</table>

aReceiver operating characteristic curve 0.79 (95% CI 0.78-0.79); Brier score 0.01.

bGP: general physician.

cNot applicable.
Table 3. The predictive model for suicide in females based on individual data from 2002 to 2010a.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Odds ratio</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.00</td>
<td>1.00</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Rural town _48m</td>
<td>0.19</td>
<td>1.21</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Material deprivation_2</td>
<td>-0.06</td>
<td>0.94</td>
<td>.38</td>
</tr>
<tr>
<td>Hospitalisation for Suicide attempt_60m</td>
<td>1.85</td>
<td>6.36</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hospitalisation for Physical health reasons_24m</td>
<td>0.35</td>
<td>1.42</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hospitalisation for mental health reasons_36m</td>
<td>0.37</td>
<td>1.45</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Duration of hospitalisation for Physical health reasons_6m</td>
<td>0.007</td>
<td>1.01</td>
<td>.01</td>
</tr>
<tr>
<td>Emergency room visits for mental health reasons_60m</td>
<td>1.13</td>
<td>3.10</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Outpatient psychiatrist visits for mental health reasons_60m</td>
<td>0.66</td>
<td>1.93</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Outpatient GPb visits for mental health reasons_36m</td>
<td>-0.01</td>
<td>0.99</td>
<td>.94</td>
</tr>
<tr>
<td>Mood and anxiety disorders_12m</td>
<td>1.05</td>
<td>2.86</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Bipolar disorders_60m</td>
<td>0.08</td>
<td>1.08</td>
<td>.31</td>
</tr>
<tr>
<td>Substance use disorders_60m</td>
<td>0.84</td>
<td>2.32</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Endocrine system disorder_60m</td>
<td>-0.25</td>
<td>0.78</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Personality disorders_60m</td>
<td>0.47</td>
<td>1.60</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Dementia_60m</td>
<td>-0.65</td>
<td>0.52</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Genito-urinary disorders_48m</td>
<td>-0.23</td>
<td>0.79</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Symptoms, Signs and Ill-defined Conditions_12m</td>
<td>0.22</td>
<td>1.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Non-intentional trauma_36m</td>
<td>0.41</td>
<td>1.51</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Respiratory disorders_24m</td>
<td>-0.10</td>
<td>0.90</td>
<td>.04</td>
</tr>
<tr>
<td>Other mental disorders_60m</td>
<td>0.10</td>
<td>1.11</td>
<td>.02</td>
</tr>
<tr>
<td>Infectious disease_36m</td>
<td>0.13</td>
<td>1.14</td>
<td>.01</td>
</tr>
<tr>
<td>Constant</td>
<td>-6.47</td>
<td><em>c</em></td>
<td>—</td>
</tr>
</tbody>
</table>

aReceiver operating characteristic curve 0.85 (95% CI 0.84-0.86); Brier score 0.005.
bGP: general physician.
cNot applicable.

Figure 1. The calibration plots for the male and female models based on individual training data from 2002 to 2010. (A) Calibration plot of male model. (B) Calibration plot of female model. Participants were grouped in 10 bins based on the risk of suicide. “Uncalibrated” is the observed risk of suicide in each bin. “Calibrated” is the predicted risk of suicide in each bin.
Table 4. The performance of the synthetic estimation models by years.

<table>
<thead>
<tr>
<th>Years</th>
<th>Model for males</th>
<th>Model for females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean average error</td>
<td>Root mean square error</td>
</tr>
<tr>
<td>2011</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td>2012</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>2013</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td>2014</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td>2015</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td>2016</td>
<td>0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>2017</td>
<td>0.02</td>
<td>0.17</td>
</tr>
<tr>
<td>2018</td>
<td>0.03</td>
<td>0.17</td>
</tr>
<tr>
<td>2019</td>
<td>0.02</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Discussion

Principal Results

The data showed a sex difference in the trend of suicide risk and considerable variations in suicide risk by health regions in Quebec. The suicide risk in males had decreased since 2002, while the risk remained stable in females. This study demonstrated the feasibility of integrating individual-, program-, health care system-, and community-level data to build accurate prediction models for suicide at the population level. The models performed well in predicting suicide at both the provincial and health regional levels. The absolute difference between the observed and predicted proportions of suicide ranged from 0 to 1 per 100,000. The RMSEs were under 0.2. The prediction models could correctly identify the health regions that were at the top risk level, and the models achieved good performance in different age groups and health regions.

Limitations

This study had several limitations. First, data about social determinants of health, medication use, the use of crisis hotlines, and access to lethal means were not available in the health administrative data. As such, we were not able to examine the extent to which these factors may improve the performance of the models. Second, although the predictors in the models were associated with suicide, causal inferences cannot be made. The goal of risk prediction models is to identify a key set of factors that in combination best predict the outcome. The models are not meant to test a hypothesis or make inferences about etiology. Third, the relationships between the selected factors and suicide are complex. The logistic regression model is a linear function. Although we found no evidence of interactions among the selected predictors, nonlinear relationships between some predictors and suicide are still possible. Future studies may test if models using machine learning techniques have a better performance.

Comparison With Prior Works

The many factors in the models (eg, older age, living in rural areas, hospitalization for suicide attempt, emergency department visits for mental and physical health problems, and diagnosis of a mood/anxiety or substance use–related disorder) were associated with increased risk of suicide, which was consistent with the literature [6]. Outpatient GP visits for mental health (in females) and physical health (in males) were negatively associated with suicide in multivariate models. Audits of suicide cases, if aggregated, pointed toward the deficits in the health, mental health, and addiction services systems [8]. The work conducted in Quebec by our group [30] recommended better detection and treatment of substance use disorders at the primary (GPs) and specialist medical care levels, access to GPs and psychotherapy for common mental disorders and substance use disorders in the primary care context, mobile crisis teams operating from the emergency room (ER), public campaign targeting men about depression and substance use disorders as treatable diseases, and increasing the specialist mental health and addictions services budgets. This may explain the
relationships between GP visits for mental health problems, patients with visits to their GP for endocrine or genitourinary system–related problems, dementia, and the decreased risk of suicide because these individuals have had the opportunity to be diagnosed with depression and be supported. Patients with infectious diseases secondary to comorbid substance use disorders may have been undetected and untreated, which may explain the negative association. Similarly, at the program level, deficits in the coordination of specialist addiction and mental health services for known patients seen at the ER were found in one-third of the cases, which may explain the negative association with ER visits. Finally, the increased mental health and addiction services was a key system-level recommendation by the Coroner’s Office to the provincial public-managed care system that can be implemented in the allocation of regional health budgets [31]. Although material deprivation was negatively associated with suicide in females, the association was not statistically significant. More granular data about socioeconomic status may provide more insights about the sex differences.

One unique feature that differentiates population risk prediction from individual risk prediction is the use of community-level characteristics as predictors. The data for estimating these community-level parameters may come from various sources. For example, Kandula et al [19] modeled county-level suicide risk in the United States using county-level predictors derived from 8 different sources (government programs, health surveys, and private organizations). For some predictors such as the prevalence of major depressive episodes, only state-level estimates were available and these estimates were extrapolated to the counties [19]. Hudson [32] explored the utility of a regression synthetic estimation model that incorporated individual data from the National Comorbidity Survey, census, and hospital administrative data to predict state-level prevalence of severe mental illness. The advantages of these population risk prediction models are the use of community-level predictors from existing sources or published research and the ability to adapt the models to the local context. Notably, our study used the regression synthetic estimation modeling approach. We used the provincial health administrative data and the CANUE indicators from a single source (ie, INSPQ). The use of a single data source may improve the efficiency of data analysis, data access, and eventually the decision-making process. On the other hand, the use of a single data source may miss some important information. In Canada, the provincial health administrative data are collected primarily for billing and administration purposes. It does not include granular data about social determinants of health (eg, race/ethnicity, poverty, employment, and housing quality), crime rate, social support, the use of crisis hotlines, access to lethal means, access to private health care, and medication use (except for children, youth, and seniors). These factors have been found to be associated with suicide risk, but pertinent data are collected and maintained by different organizations. Future studies should investigate how important data from different sources may be feasibly and efficiently integrated, which factors can improve the model performance, and the feasibility of local adaptation and implementation of the developed models.

One critical element of building risk prediction models is assessing model performance and model validation. This is to ensure that the developed model is accurate and has good performance in different populations or at different time periods. In this study, we developed the models using data from 2002 to 2010 and validated the models using the data from 2011 to 2019. Furthermore, we validated the models in different age groups and health regions. These models were designed to predict population risk and to identify high-risk regions/communities, not to be used by clinicians to identify high-risk individuals. Therefore, the focus of model performance assessment can be different. Specifically, the few existing population risk prediction models for suicide and mental illness focused on model calibration. Kandula et al [19] used symmetric proportional error (observed deaths – predicted deaths)/(observed deaths + predicted deaths) to quantify model calibration. Hudson [32] calculated the absolute difference between the predicted and observed prevalence of severe mental illness. In this study, we calculated the absolute difference between the observed and model-predicted proportions of suicide. Additionally, following the approach of Marks et al [33], we used the MAE, RMSE, and the proportion of correct identifications of high-risk regions as the indicators for model calibration. There is no consensus about the thresholds for absolute difference between predicted and observed proportions of correct identification. Consultations with knowledge users are needed to understand what indicators are informative about model performance and the level of the model error that is acceptable.

The results of this study are expected to have implications for population mental health planning. Few would deny that resource allocation should be partly driven by needs, and needs assessments typically require the knowledge of potential changes in prevalence estimates and in local population profiles (eg, their demographics, diagnoses, and mental health services use). The prediction models developed by this study will allow decision makers and mental health service planners to forecast the proportions of suicide in the years to come at the provincial (state) and health regional (county) levels based on the potential changes in local population profiles. Such profile changes may be estimated using health administrative data and national population census data. Additionally, region-specific risk estimates can help categorize health regions (eg, regions with relatively stable suicide risk, especially those that remained in the highest or lowest groups, or regions in which the largest year-to-year changes are observed) and hence help identify areas in greater need of preventive resources or, conversely, identify areas where interventions seem to be effective. Furthermore, if the data about the predictors are available on a more frequent basis (eg, monthly or biweekly), the models will support the development of nowcast suicide surveillance systems.

Conclusions

Accurate prediction of the population suicide risk can play an important role in suicide prevention. This information can allow policy and decision makers and mental health service planners to categorize regions/communities that are at high risk and to monitor changes in risk so that they may mobilize resources [31] and interventions to the right populations and the right
places at the right time. Using linked health administrative databases, our study demonstrated the feasibility and validity of developing prediction models for population suicide risk, incorporating individual-, health system-, and community-level variables. Routinely collected health administrative data are readily accessible to policy and decision makers and mental health service planners. Suicide risk prediction models based on health administrative data can provide useful information to policy/decision makers at the moment they need the information. This effort can be enhanced by timely access to other critical information at the population level. However, the methodology of population risk prediction should be further studied to enhance the validity and precision of population risk prediction.

Acknowledgments
This study is supported by a grant from the New Frontiers in Research Funds (NFRF) of the Tri-Council Secretariat, Government of Canada, and a Tier I Canada Research Chair in Health Data Science and Innovation from the Canadian Institutes of Health Research (CIHR). However, the NFRF and CIHR played no role in the study design, data collection, analysis, and interpretation.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File. 20 KB - publichealth_v10i1e52773_app1.docx ]

References


26. CANUE. URL: [https://canue.ca/] [accessed 2024-03-17]


Abbreviations

**CANUE**: Canadian Urban Environmental Health Research

**ER**: emergency room

**GP**: general physician

**INSQP**: Quebec Institute of Public Health

**LASSO**: least absolute shrinkage and selection operator
MAE: mean average error
RMSE: root mean square error
Factors Influencing Data Quality in Electronic Health Record Systems in 50 Health Facilities in Rwanda and the Role of Clinical Alerts: Cross-Sectional Observational Study

Hamish S F Fraser\(^1\)*, MBChB, MSc; Michael Mugisha\(^2\)*, BSc, MPH; Ian Bacher\(^1\), BA, PhD; Joseph Lune Ngenzi\(^2\), BSc, MSc; Christopher Seeberegs\(^3,4\), BSc, PhD; Aline Umubyeyi\(^2\), MD, PhD; Jeanine Condo\(^5,6\), MPH, MD, PhD

\(^1\) Brown Center for Biomedical Informatics, Brown University, Providence, RI, United States
\(^2\) School of Public Health, University of Rwanda, Kigali, Rwanda
\(^3\) Jembi Health Systems, Cape Town, South Africa
\(^4\) University of Cape Town School of Public Health, Cape Town, South Africa
\(^5\) Center for Impact, Innovation and Capacity Building for Health and Nutrition (CIIC-HIN), Kigali, Rwanda
\(^6\) Tulane University, New Orleans, LA, United States

\(*\) these authors contributed equally

**Corresponding Author:**
Hamish S F Fraser, MBChB, MSc
Brown Center for Biomedical Informatics
Brown University
233 Richmond Street
Providence, RI, 02903
United States
Phone: 1 4018631815
Email: hamish_fraser@brown.edu

**Abstract**

**Background:** Electronic health records (EHRs) play an increasingly important role in delivering HIV care in low- and middle-income countries. The data collected are used for direct clinical care, quality improvement, program monitoring, public health interventions, and research. Despite widespread EHR use for HIV care in African countries, challenges remain, especially in collecting high-quality data.

**Objective:** We aimed to assess data completeness, accuracy, and timeliness compared to paper-based records, and factors influencing data quality in a large-scale EHR deployment in Rwanda.

**Methods:** We randomly selected 50 health facilities (HFs) using OpenMRS, an EHR system that supports HIV care in Rwanda, and performed a data quality evaluation. All HFs were part of a larger randomized controlled trial, with 25 HFs receiving an enhanced EHR with clinical decision support systems. Trained data collectors visited the 50 HFs to collect 28 variables from the paper charts and the EHR system using the Open Data Kit app. We measured data completeness, timeliness, and the degree of matching of the data in paper and EHR records, and calculated concordance scores. Factors potentially affecting data quality were drawn from a previous survey of users in the 50 HFs.

**Results:** We randomly selected 3467 patient records, reviewing both paper and EHR copies (194,152 total data items). Data completeness was >85% threshold for all data elements except viral load (VL) results, second-line, and third-line drug regimens. Matching scores for data values were close to or >85% threshold, except for dates, particularly for drug pickups and VL. The mean data concordance was 10.2 (SD 1.28) for 15 (68%) variables. HF and user factors (eg, years of EHR use, technology experience, EHR availability and uptime, and intervention status) were tested for correlation with data quality measures. EHR system availability and uptime was positively correlated with concordance, whereas users’ experience with technology was negatively correlated with concordance. The alerts for missing VL results implemented at 11 intervention HFs showed clear evidence of improving timeliness and completeness of initially low matching of VL results in the EHRs and paper records (11.9%-26.7%; \(P < .001\)). Similar effects were seen on the completeness of the recording of medication pickups (18.7%-32.6%; \(P < .001\)).
Conclusions: The EHR records in the 50 HFs generally had high levels of completeness except for VL results. Matching results were close to or >85% threshold for nondate variables. Higher EHR stability and uptime, and alerts for entering VL both strongly improved data quality. Most data were considered fit for purpose, but more regular data quality assessments, training, and technical improvements in EHR forms, data reports, and alerts are recommended. The application of quality improvement techniques described in this study should benefit a wide range of HFs and data uses for clinical care, public health, and disease surveillance.

(JMIR Public Health Surveill 2024;10:e49127) doi:10.2196/49127

KEYWORDS
data quality; electronic health record; EHR; electronic medical record; EMR; HIV; Rwanda

Introduction

Background

By 2021, over 67% of people living with HIV were from sub-Saharan Africa [1]. Rwanda, along with other East African countries, has made great progress in the care and treatment of patients living with HIV/AIDS, including improvements in the uptake of prevention of mother-to-child transmission services and reduction in the rate of loss to follow-up for patients receiving antiretroviral therapy (ART). In 2019, a total of 84% of HIV-positive adults in Rwanda knew their status, 98% of those who knew their status were receiving ART, and 90% of those receiving ART had suppressed viral loads (VLs) [2]. HIV care has been available in Kigali (the capital) approximately from 2002. In 2005, partly through a collaboration between the Rwandan Ministry of Health (MOH) and the health care nongovernmental organization Partners in Health and Inshuti Mu Buzima (PIH-IMB), among other global partners, including the Centers for Disease Control and Prevention (CDC) and Global Fund, work began to scale up treatment to the whole population in need [3,4]. This scale-up treatment was designed around a community-based model, including routine care provision by nurses. Furthermore, it included the training and deployment of community health care workers for home follow-up and support and directly observed therapy for some patients. Ensuring long-term high-quality care for chronic diseases like HIV/AIDS also requires effective documentation of enrollment and follow-up care, with reliable access to clinical data at the point of care and quality improvement processes and reporting [5,6]. In 2005, PIH-IMB deployed a web-based medical record system, the HIV–electronic medical record (HIV-EMR), that had been developed for HIV care in Haiti [5]. These initial strategies were coupled with task shifting and scaling of diagnostics and treatment across Rwanda, reducing barriers related to geographical and financial access to the management of HIV/AIDS.

Digital Health Electronic Health Records

In the last 5 years, Rwanda has initiated digital health as one of the 7 key pillars of the Smart Rwanda Master Plan, which involves the digitalization of different processes of health care delivery and management. Recently, significant efforts were made to deploy information systems and numerous digital tools to support various aspects of service delivery in the health sector [2] with the aim of reaching the one citizen, one record principle. The focus of building the national digital health enterprise architecture was to support seamless information exchange and the need to develop Rwanda Digital Health Interoperability Standards and Health Information Exchange, now known as OpenHIE.

Electronic health records (EHRs) are widely used for collecting, storing, viewing, analyzing, and sharing of clinical data. They address critical limitations in paper records that have been used for generations to support health care and are still relied upon in many lower-income countries. The potential benefits of EHRs over paper records include the following capabilities: unrestricted access to patients’ records for routine or emergency care, often in multiple locations; secure storage of data with backups; search and retrieval of patient records and key clinical data; clinical decision support to improve quality of care; generating reports; sharing data between facilities; disease surveillance; and clinical research. The success of these activities depends on the design and cost of the software; the availability of suitable hardware and infrastructure, particularly in low-income settings; and the ability and motivation of the staff to use the system [7-9]. Effective EHRs need to address important priorities of stakeholders (value of the system) [10,11] and be usable by staff as part of their routine work, which requires attention to the user interface, clinical workflow, and training [11-13] (ease of use). Regardless of the specific system in use, one metric is critical for creating value from the system in the form of data for decision-making processes: data quality [8,14]. If an EHR implementation cannot achieve and sustain good data quality, it is unlikely to create real value for stakeholders, whether they are clinical staff caring for patients, members of the health facility (HF) management, members of ministries of health conducting disease surveillance, or researchers studying questions such as disease incidence or management.

What constitutes good data quality depends on the use of the data: the principle of “data which is fit for purpose.” Clinical care requires data that are as complete and accurate as possible, but specific variables, such as medications, past medical history, or laboratory results, are often most important [15]. Missing data erode the value of the EHR and trust in the system. This typically leads to staff questioning the value of using the systems and keeping data up to date, which can lead to a downward cycle of deterioration and sometimes abandoning the EHR [9]. Conversely, a well-maintained system that is valuable for clinical tasks and reporting can incentivize better data entry. Many strategies have been used to improve data entry, including mapping the system’s workflow to the typical task order of staff, providing warnings of inappropriate values for a form field, and reporting values inconsistent with previous entries such as body weight or age. Training staff in system use and the importance
of complete and up-to-date data are also priorities [16]. One essential strategy is monitoring data entry and completeness and periodic audits of data quality. As noted previously, data quality and completeness are typically linked. A fall off in data entry, for example, may be due to a technical issue or a lack of staff but can lead to gaps in the records and potential falls in data quality. Conversely, catching a lack of data entry quickly can identify correctable technical issues and allow early intervention to maintain momentum and confidence [17].

Many of these issues are more challenging in low- and middle-income countries (LMICs). In addition, the loss of reliable EHR data may have a greater impact on patient care and health system performance in LMICs, given the frequent lack of alternative records or strong health systems. Previous studies have investigated the benefits of different data entry and quality strategies, often in a small number of HFs [6,18]. However, many of these questions are most relevant in EHR systems deployed at scale and in longer-term use rather than pilot projects and recent deployments. There are important unanswered questions on what strategies are most effective at improving data quality, and what factors related to HFs and staff can facilitate or impede data quality [14]. Hedt et al [19] studied the quality of data in paper records of patients with HIV in Malawi using a technique termed lot quality assurance sampling (LQAS). Their study revealed high levels of completeness and accuracy of these records, all of which were reviewed by a supervising team on a quarterly basis. Furthermore, they studied how accurately the overall data quality could be predicted from a random sample of 76 records in each of the 19 HFs. This sample accurately predicted the overall data quality when a threshold of 85% on quality metrics was used to indicate a low quality.

**EHR Data Use for Clinical and Public Health**

Routine clinical data, such as those collected from EHRs, are an important component of public health systems in high-income countries [20]. Moreover, the data in EHRs can help to understand the patterns of HIV distribution, the progress in achieving national and global targets, oversight of the risks and patterns in specific population groups, and the evolution of the infection in different circumstances [21,22]. EHR data shared across HFs may help to realize continuity of care as people move from one region to another. LMICs are increasingly adopting EHRs, with HIV care being in the vanguard in Eastern and Southern Africa, providing a test case for broader EHR adoption and use. The quality and timeliness of these data are a key factor in public health decision-making and management of programs to improve population health outcomes. EHR data are used for public health surveillance, patient to provider communication, quality of care monitoring, linkage of primary care and public health system such as notifiable disease reporting [23], and strategic policy direction [24]. Health Information Exchanges linking EHR and laboratory data, for example, have a growing role in population health and disease surveillance [25]. To maximize the benefits of EHRs for clinical and public health functions, they increasingly allow data collection on behavioral, social, and environmental factors that may affect patient and population health [26]. EHR systems in LMICs, such as OpenMRS, are able to interoperate with pharmacy information systems, laboratory information systems, and mobile health (mHealth) apps, including through the Health Level-7 Fast Healthcare Interoperability Resources (HL7 FHIR; Health Level Seven International) standard, creating more comprehensive views of patient care, population, and public health across a range of communities [27].

**The OpenMRS EHR**

Recognizing the need for a general-purpose EHR system, PIH in collaboration with the AMPATH project (Academic Model Providing Access to Healthcare) in Eldoret, Kenya; the Regenstrief Institute in Indiana, United States; and the South African Medical Research Council deployed a new open-source general-purpose EHR for HIV care in 2006 [28]. This system, the OpenMRS EHR, was subsequently scaled to 43 PIH-IMB-supported HFs in Rwanda. From 2011 onward, it was rolled out to >300 MOH-run HFs. OpenMRS presently supports comprehensive services in all departments of district hospitals. Furthermore, it supports care for maternal-child health, noncommunicable diseases, and oncology in some health centers and hospitals in Rwanda and other countries [7]. OpenMRS is used in >44 LMICs [7], including support for HIV care in >5000 HFs, primarily in Kenya, Uganda, Mozambique, Nigeria, and Rwanda, and will soon be deployed widely in Ethiopia and Haiti [29]. OpenMRS continues to be upgraded with recent improvements in the user interface, data analysis and export, standardized clinical decision support systems, and improved interoperability using HL7 FHIR [7,27]. Previous studies on data quality in OpenMRS have been conducted [6], but only 1 published study by Muthee et al [14] evaluated data quality at scale in >50 HFs.

**Rwanda OpenMRS Evaluation**

This study was part of a larger process evaluation and randomized controlled trial (RCT) of OpenMRS in Rwanda. The US CDC funded a process evaluation of the OpenMRS EHR system in use at scale in Rwanda. This included (1) a user survey, (2) key informant interviews, (3) monitoring of system use and technical stability, (4) a study of the costs of development and deployment of the enhanced EHR software, and (5) the data quality study described in this paper.

Furthermore, a cluster RCT on the impact of improving workflow and adding decision support tools on the quality of HIV care is currently being completed. Three types of alerts were implemented in the trial:

1. A warning if a patient newly diagnosed with HIV was not started on ART within 2 weeks (in the same clinic)
2. A warning for a missing VL result for patients after 6 months of HIV care and annually thereafter (with a 2-month window for test results to be returned and entered into the EHR)
3. A warning for abnormal VL results (>1000) suggesting virologic failure

A total of 112 HFs were included in the RCT, half of which (56/112, 50%) were randomized to receive alert 1; 28 (25%) of these HFs also received alert 2, and 14 (12.5%) of these also received alert 3. The alerts were triggered when clinicians opened the patient summary and were also presented as a report...
of all patients currently matching the rule. The end users were trained to interpret the alerts in the patient summaries and reports. The larger study design has been described previously [4].

We aimed to measure data completeness, matching, concordance, and timeliness and to evaluate factors influencing data quality and strategies to improve it. The data studied were sourced from OpenMRS, a widely deployed EHR system supporting HIV care in Rwanda for >15 years [7,30]. It builds on a clinical trial of the effects of clinical decision support tools on HIV care. Furthermore, it includes data on the characteristics of the HFs and the staff who work there, building on the results of an EHR user survey of the same group of facilities [4].

Methods

Overview

In this paper, we describe a cross-sectional study based on the LQAS approach used by Hedt et al [19]. A total of 50 (44.6%) HFs were selected from the 112 HFs included in the clinical trial, based on the power calculations for that study, selecting a random sample from both intervention and control HFs and a representative mix of larger and smaller HFs but not hospitals. Trained research assistants from the Rwanda School of Public Health visited the 50 HFs and collected data using Open Data Kit (ODK), an open-source mobile data collection software [31]. At each HF, depending on the patient volume, 48 to 76 charts were randomly selected from the set of patients receiving HIV care based on the LQAS protocol used by Hedt et al [19].

The deidentified data were uploaded to a study database housed at the University of Rwanda School of Public Health. Data analyses were partly based on the work of Hedt-Gautier et al in Malawi [19] and Rwanda on public health reporting systems and Muthee et al [14] on data quality analysis in the KenyaEMR OpenMRS-based system.

Analyses included the following metrics:

- Comparison of the paper records and EHR records for completeness of the data items
- Comparison of the paper records and EHR records for percentage of matching of the data items
- Concordance scores comparing the combined matches of 15 key variables
- Comparison of HFs that were recorded as having high EHR uptime and availability (in the previous user survey by Fraser et al [4]) to those with poor EHR uptime
- Comparison of HFs based on the duration for which staff have worked with the EHR and their level of technological experience
- Comparison of completeness of key variables in the EHR records of intervention HFs with those of control HFs, for records completed after the intervention was implemented

Data Processing and Analysis

First, the data were processed to identify any missing values. Values were regarded as missing if the field was blank, was coded with an answer indicating that the information was not available, and, for date variables, if the date reported was impossible (ie, either too far in the future or dates related to HIV treatment that predated any HIV treatment occurring in Rwanda). The latter category helped address the fact that “missing” dates were often recorded as January 1, 1980. Then, for all 28 variables in the data set, we counted the number of records that did not have missing data in either the paper file or the EHR and recorded that value as the number filled or complete. This value was divided by the total number of records to obtain the percentage that were complete. The data were scored as “high quality” if ≥85% of the records were complete based on the LQAS protocol [19].

Then, from the subset of records that were complete, we analyzed how many of them matched in both the paper file and the EHR. Before determining the number of matches, we performed basic data standardization. Specifically, drug regimens were sometimes reported as multiple combinations, for example, the medications could have been coded as Tenofovir (300) + Lamivudine (300) and Efavirenz (600) instead of Tenofovir (300) + Lamivudine (300) + Efavirenz (600); for these cases, we standardized the drug list to 1 standard code. For 22 data points (minus VL values), data were regarded as a match if the standardized value recorded in the paper file exactly matched the standardized value recorded in the EHR. Data quality for matches was regarded as “high quality” if 85% of the records that were complete matched in both the paper file and the EHR.

For each patient, we also collected the last 3 VL results and dates, if available. There were few “matching” records for VL results. Further analysis showed that this usually occurred when the paper file had 1 newer VL result than the EHR; therefore, the most recent VL result in the EHR exactly matched the second most recent VL result in the paper file. To account for this, we aligned the records based on date; for example, if the EHR had a record for May 1, 2019, this would be treated as being the same entry as a record from the paper file that was also dated May 1, 2019. VL results recorded with a date in the paper file that did not match the date of a recorded VL result in the EHR were counted as missing in the EHR and vice versa. Because we only had the last 3 records from each file, in any case where the paper record had newer data than the EHR, there were always at least 2 missing records. To account for this, we aligned records in the EHR and paper file based on date and produced 2 derived variables: “All Viral Loads,” which summarizes all VL records but counts matches based on date rather than position, and “At Least 1 Viral Load,” which examines whether the most recent value in the EHR matches one of the 3 values in the paper file. Data completeness for filled VLs was regarded as “high quality” if ≥85% of the records existed in both files.

After the date alignment was completed, we standardized values such that VL numbers with a small difference within 1 copy/mL were counted as a match. This helped address cases where VL
results were reported with additional precision (20.6 vs 20 or 19.2 vs 20), noting that the EHR system only recorded absolute numbers (ie, 20 instead of 20.6). For all filled VLs (VLs where the date matched in both the paper file and the EHR), we counted them as a match if the standardized VL numbers in both records matched. VL matches were regarded as “high quality” if 85% of the records matched in both the paper file and the EHR.

Subsequently, we performed the same analysis for the data broken down by individual health care facilities, allowing for the analysis of potential factors affecting data quality. First, we examined the correlations between the HF characteristics and the resulting data quality. Facility characteristics were obtained from the user survey collected at the same time as this data quality study [4]. We hypothesized that the relevant factors would be (1) whether the HF was an intervention or control HF, (2) the average time (years) that the clinical users and data managers had been working with the EHR, (3) the relative technology experience of the users, and (4) how available and stable the EHR was. We hypothesized that being an intervention HF, the length of time of working with the EHR, EHR uptime, and technical experience would be positively correlated with better data quality.

To measure overall data quality, we developed a concordance score based on the work of Muthee et al [14] who compared data quality in paper and EHR records in Kenya. We narrowed the number of variables to 15 by considering only the most recent VL and drug pickup results and by dropping the variables for the health center sector, district, and facility, as these were not patient-specific data. The dropped variables were highly correlated or had little data entered in the paper file or EMR. The concordance score was calculated per record as the number of variables that exactly matched in both the paper file and the EHR after the values had been normalized, following the approach of Muthee et al [14]. This differed from our above mentioned analysis in that if the data were missing in both the paper file and the EHR, the records were counted as concordant.

Therefore, the concordance score here is a number from 0 to 15, which is the count of the number of concordant variables in a single record.

The randomization list of the control and intervention arm HFs was used to label the HFs in this study as intervention or control. For the number of years of experience with the EHR, we used the arithmetic mean of the number of years the clinical users and data managers at each HF had been using the EHR (if there was >1 individual). Technology experience was determined based on answers to the survey regarding “outside of work” use of technology, which included how frequently they (1) sent SMS text messages, (2) used the internet on a mobile phone, (3) used a computer, or (4) accessed the internet. For each question, they received 1 point for each time they said they used the technology either “most of the time” or “always,” resulting in a score from 0 to 4.

### Statistical Analysis

The study metrics are summarized in Textbox 1. We calculated the correlation between the concordance score and each HF characteristic using the Pearson correlation coefficient with Bonferroni correction applied to adjust for multiple comparisons. Subsequently, we sought evidence for the hypothesis that being an intervention site in the RCT would have a positive effect on the recording of VLs or drug dispensing episodes in the EHR. VL results and drug regimens were standardized as described in the Data Processing and Analysis section (due to different ways of recording the same data) and then counted as a match when the data were not missing and matched exactly. We could only compare data from after the intervention had been rolled out in July 2018. We then evaluated whether there was any association between the variables using a simple chi-square test to determine whether there was a correlation and using a Pearson contingency coefficient to estimate the effect size (for simple data such as a 2x2 contingency square, Pearson contingency is equivalent to both the phi coefficient and Cramer V). Data analyses were performed using R software (R Foundation for Statistical Computing).

### Textbox 1. Metrics defined in the study.

<table>
<thead>
<tr>
<th>Metric and definition</th>
<th>Data completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of fields filled divided by all fields</td>
</tr>
<tr>
<td></td>
<td>Data matching</td>
</tr>
<tr>
<td></td>
<td>Number of fields in the electronic health records that match those in the paper records divided by the number of filled fields in the paper records</td>
</tr>
<tr>
<td>Concordance scores</td>
<td>Number of variables among the 15 selected variables in which the electronic health records and paper records match in each medical record (including empty fields in both record types; Data matches and concordance scores were based on data that had been normalized as described in the Methods section)</td>
</tr>
<tr>
<td>Impact of hospital facility factors on concordance scores</td>
<td>The comparison of concordance scores in hospital facilities with and without the hospital facility characteristic (eg, high electronic health record uptime)</td>
</tr>
</tbody>
</table>
Ethical Considerations
The study was approved by 3 institutional review boards: Rwanda National Ethics Committee, Kigali (913/RNEC/2016), and the University of Leeds School of Medicine Research Ethics Committee, Leeds, United Kingdom (MREC16-176). This study was also reviewed in accordance with the US CDC human research protection procedures (approval no CGH HSR 2014-270a) and determined to be research. However, CDC investigators did not interact with human subjects or have access to identifiable data during this study.

Results
Overview
A total of 3467 records were reviewed in both the paper charts and the EHRs in the 50 HFs selected, with a mean of 69.3 (SD 8.8) records reviewed per HF. A total of 194,152 data items were collected using the ODK-based study tool. Three metrics of data quality were applied: (1) data completeness of EHR and paper records, (2) matching of the values of individual data types between EHR and paper records, and (3) concordance scores of matches for 15 regularly collected data items between EHR and paper records.

Table 1 presents all 28 variables collected from the paper and EHR records and the percentage completeness of each category. Few patients were on second-line drugs and possibly 1 was on third-line drugs (shown in the EHR only) therefore these 4 fields were not included in further analysis. Notably, paper VL records were 20% to 30% more complete than EHR records.

Table 2 presents the matching of variables between the paper records and the EHR records. Most nondate variables had fairly high levels of matching, although 5 were between 82% and 85%, just below the threshold set for high quality. Similar to the data completeness results, VL values had lower matching rates (59.5% for the most recent value and 65.8% for the oldest value). Few patients were prescribed second-line or third-line drug regimens (and their matching was low).

Additional analysis of date matches showed that for single-entry dates, such as “date of enrollment in HIV care,” matches were fairly high (60.2%-75%) compared to repeated dates, such as “date of last visit” or drug pickup date. For the “one-off” dates, most errors appeared to be mistakes in transcription (eg, small date differences and single character errors). Detailed analysis of the degree of error in these date fields showed that many incorrect one-off dates, such as “date of enrollment in HIV care,” were within 1 month of the correct value 34% to 38% of the time. Conversely, for regularly updated dates, only 10% to 13% of the errors were within 1 month of the correct value or 29% to 34% of the errors were within 3 months, suggesting that these items were not updated frequently in the EHRs.
Table 1. Completeness of all 28 variables\(^a\) in the paper records and electronic health records (EHRs) and the identification of variables included in the concordance score (N=3467).

<table>
<thead>
<tr>
<th>Number</th>
<th>Variable name</th>
<th>Variable included in concordance score</th>
<th>Completeness in paper records, n (%)</th>
<th>Completeness in EHR records, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Date of birth</td>
<td>Yes</td>
<td>3465 (99.94)</td>
<td>3466 (99.97)</td>
</tr>
<tr>
<td>2</td>
<td>Gender</td>
<td>Yes</td>
<td>3441 (99.25)</td>
<td>3467 (100)</td>
</tr>
<tr>
<td>3</td>
<td>Health center district</td>
<td>No</td>
<td>3467 (100)</td>
<td>3467 (100)</td>
</tr>
<tr>
<td>4</td>
<td>Health center sector</td>
<td>No</td>
<td>3467 (100)</td>
<td>3467 (100)</td>
</tr>
<tr>
<td>5</td>
<td>Health center name</td>
<td>No</td>
<td>3467 (100)</td>
<td>3467 (100)</td>
</tr>
<tr>
<td>6</td>
<td>Date of last visit</td>
<td>Yes</td>
<td>3464 (99.91)</td>
<td>3465 (99.94)</td>
</tr>
<tr>
<td>7</td>
<td>Date of first HIV-positive test</td>
<td>Yes</td>
<td>2711 (78.19)</td>
<td>2273 (65.56)</td>
</tr>
<tr>
<td>8</td>
<td>Date of enrollment in HIV care</td>
<td>Yes</td>
<td>3456 (99.68)</td>
<td>3457 (99.71)</td>
</tr>
<tr>
<td>9</td>
<td>Date of first prescribed ARTs(^b)</td>
<td>Yes</td>
<td>3463 (99.88)</td>
<td>3383 (97.58)</td>
</tr>
<tr>
<td>10</td>
<td>Initial WHO(^c) stage</td>
<td>Yes</td>
<td>3433 (99.02)</td>
<td>3267 (94.23)</td>
</tr>
<tr>
<td>11</td>
<td>Regimen 1 start date</td>
<td>Yes</td>
<td>3404 (98.18)</td>
<td>3385 (97.63)</td>
</tr>
<tr>
<td>12</td>
<td>Regimen 1 drugs</td>
<td>Yes</td>
<td>3411 (98.38)</td>
<td>3311 (95.50)</td>
</tr>
<tr>
<td>13</td>
<td>Regimen 2 start date(^d)</td>
<td>Yes</td>
<td>135 (3.89)</td>
<td>86 (2.48)</td>
</tr>
<tr>
<td>14</td>
<td>Regimen 2 drugs(^d)</td>
<td>Yes</td>
<td>136 (3.92)</td>
<td>86 (2.48)</td>
</tr>
<tr>
<td>15</td>
<td>Regimen 3 start date(^d)</td>
<td>No</td>
<td>0 (0)</td>
<td>1 (0.03)</td>
</tr>
<tr>
<td>16</td>
<td>Regimen 3 drugs(^d)</td>
<td>No</td>
<td>0 (0)</td>
<td>1 (0.03)</td>
</tr>
<tr>
<td>17</td>
<td>Drug pickup 1 date</td>
<td>Yes</td>
<td>3466 (99.97)</td>
<td>3465 (99.94)</td>
</tr>
<tr>
<td>18</td>
<td>Drug pickup 1 drugs</td>
<td>Yes</td>
<td>3466 (99.97)</td>
<td>3312 (95.53)</td>
</tr>
<tr>
<td>19</td>
<td>Drug pickup 2 date</td>
<td>No</td>
<td>3433 (99.02)</td>
<td>3385 (97.63)</td>
</tr>
<tr>
<td>20</td>
<td>Drug pickup 2 drugs</td>
<td>No</td>
<td>3434 (99.05)</td>
<td>3233 (93.25)</td>
</tr>
<tr>
<td>21</td>
<td>Drug pickup 3 date</td>
<td>No</td>
<td>3395 (97.92)</td>
<td>3267 (94.23)</td>
</tr>
<tr>
<td>22</td>
<td>Drug pickup 3 drugs</td>
<td>No</td>
<td>3397 (97.98)</td>
<td>3123 (90.08)</td>
</tr>
<tr>
<td>23</td>
<td>Viral load 1 date</td>
<td>Yes</td>
<td>3280 (94.61)</td>
<td>2645 (76.29)</td>
</tr>
<tr>
<td>24</td>
<td>Viral load 1 value</td>
<td>Yes</td>
<td>3284 (94.72)</td>
<td>2649 (76.4)</td>
</tr>
<tr>
<td>25</td>
<td>Viral load 2 date</td>
<td>No</td>
<td>2951 (85.12)</td>
<td>1773 (51.14)</td>
</tr>
<tr>
<td>26</td>
<td>Viral load 2 value</td>
<td>No</td>
<td>2955 (85.23)</td>
<td>1775 (51.19)</td>
</tr>
<tr>
<td>27</td>
<td>Viral load 3 date</td>
<td>No</td>
<td>2382 (68.7)</td>
<td>1116 (32.19)</td>
</tr>
<tr>
<td>28</td>
<td>Viral load 3 value</td>
<td>No</td>
<td>2386 (68.82)</td>
<td>1118 (32.25)</td>
</tr>
</tbody>
</table>

\(^a\)All variables met the 85% threshold except date of first HIV-positive test, viral load data, and second-line and third-line drug regimen data.

\(^b\)ART: antiretroviral therapy.

\(^c\)WHO: World Health Organization.

\(^d\)Very few participants had the records for these variables.
Table 2. Matching of the variables between the paper records and the electronic health records (EHRs; N=3467). Several variables were just below the 85% threshold.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Complete in paper record and EHR, n (%)</th>
<th>Matches, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of birth</td>
<td>3464 (99.91)</td>
<td>3242 (93.59)</td>
</tr>
<tr>
<td>Gender</td>
<td>3441 (99.25)</td>
<td>3326 (96.66)</td>
</tr>
<tr>
<td>Health center district</td>
<td>3467 (100)</td>
<td>3452 (99.97)</td>
</tr>
<tr>
<td>Health center sector</td>
<td>3467 (100)</td>
<td>3330 (98.05)</td>
</tr>
<tr>
<td>Health center name</td>
<td>3467 (100)</td>
<td>3467 (100)</td>
</tr>
<tr>
<td>Date of last visit</td>
<td>3462 (99.86)</td>
<td>979 (28.28)</td>
</tr>
<tr>
<td>Date of first HIV-positive test</td>
<td>1830 (52.78)</td>
<td>1102 (60.22)</td>
</tr>
<tr>
<td>Date of enrollment in HIV care</td>
<td>3446 (99.39)</td>
<td>2583 (74.96)</td>
</tr>
<tr>
<td>Date first prescribed ARTa</td>
<td>3379 (97.46)</td>
<td>2443 (72.3)</td>
</tr>
<tr>
<td>Initial WHOb stage</td>
<td>3236 (93.34)</td>
<td>2740 (84.67)</td>
</tr>
<tr>
<td>Regimen 1 start date</td>
<td>3322 (95.82)</td>
<td>2382 (71.71)</td>
</tr>
<tr>
<td>Regimen 1 drugs</td>
<td>3256 (93.91)</td>
<td>2673 (82.09)</td>
</tr>
<tr>
<td>Regimen 2 start date</td>
<td>78 (2.25)</td>
<td>42 (53.85)</td>
</tr>
<tr>
<td>Regimen 2 drugsc</td>
<td>78 (2.25)</td>
<td>37 (47.44)</td>
</tr>
<tr>
<td>Regimen 3 start date</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Regimen 3 drugs</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Drug pickup 1 date</td>
<td>3464 (99.91)</td>
<td>986 (28.46)</td>
</tr>
<tr>
<td>Drug pickup 1 drugs</td>
<td>3311 (95.5)</td>
<td>2772 (83.72)</td>
</tr>
<tr>
<td>Drug pickup 2 date</td>
<td>3362 (97)</td>
<td>698 (20.76)</td>
</tr>
<tr>
<td>Drug pickup 2 drugs</td>
<td>3212 (92.64)</td>
<td>2697 (83.97)</td>
</tr>
<tr>
<td>Drug pickup 3 date</td>
<td>3220 (92.88)</td>
<td>541 (16.8)</td>
</tr>
<tr>
<td>Drug pickup 3 drugs</td>
<td>3078 (88.78)</td>
<td>2583 (83.92)</td>
</tr>
<tr>
<td>Viral load 1 date</td>
<td>2614 (75.4)</td>
<td>847 (32.4)</td>
</tr>
<tr>
<td>Viral load 1 value</td>
<td>2622 (75.63)</td>
<td>1559 (59.46)</td>
</tr>
<tr>
<td>Viral load 2 date</td>
<td>1719 (49.58)</td>
<td>547 (31.82)</td>
</tr>
<tr>
<td>Viral load 2 value</td>
<td>1725 (49.75)</td>
<td>1095 (63.48)</td>
</tr>
<tr>
<td>Viral load 3 date</td>
<td>1033 (29.8)</td>
<td>322 (11.17)</td>
</tr>
<tr>
<td>Viral load 3 value</td>
<td>1038 (29.94)</td>
<td>683 (21.58)</td>
</tr>
</tbody>
</table>

aART: antiretroviral therapy.
bWHO: World Health Organization.
cThe only matching rate below 59% that was not a date was for regimen 2 drugs that had very little data (n=78).

Concordance Scores

The mean overall concordance score was 10.2 (SD 1.28) for 15 (68%) variables in all 50 HFs. Only 2 HFs were at or above the threshold of 85% concordance. However, 8 of the 15 variables in the concordance score (not including the date of birth) were dates (refer to Table 1 for the included variables). The concordance score would be higher if dates that match within 1 month were included.

Evaluation of Factors Associated With Concordance Scores in Each HF

Correlations between the concordance scores and most HF characteristics were not significant. For the survey response that “the EHR was available always or nearly always,” there was a positive correlation with the concordance score, with $r=0.16$ (95% CI 0.13–0.19; $P<.001$; Figure 1). There was also a significant (negative) relationship between user responses suggesting that they had higher technology experience, and concordance scores with $r=-0.11$ (95% CI –0.15 to 0.08; $P<.001$). There was no significant correlation between the concordance scores and users’ reported years of EHR use (Table 3).
Figure 1. Correlation between concordance scores and the reported electronic health record (EHR) availability or uptime (ranging from 1 to 5 Likert scale responses of EHR users).

Table 3. Association of electronic health record (EHR) performance and use characteristics with concordance scores.

<table>
<thead>
<tr>
<th>Parameter 1</th>
<th>Parameter 2</th>
<th>r (95% CI)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance score</td>
<td>EHR availability</td>
<td>0.16 (0.13 to 0.19)</td>
<td>9.37 (3326)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Concordance score</td>
<td>Technology experience</td>
<td>−0.11 (−0.15 to −0.08)</td>
<td>−6.64 (3326)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Concordance score</td>
<td>Years of EHR use</td>
<td>−0.04 (−0.07 to 0.00)</td>
<td>−2.10 (3326)</td>
<td>&lt;.21</td>
</tr>
</tbody>
</table>

Effects of Alerts in the Intervention HFs on Matching of Key Variables in EHRs and Paper Records

As the intervention was implemented in July 2018 and the data for the data quality study were collected in November and December 2018, there was a 4-month to 5-month window when the alerts were active. All intervention HFs had alerts for starting a new treatment, but few new patients were enrolled in the 5-month time frame. Therefore, there was insufficient data to study the effects of alert 1. However, there was sufficient data to compare the data quality metrics of the 11 intervention HFs that had alerts for missing VL results (alert 2) with the 39 HFs without these alerts. Figure 2 shows the concordance scores for each site color coded by intervention status. Figure 3 presents the effect of the VL alerts on overall data matches between the EHRs and the paper records, ranked by percentage matching of variables between the 2 sources. The HFs with the VL alerts (orange color) had a higher rank overall than the control HFs.

In HFs with alerts, 85 (26.7%) of 318 VL records in the EHR matched those in the paper records, but in HFs without alerts, only 120 (11.9%) of 1008 VL records matched (P<.001, chi-square test).

Furthermore, we compared the number of drug pickups in the EHRs that exactly matched with those in the paper file records, comparing the intervention HFs with VL alerts with HFs without such alerts. The results are shown in Figure 4. In HFs with alerts, 280 (36.2%) of 774 of the records of drug pickups matched, but in HFs without alerts, only 183 (18.7%) of 981 matched (P<.001). A second analysis using the chi-square test and Pearson contingency coefficient for the 25 HFs revealed χ² =67.5 (P<.001), indicating a strong effect (φ=0.20, 95% CI
0.15-0.24). Despite the significantly higher data quality at the intervention HF s, not all HF s showed this effect.

**Figure 2.** Mean concordance scores for all health facilities color coded by intervention status (maximum score=15).

**Figure 3.** Graph showing percentage data matches for viral load results between paper records and electronic health records for each of the 25 health care centers in the intervention arm, ranked by percentage of matches.
Figure 4. Graph showing percentage data matches for drug pickups between paper records and electronic health records among the same 25 intervention health facilities (HFs) as in Figure 3. The overall scores were higher than in Figure 3 but only 1 (intervention) HF score was more than the data quality threshold.

Discussion

Principal Findings

The results showed that the completeness of the paper records and EHR records was generally >85% threshold for good quality, with the exception of VL data collected in the 2 years before the study data collection (VL results 2 and 3). The completeness varied across the data items, with the best results (>99%) for EHR data seen in demographic data, date of enrollment in antiretroviral (ARV) care, date of first being prescribed ARV drugs, and date of last visit. The lowest completeness was seen in the second-line and third-line drug regimens, but very few patients were prescribed these regimens. There was clear evidence that the electronic records were not being updated in a complete and timely manner for some data items at most HFs (based on concordance scores), with only 2 HFs meeting the 85% threshold for good data quality. Furthermore, there was evidence that certain HFs were more effective in data entry and management. Using data from the previously published user experience survey by Fraser et al [4], we compared the measures of user experience with the EHRs in the same HFs included in this study. Facilities reported to have more consistent uptime and availability of EHRs showed significantly better data quality based on the concordance scores. However, HFs with higher “technology experience” scores had lower concordance scores, the reason for which is unclear.

A key scientific question for the overarching study was to determine whether the enhanced EHR in the intervention HFs improved both the usability and use of the EHR and the completeness and overall quality of key data items required to support effective HIV care. The second decision support alert (refer to the Methods section) was designed to warn if a VL result had not been uploaded to the system by 6 months after starting ARV treatment or by 12 months after the last VL result (allowing for up to 2 months transport and data entry delays). There was a 4-month to 5-month period with the intervention in place before we collected data for this data quality study, allowing the assessment of whether VL data were more complete in the EHRs from the 11 HFs that received this alert. Although data completeness and timeliness for VL data were overall <85% of the threshold, with only 11.9% completeness in the 39 control HFs lacking the VL alert, completeness was significantly higher (26.7%) in the 11 intervention HFs that had VL alerts. A similar effect was seen for drug pickup data, suggesting that the presence of VL alerts impacted the overall data entry processes or behavior and not just the specific data type in the alert. It should be noted that VL testing was infrequent before the start of the study, as the laboratory capacity was still being established, which may have affected data management and data entry in the EHRs, especially for older values. The metrics for overall data quality and completeness may have been higher if there had been regular data quality improvement visits to Rwandan MOH HFs, an effect seen in the study by Muthee et al [14] in Kenya. Another strategy for improving data quality and user experience is the implementation of EHR systems, allowing direct point-of-care data entry rather than entering data on paper and in the EHR, as observed for VL results in Kenya [32]. The findings in this study also indicate
that data quality and timeliness would have likely been improved more generally using automated tools to alert staff of missing data for key variables such as VL results, using form field validation to detect impossible values, and comparing date values with other data. The analysis of the degree of error in date fields showed that many incorrect one-off dates, such as “date of enrollment in HIV care,” were within 1 month of the correct values for approximately one-third of the times and therefore may have been adequate for the care of HIV patients seen every 2 to 3 months. Incorrect repeated dates such as “last clinic visit,” “drug pickup,” and “VL” were within 1 month of the correct date, ≤12% of the times, and therefore may be less fit for use in patient care. Overall, most data were considered fit for purpose, as confirmed by the user survey [4], which showed a reliance on the EHR data for clinical care and reporting. Regular monitoring of data quality metrics and intervention at poorly performing HFs would likely help address major deficiencies efficiently.

The implications of these results are much broader than just HIV care, which provided the environment for this study. Primary care EHR systems remain rare in LMICs, but recent progress with facility-based EHR systems, such as OpenMRS in Haiti, Sierra Leone, parts of India, and the Philippines, are laying the foundations for wider use [33]. Mobile-based and tablet-based systems that support care in communities and some HFs should provide a much wider range of data for clinical and public health use. The World Health Organization (WHO)—led Open Smart Register Project [34] has 150 million patients now, the majority being in Bangladesh. Other mHealth systems, such as CommCare [35] and ODK [31], are used in many LMICs [36]. All these systems have the ability to share data with EHRs, such as OpenMRS. They increasingly support interoperability standards such as HL7-FHIR [27] for interoperability with a range of other health information systems, including laboratory and pharmacy systems, radiology information systems, and the District Health Information System-2 [37]. FHIR-based Application Programming Interfaces and toolkits are also being built into the Apple OS and Android operating systems [36]. Projects using data science techniques and machine learning are increasing in LMICs and require high-quality data sets that are representative of all communities and groups to avoid biases or poor performance in vulnerable groups [38].

Comparison to Previous Work

Muthee et al [14] studied data quality in 53 HFs running the KenyaEMR OpenMRS-based system. They analyzed data from a data quality assessment (DQA), which had been collected by MOH lead teams that visited the HFs. They found ≥1 missing value in 735 (31%) paper forms and 747 (32%) KenyaEMR records, a higher missing rate than that in this study. Concordance scores were calculated in the same manner as in this Rwanda study, with a mean concordance of 11.9 (SD4.0) for 20 variables at baseline. Furthermore, they were able to compare the DQA results with a second DQA assessment of 27 HFs. They showed that HFs with previous DQA had significantly better data quality on the second visit, with missing data falling to 13% and concordance scores increasing to a mean of 13.6 (SD 4.2, difference in concordance scores P<.02). This result suggests that the process of measuring data quality and providing feedback to staff can have long-term benefits. The mean concordance scores of 13.6 (SD 4.2) for 20 variables (68%) at the second DQA were equivalent to those in Rwanda, which was 10.2 (SD 1.28) for 15 variables (68%). Abiy et al [39] studied data quality in an HIV treatment program in Ethiopia. They compared the paper and EHR versions of 250 records. Data completeness was 78% (95% CI 70.8%-85.1%) for paper records and 76% (95% CI 67.8%-83.2%) for the EHR. They used the κ statistic to compare variables in both records, with results ranging from κ=0.93 for demographics, κ=0.86 for WHO stage, and κ=0.83 for “general appearance.” Furthermore, Ngugi et al [40] studied the clinical use and completeness of data entry in KenyaEMR at 219 HFs in Kenya. They showed a wide variation in data entry per month in different HFs and noted that this was likely affected by “patients’ volume, frequency of patients’ visits (encounters), EHRs mode of use, and active use of the system during care.” Haskew et al [41] studied the quality of data collection and the effects of clinical alerts on HIV patient care in Western Kenya before and after the cloud-based implementation of an OpenMRS EHR. They showed that clinical alerts reduced missing data and improved the quality of care. Missing data were significantly reduced for key variables (before and after alerts implemented) including “patient source” (375 to 69 patients), “first CD4 count” (826 to 35 patients), and “first WHO stage” (2258 to 479 patients; all P<.001). The number of patients eligible for ART (based on CD4 count and WHO stage) but not yet started receiving it, fell from 1346 (29.6%) to 270 (6.2%; P<.001) [41].

A study of the health management information system in 16 health centers, 28 health posts, and 1 hospital in Ethiopia showed that training staff to fill out the data forms was significantly associated with improved data quality (odds ratio [OR] 2.253, 95% CI 1.082-4.692), and measures showing effective supervision and leadership were also beneficial [42]. A study of HIV clinical data collection systems at 21 HFs (18 in Africa) showed a reduction in missing data associated with training in data management (OR 0.58, 95% CI 0.37-0.90) and weekly hours spent by a data clerk working on the data collection system (OR 0.95, 95% CI 0.90-0.99) [16]. Previous studies suggest that poor completion or accuracy of key data items required for clinical care is likely to limit the benefit of EHR systems in day-to-day patient management, such as the functioning of clinical decision support system in this study and that reported by Haskew et al [41]. Furthermore, it impacts quality of care metrics (eg, in cancer care in the United States [43]) and efficient reporting of data for clinical teams, MOH, and funders [16,42].

Limitations

The study was limited to a subset of the MOH-run OpenMRS EHR systems for use in HIV care. The inclusion criteria for the 112 HFs in the larger study favored HFs with better information technology hardware and successful implementation of a server monitoring tool [4]. Although the intervention HFs appeared to show improved data quality and completeness for certain variables, this effect might change over time. Although the user survey showed that staff in the intervention and control HFs received similar amounts of training and were equally positive about their training, it is possible that differences in the type of...
training in the intervention HFs and emphasis on the importance of VL contributed to better data quality. More robust and longitudinal studies are needed to understand the effects of different technologies on the use of EHR, including variations within and across HFs and individuals.

Conclusions

The study found that there was generally high data completeness in both EHR and paper records in Rwandan MOH-supported HFs. The percentage matching and concordance scores for the EHR and paper records were lower, but this mainly affected the recording of dates. The considerable potential benefits of EHR systems for patient care for HIV and other diseases will only be realized with sufficient support, training and monitoring of data quality, and adequate technical support and infrastructure to ensure reliable systems. The use of “point-of-care” EHR systems rather than transcription of data from paper records can also improve user experience and data completeness. The results of this study show that automated alerts regarding data quality, completeness, and timeliness can significantly improve these metrics in remote HFs in a low-income country. However, to be fully effective, these alerts should be combined with other strategies to ensure high system uptime and a range of data quality improvement strategies, including training, regular supervision, and feedback. These improvements will likely help drive the much wider use of EHR data in LMICs for clinical care, reporting, data science, and machine learning. Furthermore, the data collected will support population and public health uses, including syndromic surveillance, notifiable disease reporting, real-time monitoring of disease burden, and forecasting resource requirements to meet care needs.

Acknowledgments

The authors would like to thank the clinic staff at the study health facilities in Rwanda for their participation in the data quality study, the staff of the Rwanda Biomedical Center and the Ministry of Health, and the informatics team at Partners in Health in Rwanda for technical assistance and advice. The authors would like to thank Dr Eric Green (Duke University) for assisting with the randomized controlled trial design and randomization. They would also like to thank Henry “Roo” Fraser for assisting with the creation of diagrams.

This research was supported by the President’s Emergency Plan for AIDS Relief through the Centers for Disease Control and Prevention under the terms of program grant U01 GH000782-01 awarded to the National University of Rwanda. Additional funding and in-kind support were provided by Brown University, Providence, Rhode Island, United States. HSFF was a Marie Skłodowska-Curie Fellow from 2015 to 2017, funded by the European Union’s Horizon 2020 Research and Innovation Program under grant 661289, “Global eHealth.”

Data Availability

The data sets generated and analyzed in this study are available from the corresponding author upon reasonable request.

Disclaimer

This program was supported by the President’s Emergency Plan for AIDS Relief through the Centers for Disease Control and Prevention. The findings and conclusions of this study do not necessarily represent the official positions of the funding agencies.

Conflicts of Interest

HSFF and CS are cofounders of the Open MRS EHR project. JR and XS worked for the Centers for Disease Control and Prevention Division of Global HIV and tuberculosis, which funded this study. None of these individuals received compensation for their time from the Centers for Disease Control and Prevention implementation science grant. HSSF received travel and accommodation support for study meetings and workshops from the grant. CS is the director of Jembi Health Systems and received funds from a grant for the software development work on OpenMRS. The findings and conclusions in this report are those of the authors and do not necessarily represent the official positions of the funding agencies. All other authors declare no other conflicts of interest.

References

1. HIV/AIDS. World Health Organization. 2023 Jan 01. URL: https://www.afro.who.int/health-topics/hivaids [accessed 2023-02-15]


34. OpenSRP. URL: https://opensrp.io/ [accessed 2024-04-02]

35. Dimagi. URL: https://www.dimagi.com [accessed 2024-04-02]


37. Integration and interoperability with DHIS2. DHIS2. URL: https://dhis2.org/integration/ [accessed 2024-04-02]


Abbreviations

ART: antiretroviral therapy  
CDC: Centers for Disease Control and Prevention  
DQA: data quality assessment  
EHR: electronic health record  
EMR: electronic medical record  
HF: health facility  
HL7 FHIR: Health Level-7 Fast Healthcare Interoperability Resources  
IMB: Inshuti Mu Buzima  
LMICs: low- and middle-income countries
A Novel Web-Based Application for Influenza and COVID-19 Outbreak Detection and Response in Residential Aged Care Facilities

Kai Hsun Hsiao¹, MBChB, MPH, PGDipTMH; Emma Quinn¹,², BSc, MPH, DrPH; Travers Johnstone¹, BSW, MPH; Maria Gomez¹, BCom, MPH; Andrew Ingleton¹, BN, MIPH; Arun Parasuraman¹, BVSc, BN, MPH, MIPH; Zeina Najjar¹, BSc, BMed, BSurg, MPH; Leena Gupta¹,², MBBS, MPH, DrPH

¹ 2

Corresponding Author:
Kai Hsun Hsiao, MBChB, MPH, PGDipTMH

Abstract

The use of innovative digital health technologies in public health is expanding quickly, including the use of these tools in outbreak response. The translation of a digital health innovation into effective public health practice is a complex process requiring diverse enablers across the people, process, and technology domains. This paper describes a novel web-based application that was designed and implemented by a district-level public health authority to assist residential aged care facilities in influenza and COVID-19 outbreak detection and response. It discusses some of the challenges, enablers, and key lessons learned in designing and implementing such a novel application from the perspectives of the public health practitioners (the authors) that undertook this project.

(JMIR Public Health Surveill 2024;10:e37625) doi:10.2196/37625

KEYWORDS

web application; digital health; communicable disease control; outbreak; surveillance; influenza; aged care; aged care homes

Introduction

Digital health technologies such as web- and mobile-based applications have proliferated over the last 2 decades [1]. Historically, these applications have been used in communicable disease surveillance and control [1-3]. However, the public health field is beleaguered by tools that are short-lived, duplicative, disconnected, of poor quality, or not suited to user needs [2,4,5]. There is also growing evidence that there are consistent barriers to effective implementation and adoption of digital health technologies [4]. Some of these include change management, stakeholder acceptance, training, supporting infrastructure, sustainable financial and human resources, and governance for quality and safety [1,4]. It is therefore critical for public health practitioners to recognize the broader set of necessary factors and considerations when planning, designing, and implementing innovative digital health interventions.

In 2017, the Sydney Local Health District (SLHD) Public Health Unit (PHU), a district-level public health authority, embarked on a digital health innovation project to design, develop, and implement a web-based application—the Influenza Outbreak Communication, Advice and Reporting (FluCARE) application—to assist staff in residential aged care facilities (RACFs) in the timely detection and response to influenza outbreaks. Early outbreak notification to the PHU has been shown to reduce outbreak duration and impact [6].

The application was launched as a pilot across the district in late 2019. In 2020, in response to the COVID-19 pandemic, rapid adaptations were made to FluCARE to incorporate COVID-19 monitoring, supported by additional procedures, guides, and training modules. FluCARE continues to be used by RACFs across the district. The purpose of this paper is to describe the design, development, and implementation of the FluCARE application, as well as key challenges, enablers, and lessons learned from the perspective of the public health practitioners (the authors) who implemented the application.

Setting and Context

FluCARE was undertaken in the SLHD, which encompasses a metropolitan area of Sydney, Australia, with a population of 640,000, of which approximately 80,000 (12%) are aged 65 years and older [7]. Within the district, there are 62 RACFs, which vary in size, care levels, and ownership, with over 4700 operational bed places in total [8].

The PHU is responsible for local communicable disease control and outbreak management within the SLHD. This includes assisting RACFs in responding to influenza outbreaks, which remains a high priority for public health due to the potential for
significant health impacts on older adult residents [6]. While RACFs are not mandated to report influenza outbreaks to the PHU, it is recommended in national guidelines [9]. The national guidelines also provide definitions for suspected and confirmed influenza cases based on their symptoms and laboratory results, as well as the criteria for declaring a potential or confirmed outbreak. Upon the notification of an influenza outbreak, the PHU assists the RACF by advising on response actions and by monitoring outbreak progress through daily line lists of cases submitted by the RACF.

### Design and Development of FluCARE

An overview of our application design, development, and implementation activity is shown in Figure 1.

**Figure 1.** FluCARE project timeline demonstrating the relationships between stakeholder engagement, design and development, and implementation support activities. Staging refers to a prelaunch phase of FluCARE testing within the SLHD hosting environment. *Web-based training modules were offered through the Moodle platform. BRD: business requirements document; dev.: development; environ.: environment; FluCARE: Influenza Outbreak Communication, Advice and Reporting application; ICT: information and communication technology; M&E: monitoring and evaluation; MVP: minimum viable product; RACF: residential aged care facility; SLHD: Sydney Local Health District; SOP: standard operating procedure; UAT: user acceptability testing.*
Conceptual Design

Application scoping and conceptual design occupied a significant portion (8 months) of our project timeline due to the need to ensure that we had the appropriate type and level of functional capability included in the application. A variety of different stakeholder consultation sessions were held with RACF and PHU staff to understand design needs (see Table 1, item 6), and through the establishment and use of a working group, we were able to finalize and agree on a minimum viable product (MVP; see Figure 1).
Table. Description of the design and technical features of FluCARE as per the WHO Mobile Health Evidence Reporting and Assessment (mERA) criteria [10].

<table>
<thead>
<tr>
<th>Item</th>
<th>mERA criteria or feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Infrastructure</td>
<td>FluCARE is a web-based application. User access requires internet connectivity and a web browser, which are typically available in the PHU and the RACFs in the district. The application is optimized for desktop and laptop computers but can also be accessed by tablet and mobile devices. Users also require access to an individual email account and/or mobile phone to receive MFA codes. While smartphone device ownership in Australia is 91% [11], their use by RACF staff at work may be limited by workplace policies and/or signal strength. The FluCARE application and associated databases were initially hosted on a local server maintained within the SLHD secure network. At the time of writing this viewpoint, the application was being transitioned into the Microsoft Azure cloud environment.</td>
</tr>
<tr>
<td>2</td>
<td>Technology platform</td>
<td>FluCARE is an ASP.NET Core MVC web application written in the C# language with the Razor View engine and Telerik controls. Sensitive data are encrypted and held in a Microsoft SQL Server database. SendGrid is used for the delivery of email notifications and MFA codes. Amazon SNS is used for the delivery of SMS MFA codes. FluCARE is hosted on IIS in a Windows environment.</td>
</tr>
<tr>
<td>3</td>
<td>Interoperability</td>
<td>FluCARE is not directly connected to any other local or state health information system. However, relevant reports can be generated and exported as a CSV file, which can be configured for upload to the NSW Notifiable Conditions Information Management System where PHUs are required to report and document RACF influenza outbreaks.</td>
</tr>
<tr>
<td>4</td>
<td>Intervention delivery</td>
<td>RACFs register for and access the FluCARE application via the application web page. The PHU promotes RACF awareness, access, and use of FluCARE via the PHU’s annual workshop on winter preparedness for RACFs, regular email reminders, and occasional update teleconferences.</td>
</tr>
<tr>
<td>5</td>
<td>Intervention content</td>
<td>FluCARE content and processes are based on the national CDNA guidelines for the management of influenza outbreaks in RACFs [9]. Application content was developed, adapted, and reviewed by a multidisciplinary working group.</td>
</tr>
</tbody>
</table>
Design and development

End-user engagement and testing were conducted at multiple points throughout the design and development process (see Figure 1):

- Initial RACF staff consultation via an open forum discussion and structured survey, and key informant interviews with PHU staff on their surveillance needs informed the conceptual design of FluCARE.
- Preproduction wireframes were presented to PHU and RACF staff (both at an annual winter preparedness workshop and in facility-based meetings) to seek feedback on user-friendliness, aesthetics, and functions prior to MVP launch (see Figure 1).
- Postproduction UAT was conducted with RACF staff focus groups (not reported in detail here).

User feedback

UAT conducted with a focus group of RACF staff found the application design to be user-friendly and accessible, the content informative, and the functionalities useful and likely to save time and reduce workload. Most importantly, the UAT found that most RACF staff users would access and use the application on their desktop, rather than on mobile devices at the bedside, which led us to optimizing the application for desktop use.

Access of individual participants

Access to the FluCARE application requires a desktop computer or mobile phone and internet connection either at work or home.

Cost assessment

Economic evaluations of FluCARE are yet to be conducted.

Adoption inputs/program entry

PHU and RACF users are required to complete training on FluCARE via either a 2.5-hour face-to-face training session or a set of web-based modules. User manuals are available within FluCARE. Designated PHU staff provide users with email and telephone support from initial registration to operational use of FluCARE.

Limitations for delivery at scale

Application development will be necessary to allow FluCARE to cater for multiple districts or jurisdictions. Currently, further scale-up of FluCARE is limited by our district’s IT infrastructure and human resource capacity to support an expanded number of RACFs and users. Implementation research is needed to inform and adapt strategies to support FluCARE implementation to other LHDs or PHUs.

Contextual adaptability

FluCARE was built as a proof of concept with a focus on influenza outbreaks in RACFs in our district. However, we believe it would be adaptable for other communicable diseases, settings, districts, or jurisdictions. For example, FluCARE was rapidly redeveloped to incorporate COVID-19 outbreak monitoring in RACFs.

Replicability

The FluCARE application and supporting processes, such as user manuals, training modules, and standard operating procedures, can be replicated for introduction into new districts or jurisdictions.
<table>
<thead>
<tr>
<th>Item</th>
<th>mERA criteria or feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Data security</td>
<td>Application-level security includes ASP.NET-secured cookie authentication, a strong password policy, MFA for user access, autolocking accounts for excess failed log-ins, and inactivity timeout. Data are secured with 256-bit encryption. System servers are secured by firewalls. Procedural protections include manual verification prior to RACF account approval, user terms and conditions, and a user management policy based on a principle of least privilege.</td>
</tr>
<tr>
<td>15</td>
<td>Compliance</td>
<td>FluCARE is compliant with the NSW Health Records and Information Privacy Act 2002 [12] and relevant NSW Health governing standards, that is, the Health Care Records Policy Directive [13] and the Privacy Manual for Health Information [14]. This was verified by an independent security audit. The content, algorithms, and advice in FluCARE align with the national CDNA guidelines for the prevention, control, and public health management of influenza outbreaks in RACFs in Australia [9].</td>
</tr>
<tr>
<td>16</td>
<td>Fidelity</td>
<td>FluCARE underwent a pilot feasibility study, which demonstrated high acceptance, utility, and safety for RACF staff in the management of influenza and COVID-19 outbreaks [15].</td>
</tr>
</tbody>
</table>

FluCARE: Influenza Outbreak Communication, Advice and Reporting.
WHO: World Health Organization.
PHU: Public Health Unit.
RACF: residential aged care facility.
MFA: multifactor authentication.
SLHD: Sydney Local Health District.
MVC: Model-View-Controller.
SNS: Simple Notification Service.
IIS: Internet Information Services.
NSW: New South Wales.
CDNA: Communicable Diseases Network Australia.
MVP: minimum viable product.
UAT: user acceptability testing.
LHD: local health district.

Through the RACF stakeholder consultation sessions (see Figure 1), the PHU identified 3 key barriers to effective influenza surveillance and outbreak response and, with the working group, designed a series of features that addressed these barriers, as shown in Table 2 below. The working group membership included clinical and epidemiological staff from within the PHU, geriatric care managers and senior information and communication technology (ICT) staff from across the district, and managers of respiratory diseases from within Health Protection NSW. The group met weekly to advise on the components or features of the application and together produced a comprehensive business requirements document from which only the essential elements for an initial MVP build had to be identified and agreed on (see Figure 1).
Table. Residential aged care facility (RACF) barriers to influenza outbreak control, and associated key design features of FluCARE\(^a\) to improve outbreak management.

<table>
<thead>
<tr>
<th>Key barriers to influenza outbreak control</th>
<th>Key FluCARE application design features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emailed or faxed line listing</td>
<td>• Web-based line list and interface for data management (see Figure 2): The interface allows RACF staff to log onto their facility-specific account, add or update line list entries, and directly submit to the PHU(^b) via the application. The PHU can immediately view and manage all RACF line lists.</td>
</tr>
</tbody>
</table>
| Complicated case definitions and outbreak criteria | • Real-time outbreak detection algorithms: FluCARE analyzes the line list data to (1) highlight entries meeting case definitions for influenza-like illness or confirmed influenza and (2) detect when the criteria for a potential or confirmed influenza outbreak are met.  
• An outbreak action checklist: Tick-box list of key response actions that the RACF should take based on guidelines. |
| Need to notify multiple stakeholders of an outbreak in a timely fashion | • Automated stakeholder notifications once the outbreak criteria have been met: Upon detecting an outbreak situation, FluCARE sends email notifications to: the PHU, including to prompt verification; key RACF staff, such as the facility or infection control manager; and any additional notification recipients, such as residents’ general practitioners, as subscribed by the RACF.  
• Daily line list report: Enables the RACFs and the PHU to report on the progress and response to the outbreak to senior management. |

\(^a\)FluCARE: Influenza Outbreak Communication, Advice and Reporting application.  
\(^b\)PHU: Public Health Unit.
Figure 2. Screenshot of the residential aged care facility (RACF) dashboard and web-based line list interface in the Influenza Outbreak Communication, Advice and Reporting (FluCARE) application. This is the most recent version of FluCARE, with rapid add-on COVID-19 functionalities visible. (Displayed names are fictional). NSW: New South Wales.

Development

Once the MVP was developed, a comprehensive audit of the application’s content, algorithm, and functions was conducted, as well as external user acceptability testing (UAT) with a focus group of self-volunteered RACF staff (see Figure 1). The UAT was designed as a guided, hands-on training session on FluCARE followed by a structured feedback session with comments obtained via moderated group discussion. This testing identified minor issues and potential improvements, such as changes to interface workflow. The MVP then underwent further cycles of testing and enhancement prior to pilot launch. After the MVP launch, an independent security assessment verified that the technical and architectural design of the application met governing standards (Table 1, item 15 and Figure 1). Further to this audit and to support appropriate local procedures in how FluCARE data were managed, the PHU underwent an internal security risk assessment and developed a risk management plan with our IT services prior to launch. This was particularly important as the application transitioned from the digital developer’s environment to being hosted by the SLHD ICT local servers (see Figure 1).

Even after launch, the FluCARE application underwent further refinements over time in accordance with PHU or RACF feedback. The most significant enhancement was a rapid 6-week
development in March 2020 to incorporate COVID-19 monitoring and outbreak detection (see Figure 1). At the time of developing the FluCARE application, there was no standardized and validated way to report the features of a web-based application, so the features of FluCARE have been described as per the World Health Organization checklist on Mobile Health Evidence Reporting and Assessment criteria above (see Multimedia Appendix 1 [10]).

### Implementation Process

Preparations for application implementation occurred in parallel to the development process, as illustrated in Figure 1. The authors designed the implementation supports based on 2 reviews of the literature [16,17] focused on the implementation and adoption of health technologies by health care professionals, as shown in Figure 3.

**Figure 3.** Overview of the FluCARE digital health implementation package as informed by Cresswell and Sheikh [16] and Gagnon et al [17]. FluCARE: Influenza Outbreak Communication, Advice and Reporting application; ICT: information and communication technology; PHU: Public Health Unit.

Existing engagements with RACFs, such as the annual winter preparedness workshop, as well as FluCARE-specific outreach, including the various UAT episodes (Table 1, item 6), were used to raise awareness and user buy-in of FluCARE (Figure 3; personal and social enablers). A range of training resources and sessions (described in Table 1, item 10; see Figure 1) were developed by the working group to support PHU and RACF staff onboarding. Operationally, a suite of policies and
procedures were developed to govern the use of FluCARE as well as to manage issues typically introduced by digital technologies, such as user management, data security, and downtime contingency plans (see Figure 1, items 1 - 8 and Figure 3; organizational enablers).

Internal IT infrastructure and operational support from the SLHD ICT Services were also critical. The SLHD ICT Services built the requisite highly available and secure server environment to host FluCARE and continue to provide services for maintenance and issues resolution (see Figure 3; technical enablers).

For ongoing monitoring, quality improvement, and safety, the PHU established a registry (issue log) to prospectively record technical or operational issues and any improvement suggestions (see Figure 1). These are routinely reviewed by the working group for further updates to the application (technical enablers). Further evaluation of the implementation of FluCARE through a feasibility study has been published [15], and effectiveness studies are planned.

Discussion

Despite substantial challenges related to the redesign of the application during the COVID-19 response, the main lessons learned across the entire design, development, and implementation time frame of FluCARE (as shown in Figure 1) were the importance of user-based design activity, information security and privacy, and organizational support (such as funding and IT infrastructure).

User-Related Lesson: User Engagement Is Critical and Should Be Embedded in All Phases of Design, Development, and Implementation

User-centric design is a well-understood principle in IT projects [3,5,18-20]. From the outset, we were mindful that FluCARE’s capacity to improve outbreak outcomes was dependent on its ability to be accepted and adopted by RACF staff. This is typically determined by factors such as an application’s perceived usefulness, reliability, and user-friendliness [5,21]. As such, input from RACF staff was pivotal. Beyond informing the core functions of the application, feedback from RACF staff via the various user engagement and testing episodes embedded in our design and development process (see Figure 1; 0-27 months) was valuable in answering key design questions encountered by the working group and digital developers, for example, whether to optimize the application for desktop versus mobile use. Lastly, while acknowledging the importance of user-based design in the development of a digital health tool, no one design approach is accepted as a gold standard [22]. However, we did not conduct our user-based design activity through a theoretical framework or research process and would have benefited from such a systematic approach.

Technical Lesson: Data Security Should Be Addressed With a Multifaceted, Risk Management Approach

Data security and information privacy is a central concern that must be thoroughly addressed to ensure user acceptability, and that ethical and professional obligations are met, particularly when dealing with personally identifiable health data [5,23]. Addressing all facets of security appropriately required not only an understanding of the legislation, policies, and procedures in place to secure and manage health data appropriately within New South Wales but also a practical risk management approach in order to mitigate any perceived risks to data breaches. Based on the recommendations from the independent security assessment, the working group developed a user access management policy and security risk management plan to help mitigate any inappropriate access or use of the application and to ensure the security of the personal health information within the application (Figure 1; 15 - 27 months). While much of the existing literature focuses on the technical aspects of security, it is important to recognize that data governance strategies, staff training, and incident detection and response plans are also essential for security [24-26]. This multifaceted security approach was applied for FluCARE, as detailed in Table 1 (item 14).

Organizational Lesson: Organization-Wide Support for Digital Innovation Should Be Fostered and Promoted

Supportive organizational culture and structures for digital health innovation are important enablers for successful design and implementation [5,27]. FluCARE development was undertaken within a forward-thinking organization where innovation in health care IT is a strategic priority [28]. The project received high-level support at initiation through the SLHD executive as well as at the PHU and IT departmental levels, which facilitated supportive resourcing from the IT department, including their de novo build of the required hosting environment for FluCARE. However, we may have potentially benefited from an organization-wide mechanism for pooling institutional experience and sharing resources across different projects, given that a number of other application developments were also occurring in the SLHD over the time period of the design, development, and implementation of FluCARE (see Figure 1). This has potential for leveraging economies of scale, reducing duplicative overhead costs, optimizing interoperability, and improving procedural efficiencies [5,29].

Conclusion

The translation of an innovative application idea into an effective public health tool is a multifaceted process, which importantly includes early and ongoing user engagement, considered but practical data governance and security measures, organizational support, and dedicated resources and IT infrastructure to support maintenance over time. From our experience, these elements are essential to ensure user-friendly and security-conscious design as well as successful application implementation and uptake.
Acknowledgments

The initial development of the Influenza Outbreak Communication, Advice and Reporting (FluCARE) application was funded by winnings from the Sydney Local Health District (SLHD) “The Pitch” innovation challenge. This project was also supported by a grant to Sydney Health Partners (MR9100001) from the Australian Government’s Medical Research Future Fund (MRFF) as part of the Rapid Applied Research Translation program. We also thank the SLHD Information and Communication Technology (ICT) Services and FAQ Interactive for their technical contributions in the development and implementation of the FluCARE application.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Checklist for Mobile Health (mHealth) Evidence Reporting and Assessment (mERA) guidelines, including mHealth essential criteria.

References


18. Principles for digital development. URL: https://digitalprinciples.org/ [accessed 2023-12-12]


Abbreviations

**FluCARE:** Influenza Outbreak Communication, Advice and Reporting  
**ICT:** information and communication technology  
**MVP:** minimum viable product  
**PHU:** Public Health Unit  
**RACF:** residential aged care facility  
**SLHD:** Sydney Local Health District  
**UAT:** user acceptability testing
Combining Digital and Molecular Approaches Using Health and Alternate Data Sources in a Next-Generation Surveillance System for Anticipating Outbreaks of Pandemic Potential

Abstract

Globally, millions of lives are impacted every year by infectious diseases outbreaks. Comprehensive and innovative surveillance strategies aiming at early alert and timely containment of emerging and reemerging pathogens are a pressing priority. Shortcomings and delays in current pathogen surveillance practices further disturbed informing responses, interventions, and mitigation of recent pandemics, including H1N1 influenza and SARS-CoV-2. We present the design principles of the architecture for an early-alert surveillance system that leverages the vast available data landscape, including syndromic data from primary health care, drug sales, and rumors from the lay media and social media to identify areas with an increased number of cases of respiratory disease. In these potentially affected areas, an intensive and fast sample collection and advanced high-throughput genome sequencing analyses would inform on circulating known or novel pathogens by metagenomics-enabled pathogen characterization. Concurrently, the integration of bioclimatic and socioeconomic genome sequencing analyses would inform on circulating known or novel pathogens by metagenomics-enabled pathogen characterization. Concurrently, the integration of bioclimatic and socioeconomic data, as well as transportation and mobility network data, into a data analytics platform, coupled with advanced mathematical modeling using artificial intelligence or machine learning, will enable more accurate estimation of outbreak spread risk. Such an approach aims to readily identify and characterize regions in the early stages of an outbreak development, as well as model risk and patterns of spread, informing targeted mitigation and control measures. A fully operational system must integrate diverse and robust data streams to translate data into actionable intelligence and actions, ultimately paving the way toward constructing next-generation surveillance systems.

(JMIR Public Health Surveill 2024;10:e47673) doi:10.2196/47673

KEYWORDS
data integration; digital public health; infectious disease surveillance; pandemic preparedness; prevention; response
Introduction

Overview
The COVID-19 pandemic has illustrated the limited capabilities of existing surveillance systems worldwide and the need to maximize the use of data to inform public health decisions [1,2]. Among the lessons learned from the pandemic is the critical importance of leveraging early alert warning capabilities to detect and respond to emerging and reemerging pathogens before they cause widespread impact. Unfortunately, existing global surveillance systems are often fragmented, lack coordination and collaboration, and do not prioritize early alerts [3,4]. This has resulted in millions of lives lost due to outbreaks of infectious diseases that usually occur following spillover from contacts at the animal-human interface. Therefore, as governments, civil society, and private entities continue to recover from the losses incurred by the COVID-19 pandemic, it is crucial to sustain the momentum generated by this global crisis and prioritize efforts to elevate health surveillance to the next level.

Developing a system focused on early warning tools is a fundamental requirement to enhance pandemic preparedness and response [2,4]. By investing in disease surveillance systems and building resilient health systems, countries can better prevent, detect, and respond to infectious diseases and other health threats. In this endeavor, we should take advantage of the improvements in several branches of technology and science, in particular digitalization of health routine records, data science, and artificial intelligence modeling, for the rapid gathering and analysis of large amounts of diverse data streams [2,4,5]. Next-generation early-alert warning systems should prioritize sensitivity by integrating various data streams for syndromic surveillance, as opposed to relying on traditional case reports. Additionally, achieving specificity through molecular pathogen characterization is crucial for efficient preemptive actions (Figure 1A). However, the process of collecting and analyzing samples is complex and expensive. To optimize logistics and reduce costs, it is recommended to identify priority areas for sample collection through risk assessment based on syndromic surveillance.

To inform the development of next-generation surveillance systems, here we outline key principles and requirements that should be addressed for creating a comprehensive and effective approach to disease monitoring. We instantiate these principles into the Alert-Early System for Outbreaks with Pandemic Potential (ÆSOP)—a system under development in Brazil that aims to improve the surveillance capacity and pandemic preparedness of this country.
Preparing for the Next Pandemic by Harnessing the Power of Data

Many countries, Brazil included, have now established national digital systems that capture routine health information, such as clinical consultations at the primary health care (PHC) level, hospital admissions, and drug prescriptions, and the COVID-19 pandemic has also accelerated integration of these digital systems within health systems and networks [6-8]. The next generation of surveillance systems should leverage this existing and emerging rich data landscape and integrate health routine databases with alternate data streams.

In ÆSOP, early warning will be achieved by monitoring of clinical-related data, which enables the mining of potentially anomalous patterns, such as unexpected surges of respiratory diseases outside the expected seasonality or in atypical age groups or locations (Figure 1B). More specifically, PHC weekly visits coded by reason for encounter, as well as drug sales related to acute respiratory syndromes and social media interactions will be monitored. Brazil’s national PHC database, SISAB (Information System for Primary Health Care; in Portuguese, Sistema de Informação em Saúde para a Atenção Básica), harbors data on the majority of public-funded PHC encounters in the country [9], coded by either the International Classification of Diseases or the International Classification of Primary Care, with 5535 out of 5570 municipalities reporting to SISAB in 2020, corresponding to 98.5% of the total Brazilian population for that year. Complementary, drug sales data encompasses approximately 95% of all drug sales in Brazil.

Continuous monitoring of clinical-related data streams in ÆSOP will enable the system to detect shifts in patterns, which can signal potential outbreaks. By issuing anticipatory warnings to health authorities, the system aims to provide early and timely alerts, as severe cases are not expected to be reported until days or weeks after an increase in mild cases seeking PHC assistance [1,10]. A proof-of-concept evaluating the early-warning capability performance of the system was conducted in Brazil,
totaling over 589 million patient visits to PHC facilities between 2017-2020 in 97% of Brazilian municipalities [11]. Of these visits, 4.24% were associated with upper respiratory infections, and results showed that rises in patient consultation in PHC encounters for respiratory disease were able to anticipate increases of severe acute respiratory syndromes, with geographical variabilities [11]. In a retrospective analysis conducted with data from Bahia, a state with 14.8 million inhabitants in the Northeast region, PHC syndromic monitoring was able to detect COVID-19 entry in 20 out of 21 regions of the state at least 1 week before the rise in COVID-19 cases, showing feasibility for use of the system [12]. The system can also be adapted to the monitoring of other syndromes, such as dengue-like illnesses, by reparametrizing the codes for reason for encounter and drugs used to treat these diseases. This flexibility by design is a key feature that enables the system to be customized to the specific needs of different regions and populations.

Different modeling frameworks were developed and applied to health data aiming at the early detection of outbreaks [13-17], including statistical, mathematical, and computational methods. Although some of them are in use in the routine of health surveillance agencies [15,16], to the best of our knowledge none has been adopted as a gold standard of undoubtedly effectiveness. However, all methods have limitations, and further developments are still needed in this field [2-4]. The proposal described herein aims to combine several approaches in a systemic logical approach combining their strengths. Table 1 presents general features of traditional health surveillance in comparison to what we envision would make a next-generation surveillance system (exemplified by ÆSOP) that overcomes identified deficiencies which hamper pandemic preparedness and response.

### Table 1. Comparison of key features in traditional disease surveillance systems with next-generation approaches.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Traditional surveillance</th>
<th>Next-generation surveillance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data sources</td>
<td>Largely focused on notifiable diseases, clinical and laboratory reports, but can also include sentinel surveillance, registries, surveys, and administrative data systems</td>
<td>Expands these streams to include syndromic data, drug sales, social media, rumors on news outlets and in the digital world, and can be fit to distinct syndromes and to accommodate country-specific data assets</td>
</tr>
<tr>
<td>Data integration</td>
<td>Limited integration of different data streams</td>
<td>Integrates diverse data streams, enabling a comprehensive view</td>
</tr>
<tr>
<td>Data sharing</td>
<td>Limited sharing of data between countries and agencies</td>
<td>Encourages data sharing within a federated approach, allowing for global data collaboration without compromising sovereignty</td>
</tr>
<tr>
<td>Interoperability</td>
<td>Limited interoperability between different health agencies and stakeholders due to lack of standards</td>
<td>Promotes collaboration among various entities by leveraging standardized vocabularies or ontologies</td>
</tr>
<tr>
<td>Early alerting</td>
<td>Typically relies on case reports and may lead to delayed detection</td>
<td>Uses syndromic data for early alerting, enhancing timely outbreak detection</td>
</tr>
<tr>
<td>Flexibility</td>
<td>Often lacks adaptability for different syndromes or regions</td>
<td>Designed with flexibility to target various syndromes</td>
</tr>
<tr>
<td>Molecular analysis (pathogen specificity)</td>
<td>Focuses on traditional laboratory-based pathogen analysis</td>
<td>High-throughput metagenomics for rapid pathogen characterization at specific regions informed by syndromic monitoring</td>
</tr>
</tbody>
</table>

This approach should overcome current public health response limitations, as digital health has been recognized as a critical element [18] and an emerging discipline to prepare for, prevent, and respond to pandemics [1,10], while the syndromic surveillance approach will allow for greater sensitivity [19]. It is also of great importance that this monitoring tool is built over existing systems, as we aim for cost-effectiveness, and no additional burden should be added to already overwrought health systems. The system aims to complement traditional surveillance methods, which in Brazil rely mainly on the mandatory notification of suspected cases of priority diseases listed by the Ministry of Health, a usually paper-based process that often incurs in notification delays.

**Combining Digital and Molecular Approaches for Achieving Both Sensitivity and Specificity**

Genomics-enabled pathogen discovery has been proposed as a game changer for infectious disease surveillance, which some—optimistically—refer to as “not yet a panacea” [20]. The genomics component of ÆSOP, however, leverages sequencing power to provide pathogen specificity for the system at a later stage, rather than initially. By building upon the existing digitally informed streams of syndromic surveillance, which allows for “broad, but swift” monitoring, and using its output to pinpoint areas for the pathogen identification strategy continuously, ÆSOP aims to eke out the costs associated with real-time event monitoring paralleled with targeted genomics.

With further reduction of DNA sequencing costs in developing countries, metagenomics may become a key epidemiological tool for rapid pathogen identification. Alerts generated by clinical-related data will be used internally to activate site-directed genomic analysis for pathogen identification. This will allow for optimizing the logistic strategy for obtaining and analyzing samples. For instance, once a signal of a potential outbreak is identified, field teams will collect clinical samples that undergo molecular characterization for pathogen identification over a small area (Figure 1C). A major challenge in the system is the timely transportation of samples to the processing sites. In large countries with logistic hindrances, it is important to have decentralized sequencing sites for shortening transportation delays. ÆSOP will leverage and use the capillarity of the existing Brazilian Genomic...
Network facilities, such as that induced by Fiocruz during the COVID-19 pandemic [21], the National Network of Health Surveillance Laboratories, and others to speed-up the identification of circulating pathogens. Similar networks have been established in other countries as well, including the COVID-19 Genomics UK Consortium [22], the Indian SARS-CoV-2 Genomic Consortium [23], and the Africa Pathogen Genomics Initiative [24], among others. These initiatives have proven to be valuable in the response against SARS-CoV-2, providing critical insights into the spread and evolution of the virus. As the number of COVID-19 cases continues to decrease in many countries, it is vital that these networks are strengthened and expanded to ensure that they can be used as an essential component of next-generation surveillance systems, allowing for identification of pathogens behind potential outbreaks.

This integrated approach combines both data- and hypothesis-driven disease surveillance approaches, enabling quick identification of anomalies through syndromic surveillance using a combination of bona fide health data and alternative data streams (data-driven approach); once anomalies are identified, targeted sample collection and modeling (hypothesis-driven approach) can be used to confirm the outbreak and determine appropriate actions. To generate risk maps, computational models will integrate socioeconomic, environmental, and mobility and transportation data. Additionally, mathematical models coupled with artificial intelligence and machine learning methods will enable the assessment of risk and routes of outbreak spread (Figure 1B and 1C).

Data Security, Privacy, and Trust

Another unique feature of ÆSOP is the use of a defined methodology for the collection, linkage, and use of relevant and appropriate data. We will use, extract, transform, and load protocols [25] for capturing and standardizing data from different sources, and machine learning methods for feature engineering and longitudinal multimodal data integration. To ensure the security and privacy of data, ÆSOP will follow strict legal and ethical guidelines established by the Center of Data and Knowledge Integration for Health, which has already implemented robust rules and regulations for working with health data, based on global best practices [26]. This approach will ensure that ÆSOP not only provides early warning surveillance for outbreaks of epidemic or pandemic potential but also protects the confidentiality of individuals’ health information.

To ensure the responsible use of data, all information collected will be subject to rigorous digital curation procedures and preserved according to established standards [26]. The resulting deidentified and aggregated data will be made available to health authorities and the scientific community through an application programming interface, allowing for integration with other relevant data sources. This comprehensive set of spatially integrated health, environmental, social, and contextual data has the potential to be a valuable resource for research and other purposes beyond outbreak surveillance.

Fit for Purpose

An essential aspect of ÆSOP is its development in close collaboration with stakeholders and end users, ensuring its adoption by health authorities and successful implementation in mitigating and preventing outbreaks. We have collected feedback and identified key requirements from end users to inform rapid response, by convening workshops and focused meetings. Ultimately, these will lead to creating and fostering local/regional communities of practice for surveillance that will involve end users and other stakeholders in the development and implementation of surveillance systems, inspired by previous experiences such as the US Centers for Disease Control and Prevention’s National Syndromic Surveillance Program [27]. These communities will provide a platform for sharing experiences, challenges, and best practices in surveillance, and will serve as a valuable source of feedback for improving the system. By engaging with end users in this way, we aim to create a sense of ownership and investment in the system, which is crucial for ensuring its long-term success and sustainability through adoption.

Balancing Surveillance Data Sharing Needs With Data Sovereignty Concerns Using a Federated Approach

A critical lesson learned from responding to the COVID-19 pandemic is the need to build trust between governments, organizations, and societies to facilitate the early sharing of accurate information for responsive global health surveillance [2,4]. However, individuals’ privacy and data sovereignty concerns cannot be ignored. While balancing these necessities is not trivial, we understand that finding a solution to safely share information to guide decisions and provide global health security is crucial. In this sense, federated approaches represent a feasible solution that balances the need for data sharing with the need to protect data privacy and sovereignty (Figure 2). By allowing data owners to maintain control over their data, a federated system can help encourage collaboration between organizations, with sharing occurring only on a need-to-know basis. This is particularly relevant in health surveillance, where potentially sensitive data are involved.

A subset of the federated approach, federated learning enables independently built models to be trained on distributed data sets without sharing of raw data; then, an updated model is sent back to a central server to be aggregated with other models, and by iteratively exchanging parameters a globally refined model emerges [28]. This allows for the development of a more robust and accurate model while preserving data privacy. Combined, federated approaches provide a powerful solution for meeting the swift exchange of information under the International Health Regulations while also ensuring compliance with local legislation on individual privacy. Common data models and standards are necessary to ensure that data can be exchanged, integrated, and analyzed effectively across different sites [29].
**Conclusions**

The COVID-19 pandemic, and those before it, evidenced the need for substantial improvements in pandemic preparedness and response, globally. We presented key features and essential building blocks of the ÆSOP system—a system being developed with a combination of data-driven and hypothesis-driven infectious disease surveillance approach for early-alert identification of areas with significant risk of reemergence of infectious disease outbreaks. ÆSOP aims to accurately anticipate an outbreak when compared to currently in-place surveillance systems. Moreover, ÆSOP’s reports should enable precision public health strategies, such as targeted early outbreak mitigation and response strategies, as well as site-directed sample collection for pathogen molecular characterization, thus improving logistics and enhancing responses. To ensure that next-generation surveillance systems will properly meet public health needs, and to obtain a broader perspective of possibilities and bottlenecks, active engagement and cocreation processes with stakeholders, end users, and multidisciplinary experts will be essential and key to building and refining enhanced surveillance systems. Final, we are aware of the revision of the International Health Regulations and the preparation of the Pandemic Treaty, for eventual adaptation of ÆSOP to these rules.

**Acknowledgments**

ÆSOP is under development by Brazil’s Fundação Oswaldo Cruz (Fiocruz) and the Federal University of Rio de Janeiro in a cocreation partnership with the Rockefeller Foundation’s Health Initiative. We acknowledge the representatives of the Secretariat of Health Surveillance (SVS-A, Ministry of Health, Brazil), the Brazilian Health Regulatory Agency (Anvisa), the National Council of Municipal Health Secretariats (Conasems), the Pan American Health Organization, and the Gorgas Memorial Institute for their thoughtful inputs during a Visioning Workshop held in Bahia. The initial concept development of ÆSOP was supported by Fundação Oswaldo Cruz (Fiocruz). The System Visioning workshop in Bahia was funded and co-organized by The Rockefeller Foundation. We thank Andres Marquez-Lara, Felipe Abreu, Jessica Fan, and Sarah Forrester-Wilson from UFacilitate team for inputs during the workshop organization and event facilitation. The valuable inputs from Krista Mizenko to the project are also
acknowledged. The funders did not interfere in the decision to submit the manuscript for publication. Figures were created with BioRender and are published under permission. Details are provided in Multimedia Appendix 1 regarding the “ÆSOP Collaborating Teams.”

Data Availability
Data sharing is not applicable to this article as no data sets were generated or analyzed during this study.

Authors’ Contributions
MB-N is responsible for the conceptualization and coordination of ÆSOP. PIPR wrote the original manuscript draft with input from IM, AIB, KvdH, and MB-N. All authors reviewed and edited the manuscript for intellectual content. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
List of members of the ÆSOP Collaborating Teams.

References
3. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. European Commission. 2021. URL: https://www.astrid-online.it/static/upload/comm/0000/communication150621.pdf [accessed 2023-12-13]


22. COVID-19 Genomics UK Consortium. URL: https://www.cogconsortium.uk/


Abbreviations

ÆSOP: Alert-Early System for Outbreaks with Pandemic Potential

PHC: primary health care

SISAB: Sistema de Informação em Saúde para a Atenção Básica; in English, Information System for Primary Health Care

Edited by A Mavragani, T Sanchez; submitted 29.03.23; peer-reviewed by S Pesãliã, C Zhao; comments to author 01.09.23; revised version received 18.09.23; accepted 22.11.23; published 09.01.24.

Please cite as:

Digital Transformation of Public Health for Noncommunicable Diseases: Narrative Viewpoint of Challenges and Opportunities

Onicio Leal Neto¹,²,³*, PhD; Viktor Von Wyl⁴,⁵*, PhD

¹Department of Computer Science, ETH Zurich, Zurich, Switzerland
²Global Health Institute, Mel & Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States
³Department of Epidemiology and Biostatistics, Mel & Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States
⁴Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland
⁵Epidemiology, Biostatistics & Prevention Institute, University of Zurich, Zurich, Switzerland

*all authors contributed equally

Corresponding Author:
Onicio Leal Neto, PhD
Department of Computer Science
ETH Zurich
Universitätstrasse 6
Zurich, 8092
Switzerland
Phone: 41 44 632 50 94
Email: onicio@gmail.com

Abstract

The recent SARS-CoV-2 pandemic underscored the effectiveness and rapid deployment of digital public health interventions, notably the digital proximity tracing apps, leveraging Bluetooth capabilities to trace and notify users about potential infection exposures. Backed by renowned organizations such as the World Health Organization and the European Union, digital proximity tracings showcased the promise of digital public health. As the world pivots from pandemic responses, it becomes imperative to address noncommunicable diseases (NCDs) that account for a vast majority of health care expenses and premature disability-adjusted life years lost. The narrative of digital transformation in the realm of NCD public health is distinct from infectious diseases. Public health, with its multifaceted approach from disciplines such as medicine, epidemiology, and psychology, focuses on promoting healthy living and choices through functions categorized as “Assessment,” “Policy Development,” “Resource Allocation,” “Assurance,” and “Access.” The power of artificial intelligence (AI) in this digital transformation is noteworthy. AI can automate repetitive tasks, facilitating health care providers to prioritize personal interactions, especially those that cannot be digitalized like emotional support. Moreover, AI presents tools for individuals to be proactive in their health management. However, the human touch remains irreplaceable; AI serves as a companion guiding through the health care landscape. Digital evolution, while revolutionary, poses its own set of challenges. Issues of equity and access are at the forefront. Vulnerable populations, whether due to economic constraints, geographical barriers, or digital illiteracy, face the threat of being marginalized further. This transformation mandates an inclusive strategy, focusing on not amplifying existing health disparities but eliminating them. Population-level digital interventions in NCD prevention demand societal agreement. Policies, like smoking bans or sugar taxes, though effective, might affect those not directly benefiting. Hence, all involved parties, from policy makers to the public, should have a balanced perspective on the advantages, risks, and expenses of these digital shifts. For a successful digital shift in public health, especially concerning NCDs, AI’s potential to enhance efficiency, effectiveness, user experience, and equity—the “quadruple aim”—is undeniable. However, it is vital that AI-driven initiatives in public health domains remain purposeful, offering improvements without compromising other objectives. The broader success of digital public health hinges on transparent benchmarks and criteria, ensuring maximum benefits without sidelining minorities or vulnerable groups. Especially in population-centric decisions, like resource allocation, AI’s ability to avoid bias is paramount. Therefore, the continuous involvement of stakeholders, including patients and minority groups, remains pivotal in the progression of AI-integrated digital public health.

(JMIR Public Health Surveill 2024;10:e49575) doi:10.2196/49575
Introduction

The SARS-CoV-2 pandemic saw the broad roll out and application of the first digital public health intervention at scale: digital proximity tracing (DPT) [1]. The general idea of DPT smartphone apps consists of leveraging the capabilities of standard Bluetooth sensors to trace proximity contacts and to notify the exposed individuals of past proximity encounters in case of a confirmed SARS-CoV-2 infection. The development and wide-ranging use of DPT apps have been a success story on several different levels. Their speedy development and release early in the pandemic reflects the prosperous international collaboration and backing by many relevant organizations such as the World Health Organization or the European Union.

To maintain the momentum for digital public health, it will be important to extend the digital transformation of public health in the domain of noncommunicable diseases (NCD). Prior to the pandemic, approximately 80% of all health care expenditures and 63.8% of the premature loss of disability-adjusted life years were associated with NCDs [2]. The challenges, but also opportunities, of NCD digital public health are quite different from the public health response to mitigate infectious diseases such as SARS-CoV-2. This perspective paper will start with an overview of current definitions of digital public health. Furthermore, in addition to requirements for digital transformation, we will explore different aspects that can guide the identification of potential use cases for digital public health applications in NCDs. The paper will close with an outlook on challenges and conditions to make further inroads on digitally transforming NCD public health.

Key Functions of Public Health

Public health is defined as the science and practice to establish conditions so that people can live healthy lives and make healthy choices. It draws on methods and knowledge of multiple disciplines such as epidemiology, medicine, psychology, law, social sciences, and many others [3]. Textbox 1 illustrates the essential public health functions that should be implemented by jurisdictions to achieve equitable health and health system access. These functions can broadly be grouped into “Assessment,” “Policy Development,” “Resource Allocation,” “Assurance,” and “Access.”

The framework emphasizes that public health involves more than just securing access to or ensuring high-quality medical care for individuals and an adequate health care workforce. Generally, public health also has a strong focus on communities and touches on the foundations for healthy living (eg, living conditions, education, or economic aspects), as well as population-based preventive measures (eg, childhood vaccinations).
Textbox 1. Institutional framework of essential public health functions.

Assessment
- Assess and monitor population health status and well-being, factors that influence health and its social determinants, community needs and assets, and health system performance and impact.
- Investigate, diagnose, and address health problems and hazards affecting the population, including public health surveillance and control and management of health risks and emergencies.
- Promotion and management of health research and knowledge.

Policy development
- Communicate effectively to inform and educate people about health, factors that influence it, and how to improve it.
- Strengthen, support, and mobilize communities and partnerships to improve health.
- Create, champion, and implement policies, plans, and laws that impact health.
- Use legal and regulatory actions designed to improve and protect the public’s health.

Assurance
- Assure an effective system that enables equitable access to the individual services and care needed to be healthy.
- Build and support a diverse and skilled public health workforce.
- Improve and innovate public health functions through ongoing evaluation, research, and continuous quality improvement.
- Build and maintain a strong organizational infrastructure for public health.
- Social participation and social mobilization, inclusion of strategic actors, and transparency.

Resource allocation
- Development of human resources for health.
- Ensuring access to and rational use of quality, safe, and effective essential medicines and other health technologies.
- Efficient and equitable health financing.
- Equitable access to comprehensive, quality health services.

Access
- Equitable access to interventions that seek to promote health, reduce risk factors, and promote healthy behaviors.
- Management and promotion of interventions on the social determinants of health.

Note: The framework is based on The Public Health National Center for Innovation [4] and Pan-American Health Organization [5].

Levels of Digital Changes

Overview
In health care and public health the levels to which specific processes and workflows, can be turned “digital” varies drastically. These levels are commonly described as digitization (moving from analogic to digital data collection), digitalization (integration of digital tools into established processes and workflows), and digital transformation (redesigning processes and workflows around digital tools). For the remainder, we will use the more neutral term “digital change,” which is intended to encompass all 3 levels.

Digital change attempts to achieve an impact on public health and health care by enabling better, more efficient, and more cost-effective public health interventions [6]. Thus it mirrors the triple aim of health care: improving patient outcomes, enhancing patient experiences, and reducing health care costs [7]. As such, digital change may also help to steer public health’s directions toward the goal of the 4Ps: preventive, predictive, personalized, and participatory [8]. By harnessing the -omics revolution, but also digital technologies such as wearables, mobile apps, and artificial intelligence (AI), health care providers can deliver preventive care by identifying potential health issues before they become severe. Predictive analytics can help identify patients at risk of developing certain conditions, allowing for earlier intervention and treatment. Personalized medicine can also be delivered through digital tools that analyze patient data to create tailored treatment plans. Finally, participatory health care allows for greater collaboration among communities, providers, and individuals, enabling these groups to take a more active role in managing their health. Consumers and patients are increasingly empowered to collect their own data and access medical information, thus gaining confidence, and claiming a role in health decision-making.
Conditions for Successful Digital Change in Health Care and Public Health

On the technical side, system readiness, interoperability, technical application programming interfaces, and nontechnical interfaces are crucial for implementing digital public health. System readiness ensures that digital technologies are compatible with existing health care systems and infrastructure [9,10]. Interoperability is essential to enable different systems to work together seamlessly, enabling for exchange of information and data between different platforms [11,12]. Interoperability also fosters the ability to collect and analyze vast amounts of health data in real time, thus allowing public health practitioners, policy makers, and citizens to gain valuable insights into health outcomes, treatment effectiveness, and disease trends [13]. As such, AI-driven analytics may also foster participation and personalization, as postulated by P4 [8]. However, ensuring that data analysis is unbiased and AI fairness is maintained is essential to avoid automating inequalities driven by AI modeling [14,15].

A further critical element in digital change routes is the scalability of technical systems and processes as health care systems look to scale their operations, widen care to more patients, and enhance the democratization of health care and public health [16]. Cloud-based platform solutions can help those public systems achieve scalability while reducing costs and improving efficiency as well as reducing disparities in public health [17-19]. Scalability also pertains to practices that involve humans-in-the-loop. Therefore, preparing the health care and public health workforce for digital changes is an important avenue to fill the gaps in skills and digital literacy in the next generation of public health practitioners, hence ensuring that health care systems can fully leverage the benefits of digital technologies [20-23]. Training programs and courses should focus on various disciplines and domains, including the Internet of Things, data analysis, data stewardship, and AI. Other essential domains for training programs and courses include cybersecurity, data privacy, and regulatory compliance where health care operators must be trained to ensure that they can protect patient data and comply with relevant regulations and legislation. Training programs should also cover soft skills such as communication, teamwork, and problem-solving, ensuring that health care providers can work effectively in interdisciplinary teams and collaborate with other stakeholders.

Creating a favorable legal and ethical context is critical for successful digital changes in public health. Legal frameworks must provide clear guidance on data privacy, security, and ownership issues [24,25]. In addition to legal and ethical considerations, financial factors also play a significant role in digital transformation in public health. Advantageous financial incentives can encourage health care providers to invest in digital technologies, enabling them to deliver better, more efficient, and more cost-effective care. Reimbursement and remuneration systems must also be adapted to account for incentivizing the development and use of digital technologies, ensuring that technology and service providers are fairly compensated for these technologies [26]. These models must also consider the value of these technologies, such as improved health outcomes, reduced health care costs, and increased user (ie, citizens or patients) satisfaction [27,28].

Yet, it will also be essential to plan for unexpected events or circumstances that could derail digital transformation efforts. By establishing a robust risk management plan and building flexibility into digital transformation initiatives, health care providers can minimize the impact of unforeseen events and ensure that their digital transformation efforts continue to deliver value and benefits to patients and health care providers alike [29-33].

Digital Change + Public Health = Digital Public Health?

The explicit connection between “digital” and “public health” is quite recent and well summarized by Iyamu et al [34], who trace “digital public health” back to a report by Public Health England from 2017 [35]. This report essentially calls for a reimagining of public health using digital tools, thus calling for a digital transformation of public health, a view that has recently been shared by representatives of the European Public Health Association [34]. As part of the transformation process, some authors suggest not only focusing on the technical possibilities but also looking for inspiration from the start-up culture of the tech industry, with a stronger emphasis on experimenting and adapting (pivoting) rather than waiting until scientific evidence has solidified. Other definitions describe digital public health as an “asset” toward achieving traditional public health goals, for instance by leveraging new data, methods, and work processes that have been given rise by digitization [36]. Therefore, these existing definitions appear to differ mainly in the meaning and role of “digital” on the digitalization-digital transformation spectrum, as already pointed out by Iyamu. But all authors agree that a consequential digital public health goes way beyond the digitization of existing services and interventions—they require rethinking and redesigning full health care delivery and public health services pathways.

Considering the framework of essential public health functions (Textbox 1), there are likely some domains that lend themselves more easily to a digital change, because digital data acquisition and analysis are at their core. Matching public health functions with suitable goals and purposes of digital change processes is critical for reaping the benefits of digital technologies. One of the key areas where digital transformation can be leveraged is in complex data analysis [37,38]. By contrast, the “digitalization or digital transformation question” is less easily resolved for functions such as building a sufficient health care workforce or equitable health care access. This is visualized in Figure 1 as conditions and technologies for digital change. In the figure, the “Public Health Functions” (Textbox 1) are depicted in the inner circle, and possible examples of digitally changed essential public health function processes are shown in the outer circle. The size of the circles in the outer circle illustrates our broad assessment of the achievability of digital change for specific public health functions, by taking the bespoke conditions for successful change into account, namely technical readiness, reliance on human interactions, or market interest in existing public health function processes. In general, processes relying...
on data collection and analysis or digitally facilitated mass communication lend themselves more easily to digital changes, while the creation and maintenance of an adequate workforce for health care and public health will likely continue to depend on human factors in the foreseeable future.

**Figure 1.** Conditions and technologies for digital change and concrete examples based on the public health functions. AI: artificial intelligence.

---

**Goal and Purpose of Digital Change**

To make further inroads toward digital change in public health functions and NCD public health in particular, it seems sensible to distinguish between the goal and purpose of making these functions and processes more digital. By goal, we mean the high-level public health goals of broad health care quality, access, efficiency, and equity of health promotion and health care provision, which can be fostered by digital changes. For instance, in light of the shortage of mental health care providers in many settings, apps for mental health assessments may ease access to at least some basic level of mental health care. Indeed, 24 of the 54 reimbursable digital health apps [39] (as of August 30, 2023) in Germany address mental health challenges including stress, burnout, or depression.

By contrast, we consider the purpose of digital change as improvements in the public health or health care service delivery process by digital change. These process attributes can, for example, be defined in analogy to the “triple aim” of health care innovation whether digital change renders the process of providing and delivering existing or novel public health measures and health care in a more effective, efficient, or user-friendly manner.

An alternative view on process attributes that are influenced by digital changes is offered by Peter Diamandis’ 6Ds of exponential technologies [16,40]. This concept identifies 6 categories of process attributes—digitization, deception, disruption, demonetization, dematerialization, and democratization—which also outlines the desired trajectory of a digital change process. Initiating digital changes in the health care sector may initially elicit resistance in some actors, who may initially respond with incredulity about the potential benefits of these changes in the short term. As the proposed innovation begins to ramp up, it reduces inefficiencies and optimizes current systems, disrupting the same old processes. Demonetization occurs as health care costs shift away from the fee-for-service paradigm, requiring new ways to deal with the small costs of care. Dematerialization further changes the paradigm by enabling care access outside of formal settings, such as through mobile health apps for therapeutics. Finally, democratization widens health care access thanks to these new characteristics.

Therefore, digital changes in the health care and public health sector need to be particularly mindful of 1 particular 6D [16] the creation of a digital divide due to a lack of democratization, meaning that digital changes can impose high hurdles to health
promotion or health care access for subgroups of individuals with lower digital literacy or different language backgrounds. Very often, these are also individuals who may benefit most from prevention or health care. Such barriers can be overcome, for example, by making user interfaces as simple as possible and by providing personal support, which can also be assisted by digital means. Therefore, it is advisable to include equity aims in digital change processes, and to define the purpose of digital change as a “quadruple aim.”

**Examples of Leveraging Digital Advances in a Pragmatic Digital Public Health**

Some digital advances in public health are already contributing toward making health care more equitable to underserved populations in low- and middle-income settings. Telemedicine, for example, enables patients to consult with health care providers remotely, breaking down geographic barriers to care. Digital technologies have also enabled health care professionals to reach more patients with chronic conditions, offering remote monitoring and management tools that improve patient outcomes while reducing costs. Bringing a spotlight on sub-Saharan countries, these innovations in health care access also were driven by the application of drones in delivering medicines, vaccines, and diagnostics tests to remote areas [41,42]. Furthermore, drones have been seen playing an important role in identifying malaria-vector breeding sites in rural areas as an important tool for disease surveillance [43].

As another example yet focused on low- and middle-income countries, Reach Digital Health [44] is leveraging digital technology to transform patient care, streamlining care requests, check-ins, follow-ups, and symptom assessments for a more patient-focused approach. Key initiatives include MomConnect [45], a WhatsApp smart bot that provides maternal and child health information in local languages to pregnant women in Sub-Saharan Africa, and Young Africa Live [46], a digital platform that empowers youth to make informed decisions about their sexual health. These innovative interventions, aimed at making health care more accessible and proactive, highlight the transformative potential of digital health in addressing public health challenges globally.

Expanding these digital advancements for a global health perspective, a notable instance of this progression is the “Global Flu View” platform [47]. This tool, which provides real-time tracking of global influenza activity based on participatory surveillance, signifies a transition from traditional to digital epidemiology. It presents unmatched opportunities for early detection, timely intervention, and strategic planning, all of which are crucial for managing influenza outbreaks and potential pandemics. Beyond its data visualization capabilities for influenza-like illness trends, Global Flu View also promotes knowledge sharing among public health practitioners [48]. As such, Global Flu View can also act as a template for similar collaborations in the field of NCD by acting as a code lab repository for exchanging scripts and algorithms for data analysis, facilitating peer collaboration, and fostering continuous learning and skill development of researchers and health care professionals. Furthermore, these digital platforms do not just address immediate public health needs; they also provide a valuable resource for training future digital epidemiologists. Through hands-on interaction with real-time health data, trainees can develop skills to transform intricate health data into actionable public health interventions.

**Emerging AI Trends and Their Possible Role in Digital Public Health**

In addition to automating complex data analysis and predictions, AI-driven digital transformation can replace repetitive tasks in stable environments, such as administrative tasks and data entry [18]. This frees up public health professionals and health care providers to focus on personal interactions with citizens and patients. Along those lines, it is vital to identify which human interactions or tasks cannot be easily replaced by digital technologies and should be strengthened, such as counseling and emotional support [19]. AI can also provide citizens and patients with new tools to take a more active role in managing their health [49]. However digital health is not a replacement for doctors or other health professionals, but rather a companion to guide individuals through the health care system. By leveraging digital tools, citizens and patients can take a more active role in managing their health and accessing care. This can help to address some of the challenges associated with aging, such as chronic disease management, as well as promote health equity by improving access to care for underserved populations. Additionally, digital health can help to promote sustainability in health care by reducing the need for in-person visits and streamlining care delivery.

Recent advancements in large language models (LLMs) have the potential to greatly enhance public health interventions by providing tailored, accessible information at scale [50]. LLM technology facilitates the dissemination of health information in a way that can be easily adapted to a population’s education level, cultural background, or general preferences. This can be particularly beneficial in public health campaigns, where the goal is to effectively communicate with diverse audiences who may find traditional medical jargon intimidating or confusing [51]. LLM technology, through its natural language processing capabilities, allows for the analysis of population-level data, enabling the creation of public health recommendations and advice that are contextually relevant. It can help communities better understand their collective risk factors or prevalent health conditions, thereby promoting health literacy, preventive measures, and access to care. The capability of LLM to interact in a more human-like chat can also revolutionize how public health messages are communicated, making them more engaging and comprehensible. Moreover, with its ability to provide 24/7 service, LLM has become an efficient, ever-present tool for promoting lifestyle changes and mitigating noncommunicable diseases at the population level. There is already evidence from health care settings illustrating the benefits of LLM technology in enhancing patient interactions. Applying this technology on a broader scale could potentially have transformative impacts on public health communication strategies, risk mitigation, and health promotion. Given the long time horizon of emergence and living with chronic diseases, a knowledgeable, tireless, emphatic, efficient, and 24/7-available companion would be a highly welcomed tool to support lifestyle changes or health promotion to avoid NCDs.
The opacity of some AI models in how predictions are achieved (eg, in neural networks) is a common point of critique and reason for nonadoption. Explainable AIs enable public health experts, health care providers, and patients to understand how AI-based systems arrive at their decisions and recommendations [52-57]. This can help to address issues of bias and discrimination in health care and public health, which can arise when algorithms are based on flawed or incomplete data sets. By providing transparency and accountability, explainable AIs can help to ensure that health care automation is more equitable and unbiased, although this area is still underproof and requires more evidence to achieve its pledge [58,59]. Machine unlearning is another emerging area of research in digital health that has the potential to promote equity and fairness in health care automation [60-62]. Machine unlearning involves the intentional removal of certain data points or variables from a machine learning algorithm in order to correct biases or errors. For example, if an algorithm is found to be discriminating against certain patient populations based on race or ethnicity, machine unlearning can be used to remove those variables and correct the bias. By doing so, machine unlearning can help to ensure that health care automation is more equitable and inclusive.

However, while explainable AIs and machine unlearning can help to promote equity and fairness in health care automation, they also raise important questions about the role of humans, liability, and trackability in the digital health space. For example, who is responsible if an AI-based system makes a mistake or causes harm? How can we ensure that these systems are transparent, accountable, and auditable? And what role do humans play in overseeing and monitoring these systems? These problems are even exacerbated in any application of AI in public health decision-making or prediction at the population level. While problematic treatment decisions in health care will be spotted and corrected through close patient monitoring, similar monitoring for larger populations is much more complex. Furthermore, AI-driven public health decisions may impact specific population subgroups differently. Therefore, a decision of whether AI-supported decisions are good or bad requires a much more complex framework that includes considerations of equity, vulnerability, effectiveness, and efficiency. As such, it is important to consider not only the technical aspects of AI-supported public health, but also the social, legal, and ethical implications of these emerging technologies. By doing so, we can ensure that digital health remains an inclusive, equitable, and accessible space for all patients and that these new technologies truly serve as companions to guide individuals through the health care system.

Federated learning and swarm learning hold transformative potential for public health, providing a path to harness vast, diverse data sets while preserving data privacy and data ownership. Federated learning involves training machine learning models across multiple decentralized instances with a central trusted server [63]. This offers a powerful solution for public health systems grappling with the need to analyze large, varied data sets while respecting privacy concerns. Swarm learning, in contrast, uses a completely decentralized network of instances to collaboratively train machine learning models in a peer-to-peer manner [64]. This becomes especially valuable in public health contexts where data are highly decentralized and distributed across numerous devices and locations. Swarm learning allows public health practitioners to harness the power of distributed data to create more precise and effective machine learning models, ensuring data privacy and security. Adopting these approaches can have profound benefits for public health systems aiming to enhance their digital health initiatives [65]. By refining algorithms at the population level, as well as in specific subpopulations, public health systems can improve the precision of disease prediction and early detection. This also allows for the identification of opportunities for targeted interventions and treatments, which in turn can help reduce health care costs, enhance health outcomes, and promote health equity and accessibility. Moreover, federated learning and swarm learning provide tools for developing more effective public health policies and interventions. By analyzing patterns and trends in population-level health data, public health systems can identify high-risk groups and more effectively direct interventions. This ability can play a crucial role in preventing the spread of infectious diseases, reducing health care costs, and enhancing public health outcomes overall.

A Pragmatic Vision for Making Inroads to Digital Change of Public Health for NCD

Digital change offers an unprecedented opportunity to address the substantial burden of NCDs at scale. However, the application of digital technologies needs to be thoughtfully aligned with public health functions, goals, and broader societal implications. Digital platforms can serve as tools for better public health monitoring, decision-making, as well as public engagement, and education. On the individual level, prevention requires behavioral changes and continuous personal efforts. One of the key downstream factors in a micro-level (individual) perspective is the psychosocial mechanisms mediated by social support, social influence, and social engagement [66]. Having these aspects addressed might mean a successful impact on the digital change of public health. On one hand, personalized prevention strategies have proven effective but lack scalability. On the other hand, digital solutions, such as telehealth platforms and health monitoring applications, may offer more scalable approaches by improving efficiency, effectiveness, and user experience.

While digital change opens new avenues, it simultaneously poses challenges related to equity and access. Populations that are (1) economically disadvantaged, (2) geographically isolated, or (3) digitally illiterate are at risk of being left behind in the digital transformation journey. Any initiative must include a comprehensive strategy to ensure that these changes do not exacerbate health inequities but rather work toward eliminating them. Furthermore, effective digital and nondigital NCD prevention at a population level demands a societal consensus. It often requires a willingness by people not directly at risk to participate in a larger social contract. For example, the legislation of smoking bans and “sugar taxes” may affect individuals who do not directly benefit from such measures. Therefore, stakeholders, from policy makers to health care providers and the general public, need to have realistic
expectations about the benefits, potential harms, and costs of digital changes in health care or public health service delivery. Ongoing monitoring and evaluation mechanisms need to be in place to gauge the broader-scale implications of digital change.

Conclusions

Digital health and AI in particular have great potential to make public health and health care of NCDs more efficient, effective, user-friendly, and equitable: the quadruple aim. Digital change in any of the 5 essential public health function domains should be purposeful and bring improvements in at least 1 dimension of the quadruple aim while not significantly compromising the other dimensions. But ultimately the success of digital public health and AI-supported public health measures will require clear criteria, monitoring, and benchmarks on how to achieve benefits for the majority while not putting minorities and vulnerable groups at a disadvantage. As such, AI explainability and unlearning of discriminatory or biased decision-making will be even more crucial for population-level decisions (which are common in public health, eg, regarding resource allocation) than in individual-level health care. Therefore, the close involvement of stakeholders, including patients, citizens, and vulnerable and minority populations, will be crucial in the development, roll out, and monitoring of AI-powered digital public health.

Conflicts of Interest

None declared.

References


44. Reach: Health Made Possible. 2023. URL: https://www.reachdigitalhealth.org/ [accessed 2024-01-09]


46. Young Africa Live. 2023. URL: https://www.youngafricaalive.org/ [accessed 2024-01-09]


Abbreviations

AI: artificial intelligence
DPT: digital proximity tracing
LLM: large language model
NCD: noncommunicable disease

©Onicio Leal Neto, Viktor Von Wyl. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 25.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Defining the Dimensions of Diversity to Promote Inclusion in the Digital Era of Health Care: A Lexicon

Yashoda Sharma¹, PhD; Anindita Saha², BSc; Jennifer C Goldsack¹, MChem, MA, MBA

¹Digital Medicine Society, Boston, MA, United States
²Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, United States

Corresponding Author:
Yashoda Sharma, PhD
Digital Medicine Society
90 Canal Street
4th Floor
Boston, MA, 02114
United States
Phone: 1 765 234 3463
Email: yashoda@dimesociety.org

Abstract

The pandemic provided a stark reminder of the inequities faced by populations historically marginalized by the health care system and accelerated the adoption of digital health technologies to drive innovation. Digital health technologies’ purported promises to reduce inefficiencies and costs, improve access and health outcomes, and empower patients add a new level of urgency to health equity. As conventional medicine shifts toward digital medicine, we have the opportunity to intentionally develop and deploy digital health technologies with an inclusion focus. The first step is ensuring that the multiple dimensions of diversity are captured. We propose a lexicon that encompasses elements critical for implementing an inclusive approach to advancing health care quality and health services research in the digital era.

(JMIR Public Health Surveill 2024;10:e51980) doi:10.2196/51980

KEYWORDS
digital medicine; inclusion; digital health technology/product; digital health; digital technology; health care system; innovation; equity; quality; disparity; digital era; digital access; digital literacy

Introduction

A Lack of Diversity Limits Health for All

As the health care industry undergoes digitization, leaders across the field have the opportunity to develop approaches that can advance access and increase equity and quality. To do this, we need to examine the source of health disparities in the current system: the underrepresentation of specific populations as participants in research and inadequate health care access for all patients. By recent accounting, less than 3% of published genome-wide association studies include data on people of African, Hispanic, or Latin American ancestry, and 86% of clinical trial participants are of European ancestry [1]. This type of exclusion and the accumulation of health disparities as a consequence of inadequate access to care are deeply rooted in systemic and structural racism [2].

The pandemic’s inequitable toll on public health and health care was fueled by knowledge gaps resulting from decades of exclusion and inadequate care for the most vulnerable populations. COVID-19 infection and death rates, one example of public health consequences of health disparities, highlighted the dire consequences of a lack of diversity in research. Groups historically underserved based on the dimensions of race and ethnicity are disproportionately affected and dying at higher rates than their distribution in the population. Additional dimensions of diversity, such as lower socioeconomic status and limited English proficiency, continue to influence health burdens, health outcomes, and overall quality of life [3-5].

Currently, federal policies driving diversity initiatives for patients in health care and participants in research focus on race and ethnicity [6-9] and are intended to encourage and support research that includes more racial and ethnic groups, as well as women. Some progress has been made; however, the focus has been narrow and incomplete. Attention to race, ethnicity, or sex is not sufficient to promote broad health equity. We propose widening the focus of health equity to better capture the many dimensions of diversity. Age, gender, sexual identity, socioeconomic status, educational attainment, physical and mental health, and a range of other factors contribute to disparities in health outcomes, including access to care, health behaviors, and the distribution of health care resources. By addressing these additional dimensions of diversity, we can work toward more equitable and inclusive health care systems.

(Access to the full article is provided by the publisher or through libraries.)
cognitive abilities, access to care, and geography also impact care and quality of life.

Social determinants of health (SDOH), that is, the conditions and environment in which people live, work, learn, and play, significantly contribute to health inequities [10]. When overlaid with the intersectionality of race, ethnicity, age, gender, and sexual identity, SDOH expands the categorization for diversity and populations underserved by health care [11]. The All of Us Research Program [12], which was developed to increase representation in biomedical research, expanded the criteria for diversity to reflect this intersectionality and interdependence of factors affecting health. Now, a similar expansion is necessary to account for the many factors that contribute to a lack of digital equity and inclusion, including SDOH [13], digital access and digital and health literacy [14-16], and community lived experiences [17].

Define Diversity to Promote Inclusivity

The digitization of health care and research has the potential to transform how we care for people and develop new medical products [18,19]. The Digital Medicine Society, in collaboration with the Center for Devices and Radiological Health of the Food and Drug Administration (FDA) and with guidance issued by the FDA [7,8], has been at the forefront of implementing cross-disciplinary approaches to advance the ethical, effective, equitable, and safe use of digital technologies to redefine health care and improve lives by addressing relevant evidentiary, security, ethical, regulatory, and legal issues [20]. It is with this cross-disciplinary, multistakeholder approach that we propose a complete lexicon of dimensions of diversity that must be considered to ensure inclusion in the digital era of health care. This lexicon is intended to support not only efforts to increase diversity and promote inclusion but also to ensure that health disparities are not exacerbated by these new technologies.

The COVID-19 pandemic emphasized 2 elements that are critical for equity in the digitization of health care and research. The first is the dire consequences of health care built on research with limited populations and confounded by the inequities of care caused by decades of structural racism [21]. Limited research with Black/African American and Hispanic people has resulted in less access to safe and effective medications and a higher underlying disease burden, which increased the risk of a more severe COVID-19 prognosis or higher rates of death [22]. The second is a glimpse at the potential of digital health solutions. However, as digital health solutions have accelerated the transformation of health care, the decades-long systemic barriers to health equity continue to hamstring the potential for effective care for all patients. We saw this for telehealth visits, which were higher among those identifying as White and earning at least US $100,000 [23], and low vaccine distribution and adoption due to early reliance on technology [24]. This underscores is a clear example of the need for inclusivity; it is not sufficient to implement technologies and to hope people will access and use them.

An intentional commitment to inclusion is critical during the development and deployment of digital health solutions to facilitate and advance equity. The All of Us Research Program definition of “underrepresented in biomedical research” [12] must be expanded to include digital inclusion to continue to address health disparities centered on demographics, as well as environmental and lifestyle factors.

New Dimensions of Diversity for Inclusion in the Digital Era of Medicine

Digital health is the umbrella term for the intersection of technology and health care [25]. Digital medicine is the field of evidence-based digital health tools that measure or intervene in the service of health to support the practice of medicine broadly. To advance health equity, we need to focus on inclusion in digital medicine [26].

Inclusion in digital medicine means (1) being cognizant of characteristics of different populations, and (2) tailoring solutions to ensure that digital health products meet the needs of and benefit, all individuals and communities. This entails not just addressing the needs of individuals who face barriers to digital technology use but also addressing the historical, institutional, structural, and discriminatory forces that created and continue to perpetuate the digital divide [27] and health inequality. We combined the dimensions of age, race, ethnicity, education, socioeconomic status, religion, ability, location, gender, sexual preference, language, ability, and digital technology access and literacy and propose a new, expanded lexicon of dimensions in diversity suitable for the digital era of health (Table 1).
Table 1. A newly expanded lexicon of dimensions in diversity suitable for the digital era of medicine.

<table>
<thead>
<tr>
<th>Diversity dimensions</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Pediatric and adolescent populations and adults older than 65</td>
</tr>
<tr>
<td>Annual household income</td>
<td>Individuals with annual incomes equal to or below 200% of the Federal poverty level</td>
</tr>
<tr>
<td>Digital technology access</td>
<td>Communities with limited access to high-speed internet, such as broadband, or access to digital technologies, such as computers and tablets</td>
</tr>
<tr>
<td>Digital technology literacy</td>
<td>Individuals or communities not well versed in the use of digital technology (eg, connecting to the internet and Bluetooth pairing)</td>
</tr>
<tr>
<td>Disability</td>
<td>Individuals with either a physical or cognitive disability, including visual, auditory, and mobility</td>
</tr>
<tr>
<td>Educational attainment</td>
<td>Individuals with less than a high school degree or equivalent and individuals with limited health literacy</td>
</tr>
<tr>
<td>Gender identity</td>
<td>Individuals who identify as a gender variant, nonbinary, transgender, or something else</td>
</tr>
<tr>
<td>Geography</td>
<td>Individuals who reside in rural or nonmetropolitan areas, individuals residing in areas with limited internet access, and individuals who are homeless</td>
</tr>
<tr>
<td>Language</td>
<td>Individuals with limited English proficiency (written or spoken)</td>
</tr>
<tr>
<td>Race and ethnicity</td>
<td>Individuals who identify as other than White and non-Hispanic based on their ancestry (eg, African Americans/Black, Asian, Hispanic/Latinx, Native Hawaiian or Pacific Islander, and Middle Eastern or North African)</td>
</tr>
<tr>
<td>Cultural practices</td>
<td>Individuals or communities that may abstain from accessing and using digital technologies (eg, some religions discourage the use of technology on certain days)</td>
</tr>
<tr>
<td>Sex identified at birth</td>
<td>Individuals who are neither male nor female (eg, intersex)</td>
</tr>
<tr>
<td>Sexual orientation</td>
<td>Individuals who identify as asexual, bisexual, gay or lesbian, or something else</td>
</tr>
</tbody>
</table>

aDefinitions adopted from Mapes et al [12]. The other definitions were developed by members of the Digital Health Measurement Collaborative Community [28].

Digital medicine can prioritize inclusion through a human-centered, fit-for-purpose lens; that is ensuring that digital products do what they claim for all intended users and serve all members of the populations who can benefit from them. A focus on inclusion will capture diverse populations and lead to more equitable solutions and outcomes for health care. This will also address a big challenge for digital health technologies: trust [29]. Trust is important for full participation; full participation leads to health equity—the state in which everyone has an opportunity to attain their highest level of health [30]. Mistrust of health care systems has been growing for years and has been amplified with the COVID-19 pandemic. Digital health product developers have also contributed to this mistrust, especially when people do not see technology designed for them. A clear example is the photoplethysmography (PPG) optical sensors used in many fitness trackers and pulse oximeters. While PPG sensors have been shown to be inaccurate in people with darker pigmentation [31-33], the technology persists in many widely available products. Developers are evolving fitness trackers [34] with PPG sensors to play a bigger role in collecting relevant health data; however, there are few indications that the limitations of this technology for inclusivity are being addressed.

In the digital era of health, digital access and literacy have emerged as defining factors for equity. Access to high-speed internet, reliable and secure Wi-Fi, and tools such as computers, tablets, and smartphones are greatly impacted by socioeconomic status [35]. Digital health literacy, from recognizing the relevance and value of digital health products in your care to having products in your native language, will dictate the type and level of care you receive. There are currently several limitations to inclusion including resource allocation, data and knowledge gaps, and education and training. A multistakeholder approach which includes health care institutions, policymakers, and research and education systems is required to overcome current limitations.

The proposed new complete lexicon for dimensions of diversity for the digital era of health can effectively support intentional approaches to digital health solution design and deployment [28] for inclusivity. The lexicon is designed to support the intersectionality of specific populations and communities that the field of digital health is intended to serve, allowing for some level of customization for each person. The dimensions of diversity complement Richardson et al’s [36] framework for digital health equity, which expanded the National Institute on Minority Health and Health Disparities Research Framework. Our lexicon provides a level of granularity that those developing or deploying (either in clinical care or research) digital health products should sufficiently represent to ensure an inclusive experience. The lexicon has been applied to the digital health measurement market opportunity calculator, which enables digital health measurement product developers to build a business case for incorporating inclusive practices into their research and development processes [37]. Users of the market opportunity calculator select a dimension of diversity, for a health condition, and receive an output estimating the increase in engagement size and potential increase in market value when that product is made with an inclusive lens for that dimension. The lexicon can also be applied to the Partner, Identify, Demonstrate, Access, Report equity framework for behavioral digital health interventions [38] to focus efforts on specific
diversity dimensions and define specific actions for each point of the framework. These equity frameworks provide a structure whereby our lexicon can be added for inclusivity. Together these can lead to transformational changes that will advance health equity.

**Insist on Action**

Health equity in clinical research was strengthened by the 2023 omnibus spending bill and the FDA’s diversity plan [39] and is projected to continue in the next year. In a recent Deloitte report, health care and life science leaders rank health equity as a top 10 priority [40]. Partnerships across health care, especially health care providers, communities, and digital health product developers will be instrumental in advancing equity and driving inclusion. A level of accountability will be required to demonstrate early progress; the lexicon for the dimensions of diversity can easily be customized in a dashboard [41] to track and inform strategies and resources for more inclusive digital product development and deployment.

The Digital Medicine Society is hosting the Digital Health Measurement Collaborative Community (DATAcc) [28]. Advancing health equity with collaborative communities is a priority at the FDA’s Center for Devices and Radiological Health for addressing health care challenges. DATAcc convenes industry leaders, academic and clinical researchers, patient and community organizations, health systems, and government/regulatory representatives to develop and demonstrate best practices to advance harmonized approaches to speed the use of digital health measurement to improve health outcomes and health equity. Building on the medical device product lifecycle [42], DATAcc created resources to action the expanded dimensions of diversity to drive inclusivity in both digital health technology development and deployment [37].

Recognizing that the efforts to drive inclusion falls on everyone in the health care and health research ecosystems, DATAcc also designed resources to guide clinical care and research teams in assisting patients with understanding complex concepts associated with digital health product use, such as data privacy and security, and end user agreements. With widespread use, validation, and verification inclusive product development and deployment resources can lead to industry-wide adoption and formalization of inclusivity in digital health care. Thereby, building digital medicine on the foundation of inclusion.

**Conclusions**

Leaders in health care are at a crucial juncture that will shape the future of digital medicine. We can develop technologies in a way that is more inclusive to ensure that advances in digital medicine are available and used by all populations. We also recognize that inclusive product development is necessary but not sufficient; end users (patients and research participants) need complete information and training on these products so they can be informed. This has to be an ongoing, ever-evolving process that can grow and change with the adoption of new, more complex technologies. The proposed lexicon offers leaders and organizations the granularity they need to demonstrate their health equity efforts. Inclusivity is the key to accelerating health equity so that diverse populations become more integrated and are better served by the health care system.

**Acknowledgments**

The authors would like to thank members of the Digital Health Measurement Collaborative Community (DATAcc) for providing valuable insights on the current status of inclusion with digital health measurement product development and deployment, and opportunities for improvements.

**Data Availability**

Data sharing is not applicable to this article as no data sets were generated or analyzed during this study.

**Disclaimer**

The Food and Drug Administration (FDA) participates as a member of the Digital Health Measurement Collaborative Community (DATAcc). This article reflects the views of the authors and should not be construed to represent the FDA’s views or policies.

**Authors’ Contributions**

All authors contributed to conceptualization and design of the work, writing, research, interpretation, literature searches, visual display, and approval for final submission.

**Conflicts of Interest**

None declared.

**References**


8. Enhancing the diversity of clinical trial populations—eligibility criteria, enrollment practices, and trial designs guidance for industry. US Food and Drug Administration. URL: https://www.fda.gov/media/127712/download [accessed 2023-04-13]


Abbreviations

- DATAcc: Digital Health Measurement Collaborative Community
- FDA: Food and Drug Administration
- PPG: photoplethysmography
- SDOH: social determinants of health

Edited by A Mavragani; submitted 18.08.23; peer-reviewed by D Verran, D Dinh; comments to author 08.10.23; revised version received 08.11.23; accepted 07.01.24; published 09.02.24.

Please cite as:

Sharma Y, Saha A, Goldsack JC

Defining the Dimensions of Diversity to Promote Inclusion in the Digital Era of Health Care: A Lexicon

JMIR Public Health Surveill 2024;10:e51980
URL: https://publichealth.jmir.org/2024/1/e51980
doi: 10.2196/51980
PMID: 36335012
Viewpoint

Designing Electronic Data Capture Systems for Sustainability in Low-Resource Settings: Viewpoint With Lessons Learned From Ethiopia and Myanmar

Natalie Benda¹, PhD; Kylie Dougherty¹, RN, BSN, MPhil; Abebe Gebremariam Gobezayehu²,³, MD; John N Cranmer²,³,⁴, MPH, MSN, ANP-BC, DNP; Sakie Zawtha⁵, MD; Katerina Andreadis⁶, MS; Heran Biza², BA; Ruth Masterson Creber¹, RN, MSc, PhD

¹School of Nursing, Columbia University, New York, NY, United States
²Emory-Ethiopia Partnership, Bahir Dar, Ethiopia
³Bahir Dar University, Bahir Dar, Ethiopia
⁴Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, United States
⁵Health and Hope Myanmar, Lailenpi, Myanmar
⁶New York University Grossman School of Medicine, New York, NY, United States

Corresponding Author:
Natalie Benda, PhD
School of Nursing
Columbia University
560 W 168th St
New York, NY, 10032
United States
Phone: 1 212 305 5756
Email: nb3115@cumc.columbia.edu

Abstract

Electronic data capture (EDC) is a crucial component in the design, evaluation, and sustainment of population health interventions. Low-resource settings, however, present unique challenges for developing a robust EDC system due to limited financial capital, differences in technological infrastructure, and insufficient involvement of those who understand the local context. Current literature focuses on the evaluation of health interventions using EDC but does not provide an in-depth description of the systems used or how they are developed. In this viewpoint, we present case descriptions from 2 low- and middle-income countries: Ethiopia and Myanmar. We address a gap in evidence by describing each EDC system in detail and discussing the pros and cons of different approaches. We then present common lessons learned from the 2 case descriptions as recommendations for considerations in developing and implementing EDC in low-resource settings, using a sociotechnical framework for studying health information technology in complex adaptive health care systems. Our recommendations highlight the importance of selecting hardware compatible with local infrastructure, using flexible software systems that facilitate communication across different languages and levels of literacy, and conducting iterative, participatory design with individuals with deep knowledge of local clinical and cultural norms.

(JMIR Public Health Surveill 2024;10:e47703) doi:10.2196/47703

KEYWORDS

low and middle income countries; LMIC; electronic data capture; population health surveillance, sociotechnical system; data infrastructure; electronic data system; health care system; technology; information system; health program development; intervention

Introduction

Information systems supporting electronic data capture (EDC) for research and clinical operations are crucial to the development, evaluation, and sustainability of population health programs, facilitating health improvements supported by data. In some low- and middle-income countries (LMIC), EDCs can trend health outcomes and inform decision-making, thereby improving the efficiency and effectiveness of health services [1,2]. EDC can support health intervention and policy development along with determining the allocation of critical resources [3]. With the expansion of internet services globally,
using EDC systems is a cost-effective way to increase the amount and quality of data that can be used in LMIC [4,5]. However, EDC will only be useful in promoting positive health outcomes when the system matches the required users, tasks, and environment [6]. Implementing EDC systems to monitor health indicators is a complex process. It requires that EDC systems be developed to (1) suit the needs of the local context, (2) support long-term sustainability, and (3) accommodate routine adaptations [7-9]. Barriers to successful development include limited financial capital for implementation, low utilization of local expertise, and narrow conceptualizing of EDC technologies while ignoring relevant sociotechnical factors that should drive EDC decisions [10].

Published information on health interventions in LMIC typically focuses on the results of the interventions themselves rather than the EDC used for population health surveillance. Therefore, guidance for developing EDC systems is limited [10]. One qualitative study, for example, evaluating the implementation of an EDC, described the lack of trained information system specialists, inconsistent definitions of variables, insufficient data validation, managers not valuing the data, and communication issues as the core challenges to EDC implementation [10]. One systematic review found 132 published studies that used health information systems for data collection in LMIC [3]. The study reported on data analysis methods and data quality, noting that data completeness was a commonly cited issue with EDCs and exclusion, followed by imputation, was the most common means for handling missing data [3]. These publications highlight a common pattern among studies regarding EDC in LMIC, that is, if they go beyond simply describing the impacts of the interventions, the primary focus involves data quality and completeness. Although quality and completeness are important elements to evaluate with EDC, there remains a gap in describing the nature of the EDC systems and how their components were assembled and implemented.

Due to this gap in evidence, the objective of this viewpoint is to provide practical recommendations for designing EDC systems for research and health operations to sustain data-informed population health interventions globally. We describe 2 exemplar cases in Ethiopia and Western Myanmar/Eastern India to illustrate how robust EDC systems may be designed to support population health surveillance, specifically during humanitarian crises (ie, civil unrest and the COVID-19 pandemic). The research activities related to these cases have been described in other publications [11,12]. In this viewpoint, we describe case descriptions using a sociotechnical conceptual model to determine barriers and facilitators of successful development and use of EDC. We used this model because it provides a comprehensive framework that can be used to guide and evaluate EDC design and implementation [13].

Context and Project Overviews

Ethiopia Case Description–The Saving Little Lives Initiative
As of 2019, Ethiopia had high ratios of neonatal mortality (33 per 1000 live births) and infant mortality (47 per 1000 live births) [14]. Globally, Ethiopia has the fourth-highest number of newborn deaths [15]. Further, the Sub-Saharan Africa region has the highest neonatal mortality ratio in the world [15].

Ethiopia-specific causes of neonatal death include respiratory distress (45%), infection (30%), and birth asphyxia (13%) [14]. To improve newborn and infant health, the Emory-Ethiopia Partnership began work in Amhara in 2011. This case description focuses on one of Emory-Ethiopia’s efforts in the region—Saving Little Lives (SLL). SLL is a flagship health program of the Ethiopian government that spans 290 targeted hospitals across 4 regions. SLL is an aggressive scale-up study designed to reduce population-level mortality by 35% through improved newborn care. SLL uses a package of targeted and synergistic interventions to promote survival. SLL consortium partners are supporting the regional health bureaus to scale up 5 key interventions, as follows: (1) respiratory support (including resuscitation for newborns with asphyxia and continuous positive airway pressure for respiratory distress), (2) kangaroo mother care (KMC), (3) sepsis management, (4) feeding support, and (5) cross-cutting quality improvement methods. The project covers a population of over 37 million individuals in 4 Ethiopian regions and 285 woredas (districts). Emory-Ethiopia is the lead partner for the Amhara region in this national SLL consortium. These SLL activities build upon the same consortium’s scale-up research on KMC conducted in the same regions [12,16].

As of September 2022, Ethiopia had nearly 500,000 confirmed cases and over 7500 deaths from COVID-19. In November 2020, following claims that a minority party, Tigray People’s Liberation Front, had attacked a federal army base, the majority party ordered a military offensive in Tigray [17]. This ongoing conflict has led to the forced displacement of over 2 million civilians and a food crisis affecting at least 9.4 million northern Ethiopians—primarily in the Tigray region [18]. Furthermore, the crisis has led to an increase in gender-based violence and a decrease in access to proper health services for women and children [18]. Research associated with this case was approved by institutional review boards (IRBs) in the United States and Ethiopia.

Eastern India and Western Myanmar Description–The Mobile Health (mHealth) and Mobile Ultrasound for Mothers (mMUM)
Western Myanmar has one of the highest maternal mortality ratios in Asia, ranking 148 out of 187 on the 2019 Human Development Index [19,20]. Among women of reproductive age, maternal deaths account for 1 in every 10 deaths, with 75% occurring during delivery or immediately after birth [19].

On February 1, 2021, the Tatmadaw (Myanmar’s military) seized control of the government in a coup d’État. Health care professionals across Myanmar have participated in the Civil Disobedience Movement in response to the overthrow of a
The various aspects of research associated with this case have been approved by IRBs in the United States and India.

**Information Infrastructure Design and Adaptations**

Sociotechnical conceptual models may provide a structure for discussing EDC considerations in LMIC—particularly those settings impacted by civil conflict. Figures 1 and 2 show the information infrastructure for each case using Sittig and Singh’s “sociotechnical model for studying health information technology in complex adaptive healthcare systems” [13]. Labeled boxes represent established components in the Sittig and Singh’s model, and the descriptions and symbols highlight how these components were operationalized in each context. Tables 1 and 2 further describe the actions taken to create and sustain these EDCs during dual humanitarian crises organized by the constructs in the Sittig and Singh’s model.
Figure 1. Electronic data capture infrastructure for the Saving Little Lives (SLL) initiative.
Figure 2. Electronic data capture infrastructure for the Mobile Health and Mobile Ultrasound for Mothers (mMUM) study. NGO: nongovernmental organization.
Table 1. Electronic data capture (EDC) design and implementation features to support population health surveillance for the Saving Little Lives (SLL) initiative [16,24,25].

<table>
<thead>
<tr>
<th>Sociotechnical dimension</th>
<th>SLL EDC implementation activities and design features</th>
</tr>
</thead>
</table>
| Hardware and software computing infrastructure | • Used KoBo Toolbox for easy-to-design, adaptable data collection forms.  
• Leveraged ODK\(a\) for open access sharing, allowing for the easy addition of more measures as needed to streamline data collection updates across sites.  
• Collected data primarily via tablet and used paper to list mothers and babies discharged for further follow-up on the 29th day of life.                                                                                                                                                                                                 |
| Clinical content                     | • Designed content and metrics to ensure information was relevant to the local context while promoting internal and external validity (eg, incorporating clinical cascade metrics for newborns and mothers into routine SLL data collection) [16,24].  
• Developed programs and data collection to adhere to and be evaluated against standard quality guidelines, such as BEmONC\(b\) quality or newborn cascades.                                                                                                                                                        |
| Human computer interface             | • Data collection forms developed with local partners and researchers based in the United States.  
• Embedded Amharic language into data collection (eg, patient consent forms).                                                                                                                                                                                                                                                                                       |
| People                               | • The SLL initiative funds government-based evaluation officer positions with responsibilities including primary data collection, data training, and ensuring facility-specific data are submitted to central database.  
• Implementation officers support the training and implementation of SLL at the facility.  
• Program learning officers collect qualitative information (eg, enablers and barriers) from different units and care providers.                                                                                                                                                                                                                   |
| Workflow and communication           | • Evaluation officers complete real-time data collection of project indicators; program learning team obtains qualitative information from families and clinicians for continuous improvement.  
• Existing government staff at hospitals collect and enter facility-specific data into the electronic national monitoring system (ie, DHIS2\(c\)).  
• Community-based health extension workers collect governmental surveillance, service provision data, and other community needs assessments using the community health information system. Community health workers record data on paper. These data are then submitted to their supervising health center, which, in turn, forwards them to the woreda (district) health office. The woreda health office aggregate these data electronically and shares them with the primary health care unit. These electronic data directly enter the national database, and the regional government can query this database for their region. |
| Internal and organizational policies, procedures, and culture; external rules, regulations, and pressure | • Funding embedded government staff members (eg, evaluation officers) enables the synchronization of SLL data collection with government health surveillance.  
• SLL deliberately involves university-based colleagues (eg, Emory University) so that the program may generate transferrable research findings in addition to its local public health and clinical impact.  
• All SLL-specific data collection and procedures were approved by 3 institutional review boards from Emory University, Addis Ababa University, and Amhara Public Health Institute.                                                                                                                                 |
| System measurement and monitoring    | • Leveraging DHIS2 surveillance data collected via the government’s Health Management Information System.  
• Qualitative and quantitative evaluation via government-embedded evaluation officers who collect real-time SLL data outside the government’s DHIS2 system.  
• Data collected by evaluation officers, program learning officers, health care providers, and health extension workers integrated by the Amhara SLL research team for comprehensive SLL program evaluation at multiple levels (eg, district, regional, or national levels).                                                                                            |

\(a\)ODK: Open Data Kit.  
\(b\)BEmONC: Basic Emergency Obstetric and Neonatal Care.  
\(c\)DHIS2: District Health Information Software 2.
The SLL Initiative (Ethiopia)—Initial EDC System

The EDC used to evaluate SLL’s progress stems both from government health systems and SLL-dedicated staff—the program learning team and evaluation officers (Figure 1). Health Information technicians and other care providers at health facilities routinely collect government data, including information on health service utilization and clinical outcomes (eg, newborn mortality and postnatal care uptake) using the government’s District Health Information Software 2 (DHIS2) [21]. Evaluation officers are government-based staff, funded by SLL, who capture information on SLL-specific quality indicators related to clinical care (eg, uptake of KMC and provision of skin-to-skin contact). Evaluation officers visit SLL facilities and enter these data in the KoBo Toolbox (described below) with measures aggregated in a centralized data warehouse [22]. As partners with the Ethiopian Ministry of Health, SLL partners have access to routine government data for their region from the DHIS2 system. The SLL program staff access both these routine DHIS2 data and SLL-specific program data. These multisource data facilitate the optimization of interventions using both internal and external data sources. Refinement and utilization of EDC procedures have also directly informed core content areas in the national Basic Emergency Obstetric and Neonatal Care guidelines [23].

The data collection platform for SLL has been developed using KoBo Toolbox and an Open Data Kit (ODK) tool. The KoBo Toolbox allows for open-source sharing of input forms and enables users to design data collection forms. Forms designed in the KoBo Toolbox can be imported into ODK, which provides support. The KoBo Toolbox allows for open-source sharing of input forms and enables users to design data collection forms. Forms designed in the KoBo Toolbox can be imported into ODK, which provides

Table 2. Electronic data capture (EDC) design and implementation features to support population health surveillance for the Mobile Health (mHealth) and Mobile Ultrasound for Mothers (mMUM) study [25].

<table>
<thead>
<tr>
<th>Sociotechnical dimension</th>
<th>mMUM EDC implementation activities and design features</th>
</tr>
</thead>
</table>
| Hardware and software computing infrastructure | • Selected data collection tools (ie, ArcGIS and Survey123) that work in settings with limited internet connectivity or no internet.  
• Purchased chargers that worked with local power sources (ie, solar, due to insufficient access to electricity).  
• Selected software with mapping capabilities to locate, support, and track transient populations. |
| Clinical content | • Developed data collection content with local personnel to ensure clinical and social questions were asked in a culturally sensitive manner.  
• The content of the collected data was adapted to suit the situation in a new country with different infrastructure [25].  
• Added medical anthropologist to the team to provide insights into how clinical problems are perceived and communicated in the given setting. |
| Human computer interface | • Chose a tool that also supported icon-based communication to provide context clues to users with lower literacy.  
• Chose a tool that could provide an interface not only in multiple languages but also supported languages with non-Roman characters. |
| People | • Used iterative participatory design process to create a culturally appropriate user interface.  
• Worked with local stakeholders to devise a plan to motivate data collection, specifically building on the strong sense of community and crafting messaging regarding how data collection can support the community. |
| Workflow and communication | • Local teams and community health workers handled most of the data collection due to lack of regular government surveillance data that may be used to track outcomes.  
• Planned to complete research data collection with those who worked most closely with patients (local nongovernmental organization and community health workers) but also involved higher-level providers (eg, midwives) who would do more advanced procedures (eg, ultrasound).  
• Pivoted to work in tandem with government-funded community health workers in India but are focusing the local team’s efforts on refugee camps, where existing health workers are struggling to have the bandwidth to provide support. |
| Internal and organizational policies, procedures, and culture | • Had to determine explicit language and marking for ownership of data collection devices, such as phones and ultrasound machines.  
• Added security protections to data collection devices in case they were stolen or intercepted, including encryption, dual-factor authentication, and remote wipe capabilities. |
| External rules, regulations, and pressure | • Developed memoranda of understanding with local governments to conduct data collection and research operations.  
• Adapted goals and data collection procedures based on local laws. |
| System measurement and monitoring | • Developed a procedure for assessing data quality using plausibility ratings by local experts since there is no ground truth with which we can compare the data collected.  
• We have been and will continue to collect qualitative data regarding how research data collection may impact workflow and productivity.  
• used government-collected data in the new location [25] to assess the most pressing health issues (eg, anemia), identify areas of greatest need, and detect possible shortcomings in data collection. |
Evaluation officers collect SLL-specific data using tablets with the ODK interface and forms and submit the data to the KoBo server. Each of them has a power bank for charging and airtime to upload data into the web-based server. All the collected data are fed into a centralized database to support collaboration and evaluation at multiple levels (ie, district, regional, and national levels as well as for scientific analysis).

Evaluation officers also use a collateral data collection strategy. They have printed tracking sheets that include information such as the name of the mother, her address, her phone number, or family members’ phone numbers, allowing them to call and gather information on the status of the mother and newborn.

Table 1 describes the features of the SLL EDC infrastructure.

The mMUM Study (Western Myanmar/Eastern India)–Initial EDC System

To design the mMUM study’s EDC system, we selected the ArcGIS (Esri) mapping software coupled with a survey platform (Survey123) [26] allowing community health workers and local nongovernmental organization team members to easily collect data in areas with low internet connectivity or no internet (Figure 2). ArcGIS uses cloud-based data storage. This helped support the local data collection and analysis from team members not working at the local sites. However, the location of the entity providing the cloud storage (ArcGIS in this case, located in the United States) came into question during the local IRB approval process. Cloud data storage approval was eventually granted, as only coded data and not personal health identifiers would be saved. The data collection forms allow for the inclusion of different languages and pictures to facilitate form comprehension. The survey software supports mobile apps for field-based data capture. The mapping software facilitates aggregation and embeds data into high-quality geographical information systems for tracking transient populations (eg, displaced person groups).

Table 2 provides additional details regarding the EDC features for the mMUM study.

The mMUM study required further adaptations following the move from Myanmar to India, facilitated by the design of the EDC system in place, as highlighted based on relevant sociotechnical dimensions described below.

Clinical Content

Our team has worked with local organizations from Myanmar and those native to India to update the content of the questions asked to patients in the data collection instruments. For example, new questions related to refugee status have been added, and questions pertaining to access to health facilities have been omitted, considering the stability and well-established nature of health facilities in India.

Human-Computer Interface

Our selection of a tool that could be translated into multiple languages has been helpful in relocation efforts, as the materials had to be translated into an additional language to support the Indian health workers.

Workflow and Communication

Based on discussions with Indian government officials, local team efforts leverage the medical expertise of those who are also Burmese refugees. This enables them to collect data and provide interventions to pregnant refugees, a population the Indian health system has found challenging to support due to the large and initially unexpected influx of patients in need. The Burmese aid workers also have the linguistic and cultural competency to support the unique needs of the refugee patients.

External Rules, Regulations, and Pressure

Our efforts have involved gaining memoranda of understanding with state-level government health officials in both Myanmar (initially) and India (adaptation). We have also had to adapt our interventions slightly, as in Myanmar we had planned to use portable ultrasound to serve rural areas. Portable ultrasound, however, is illegal in India due to issues with unregistered devices being used for gender selection (ie, female feticide). Therefore, our team has shifted focus to ensuring women are aware of where they can access legal ultrasounds and have transportation. We also plan to implement additional interventions that leverage the information we are collecting to better support the needs of pregnant refugees.

System Measurement and Monitoring

Myanmar did not have enough frequent public health surveillance data to help inform and pinpoint our team’s efforts. However, the population health surveillance in India is much more robust (and publicly available), allowing us to use this information to determine which districts are most in need.

Discussion of Lessons Learned

Overview

The 2 cases discussed involve curating EDC, considering various interacting sociotechnical dimensions to design EDC infrastructure for population health surveillance. This infrastructure is aimed at supporting research and clinical activities in 2 LMIC. Robust system design allowed our team to continue operations and make adaptations in varied capacities through close collaboration with local partners. This provided crucial data-informed health information relevant to vulnerable newborns, women, and families in these 2 LMIC despite ongoing humanitarian crises. The ability of both the mMUM and SLL projects to continue despite the dual humanitarian-COVID-19 crises was facilitated by the design of sustainable EDC infrastructure, which provided a foundation for each project’s interventions. There are transferrable lessons on optimizing EDC design for research and public health or clinical practice in LMIC, emerging from these 2 cases.

Despite differences in context, common lessons may be drawn from the 2 exemplar cases. Table 3 provides recommendations synthesized from our collective experience to guide EDC development and implementation in LMIC in future cases, organized using the dimensions of Sittig and Singh’s framework [13]. Both cases had strong foundations based on technical dimensions (eg, hardware, software, and human-computer interfaces) that supported necessary social, cultural, and
organizational implementation. Multiple recommendations stress the importance of directly involving those with a deep understanding of clinical context, people (end users), and cultural norms in the EDC development and implementation process. These recommendations align with the current literature, underscoring the importance of leveraging an iterative design approach when developing EDC and involving local end users and experts [27-29]. Others have also described the benefits of KoBo’s ODK and its ability to support EDC in settings with low internet connectivity [28]. However, previous studies typically describe EDC in the context of a single, specific system implementation without describing general important considerations, as is done in this viewpoint. Both cases involved engaging with government officials and data, although the SLL team has greater integration with the Ethiopian government. The mMUM team has gained approvals to work with government-funded community health workers in certain districts and has been able to use government-collected outcomes for maternal and child health in India to guide EDC development and variable selection. However, direct integration, as accomplished by the SLL team, is more challenging due to the displacement of some team members from Myanmar.

### Table 3. Lessons learned for electronic data capture (EDC) development from the described case studies.

<table>
<thead>
<tr>
<th>Sociotechnical dimension</th>
<th>Recommendation for EDC Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware and software computing infrastructure</td>
<td>Consider data collection tools that function in scenarios with low or no internet connectivity.</td>
</tr>
<tr>
<td>Clinical content</td>
<td>Ensure there is an understanding of important perceptions of the clinical conditions evaluated and how clinical concepts may be best translated into local languages.</td>
</tr>
<tr>
<td>Human-computer interface</td>
<td>Conduct iterative participatory design with end users who deeply understand the context to ensure system usability and feasibility.</td>
</tr>
<tr>
<td>People</td>
<td>Ensure some of those involved in system development are familiar with and work directly in the local settings (similar to the “clinical content” dimension).</td>
</tr>
<tr>
<td>Workflow and communication</td>
<td>Engage in regular, continued communication among partners (both local and abroad, as applicable) throughout implementation.</td>
</tr>
<tr>
<td>Internal and organizational policies, procedures, and culture</td>
<td>Involve those who can support ensuring buy-in from necessary local stakeholders, such as community health workers and patients (similar to the “clinical content” and “people” dimensions).</td>
</tr>
<tr>
<td>External rules, regulations, and pressure</td>
<td>Leverage local and governmental partnerships to understand and work within the target regions’ regulations.</td>
</tr>
<tr>
<td>System measurement and monitoring</td>
<td>Define data security and storage procedures in advance.</td>
</tr>
<tr>
<td>System measurement and monitoring</td>
<td>Use governmental data for measurement (in the SLL(^b) case) or as a gold standard comparison point (in the mMUM(^c) case) when available.</td>
</tr>
</tbody>
</table>

\(^a\)DHIS2: District Health Information Software 2.  
\(^b\)SSL: Saving Little Lives.  
\(^c\)mMUM: mHealth and Mobile Ultrasound for Mothers.

Compared to SLL, the mMUM study required more extensive adaptations to continue the project. This may be due to several factors, including the project being in earlier phases and less well-established than SLL as well as the need to relocate to a different country altogether due to a national government coup. Myanmar also represents a particularly challenging context from the perspective of developing EDC systems, as the country was essentially closed to external global influence until 2014. Consequently, other global countries with more technology-related partner engagement may have further developed EDC systems at baseline compared to Myanmar.

### Conclusions

This viewpoint addresses a gap in the literature by providing concrete recommendations for the development and refinement of EDC in LMIC, supported by case descriptions from 2 projects. In each case, the EDC infrastructure developed allowed the teams to continue clinical and research-based operations through humanitarian crises, including the COVID-19 pandemic and civil unrest. Common themes from each case have been presented as a series of recommendations for future EDC development and implementation in LMIC. Local system descriptions and recommendations leverage Sittig and Singh’s “sociotechnical model for studying health information technology in complex adaptive healthcare systems” [13], suggesting this model may be helpful for LMIC settings in...
addition to its previous use in high-income countries. Key elements of our recommendations include selecting hardware that accommodates local infrastructure, selecting systems that allow for communication across language and literacy levels, conducting iterative design with those with deep local and contextual knowledge throughout the design process, as well as gaining approval and buy-in from governmental entities.

Acknowledgments

The exemplar case based in Ethiopia gratefully acknowledges the members of the Emory-Ethiopia Partnership.

The exemplar case based in Myanmar/India gratefully acknowledges the contributions of the former Myanmar-based not-for-profit, Health and Hope Myanmar, and the Ferrando Integrated Women Development Centre in India.

Conflicts of Interest

None declared.

References


Abbreviations

- **DHIS**: District Health Information Software
- **EDC**: electronic data capture
- **IRB**: institutional review board
- **KMC**: kangaroo mother care
- **LMIC**: low- and middle-income countries
- **mHealth**: mobile health
- **mMUM**: mHealth and Mobile Ultrasound for Mothers
- **ODK**: Open Data Kit
- **SSL**: Saving Little Lives

Edited by A Mavragani; submitted 29.03.23; peer-reviewed by O Ogundaini, M Platt; comments to author 28.06.23; revised version received 19.08.23; accepted 12.12.23; published 12.02.24.

Please cite as:

Designing Electronic Data Capture Systems for Sustainability in Low-Resource Settings: Viewpoint With Lessons Learned From Ethiopia and Myanmar
JMIR Public Health Surveill 2024;10:e47703
URL: https://publichealth.jmir.org/2024/1/e47703
doi:10.2196/47703
PMID:38345833
Viewpoint


Lara Kufoof, MSc; Rana Hajjeh, MD; Mohannad Al Nsour, PhD; Randa Saad, MD; Victoria Bélorgeot, MSc; Abdinasir Abubakar, MSc, MD; Yousef Khader, ScD; Salman Rawaf, PhD

2Department of Program Management, Regional Office for the Eastern Mediterranean, World Health Organization, Cairo, Egypt
4Department of Research and Policy, Global Health Development, Eastern Mediterranean Public Health Network, Amman, Jordan
5Regional Office for the Eastern Mediterranean, World Health Organization, Cairo, Egypt
6Department of Public Health, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan
7Department of Primary Care and Public Health, School of Public Health at Imperial College London, London, United Kingdom

Corresponding Author:
Yousef Khader, ScD
Department of Public Health
Faculty of Medicine
Jordan University of Science and Technology
Alramtha-Amman Street
Irbid, 22110
Jordan
Phone: 962 796802040
Email: yskhader@just.edu.jo

Abstract

The COVID-19 transmission in the Eastern Mediterranean Region (EMR) was influenced by various factors such as conflict, demographics, travel and social restrictions, migrant workers, weak health systems, and mass gatherings. The countries that responded well to COVID-19 had high-level political commitment, multisectoral coordination, and existing infrastructures that could quickly mobilize. However, some EMR countries faced challenges due to political instability and fragile health systems, which hindered their response strategies. The pandemic highlighted the region’s weak health systems and preparedness, fragmented surveillance systems, and lack of trust in information sharing. COVID-19 exposed the disruption of access and delivery of essential health services as a major health system fragility. In 2020, the World Health Organization (WHO) conducted a global pulse survey, which demonstrated that the EMR experienced the highest disruption in health services compared to other WHO regions. However, thanks to prioritization by the WHO and its member states, significant improvement was observed in 2021 during the second round of the WHO’s National Pulse Survey. The pandemic underscored the importance of political leadership, community engagement, and trust and emphasized that investing in health security benefits everyone. Increasing vaccine coverage, building regional capacities, strengthening health systems, and working toward universal health coverage and health security are all priorities in the EMR. Emergency public health plays a key role in preparing for and responding to pandemics and biological threats. Integrating public health into primary care and investing in public health workforce capacity building is essential to reshaping public health and health emergency preparedness.

(JMIR Public Health Surveill 2024;10:e40491) doi:10.2196/40491

KEYWORDS
COVID-19; integration; pandemic preparedness; primary health care; public health

Introduction

The World Health Organization (WHO) was first alerted to cases of pneumonia of unknown origin on December 31, 2019. By January 30, 2020, the WHO had declared the novel coronavirus outbreak a public health emergency of international concern [1]. The WHO’s Director General described this outbreak of COVID-19 as a pandemic on March 11, 2020 [2]. Since then, many countries have announced several restrictive public health measures to contain the virus, such as travel bans,
border restrictions, lockdowns, and mandatory quarantines [3]. As a result, global economic growth was severely impacted, with the global gross domestic product dropping by 6.7% in 2020 [4,5]. These measures, along with the strain on resources due to the care of those infected with COVID-19, have disrupted the access to and delivery of essential health services [6]. Even the strongest health systems were heavily impacted and overwhelmed by the pandemic.

In 2020, the WHO conducted a global pulse survey to understand the impact of COVID-19 on health systems. Almost 90% of the 105 engaged countries reported interruptions in different services, ranging from routine and elective service delivery to critical care, especially in low- and middle-income countries [6]. Financial constraints, supply chain disruptions, redirection of services to the care of patients with COVID-19, and workforce unavailability affected access to essential health services [5]. The pandemic revealed that no country was sufficiently prepared against biological threats. Many risks and gaps were identified in the current public health system that hindered countries’ capacities for response. These challenges and gaps called for increased investments and stronger political will to enhance health emergency preparedness [7].

In January 2020, the WHO activated its incident management system at all 3 levels of the organization (global, regional, and country levels), in line with the WHO’s emergency response framework [3]. This system safeguards the coordination of response actions during public health emergencies [3]. The WHO’s Eastern Mediterranean Region (EMR) Incident Management Support Team (IMST) for COVID-19 was activated on January 22, 2020, as a coordination mechanism providing technical, strategic, and operational support to EMR countries. It has been operational for over 2 years as the WHO’s longest-running IMST. On January 29, 2020, the first cases of COVID-19 were reported in the EMR, and by April 10, 2020, all 22 EMR countries and territories had reported COVID-19 cases [8].

The EMR is composed of 22 countries, categorized into 3 groups based on socioeconomic development. Group 1 includes Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates (UAE), which have the most resources and are all high-income countries. Group 2 includes Egypt, Iran, Iraq, Jordan, Lebanon, Libya, Morocco, Palestine, Syria, and Tunisia, which have the next most level of resources and are upper-middle- or lower-middle-income countries. Group 3 includes Afghanistan, Djibouti, Pakistan, Somalia, Sudan, and Yemen, which have the least resources and are all either lower-middle- or low-income countries [9]. The diversity of country incomes, emergencies, cultures, and health system capacities in the EMR led to varying response capacities, COVID-19 knowledge and risk perceptions, and socioeconomic impacts, which have usually been substantial [10].

This viewpoint highlights the challenges that faced EMR countries and their achievements and lessons learned during the COVID-19 pandemic. It provides an overview of the unequal COVID-19 vaccination coverage in the region and discusses the methods and approaches of how to reshape public health in the region and strengthen health emergency preparedness, providing recommendations for the way forward.

### Challenges in the EMR

Over 50% of the countries in the region are affected by complex emergencies, either directly or indirectly. The EMR holds 9 major humanitarian emergencies, 102 million people needing humanitarian assistance (37% of the global total), and over 32 million refugees and internally displaced persons [8].

Political instability, fragile health systems in some EMR countries, multiple disease outbreaks, and poor accessibility and availability of basic health care services have hampered the effectiveness and efficiency of the strategies adopted to combat COVID-19 [11]. The COVID-19 response in the EMR was influenced by multiple factors affecting viral transmission, such as state fragility and conflict; demographics, as this region has a younger population than most; early applications of travel and social restrictions, which limited spread at the beginning of the pandemic; large numbers of migrant workers; mass gatherings; and pilgrimage. Saudi Arabia made an unprecedented decision to downsize Hajj and suspend Umrah, and other EMR countries implemented public health and social measures. However, as “COVID-19 fatigue” set in, many countries loosened their restrictive measures, which often affected the disease trend [10]. Although some countries used and built on past health emergency experiences and systems, such as Saudi Arabia applying lessons learned from the Middle East respiratory syndrome experience [12,13], other countries were poorly prepared. For example, 6 EMR countries still lack national infection prevention and control guidelines, and 5 (Afghanistan, Iraq, Libya, Palestine, and Tunisia) developed their infection prevention and control guidelines in the past year only, with the support of the WHO [14]. The COVID-19 pandemic has exposed gaps in the health systems at multiple levels globally, even in high-income countries with strong health systems. The Islamic Republic of Iran, for example, which had a strong existing health care system, witnessed the largest rate of infections and deaths [15].

The pandemic also highlighted the weak epidemiological capacity of the region [13]. Data were generated but were not always analyzed, interpreted, and used as evidence for action. Additionally, surveillance systems are fragmented, with many being old and paper based [16]. There is also an issue of trusting the shared information in the region because of the underreporting of some countries or their hesitancy to share data [17,18].

Several countries within the EMR encountered obstacles in addressing the COVID-19 pandemic. For instance, Pakistan struggled due to its fragile health care infrastructure, characterized by shortages of health care professionals, hospital beds, and essential medical equipment necessary for treating patients with COVID-19 [19]. Similarly, Iran encountered difficulties in delivering crucial medical and humanitarian supplies due to economic sanctions [20]. Additionally, Yemen and Syria faced difficulties managing the pandemic within the context of ongoing conflict, displacement, and the challenges of maintaining health care infrastructure and resources [21].
**Achievements and Lessons Learned**

Despite the abovementioned challenges, there are many successes and lessons learned. First, the EMR countries that succeeded in facing COVID-19 had high-level political commitment and high-level multisectoral coordination. Often, the heads of state or prime ministers led the multisectoral committees or crisis committees, and this is an example of leadership. Also, regional laboratories were quick to mobilize and build on previous infrastructure and systems, such as those dedicated to influenza, polio, and other communicable diseases [10, 22]. Polio response teams were used in the field and played major roles in COVID-19 vaccination and community mobilization [22]. Before the COVID-19 pandemic, the region had expanded its influenza network due to the influenza pandemic potential. The WHO supported the strengthening of influenza surveillance and testing for respiratory infections. As a result, most countries had the capacity to test for respiratory viruses in a timely manner, allowing the successful use of reverse transcription polymerase chain reaction (RT-PCR) tests for SARS-CoV-2 testing in influenza labs and national influenza centers [23]. RT-PCR capacity for SARS-CoV-2 was expanded quickly to subnational levels across all 22 EMR countries. The WHO ensured the quality of SARS-CoV-2 testing by encouraging all countries to participate in the external quality assessment program for national and subnational laboratories. The WHO’s logistics hub in Dubai, UAE, which is the WHO’s largest repository of medical equipment and supplies globally, was also an asset to the region as it moved thousands of tons of supplies to the world and the region, including millions of RT-PCR tests [24].

Due to a lack of high-quality data, epidemiologists had to consider a variety of indicators [25], including testing rates and measuring the burden on health systems-related indicators, such as hospital occupancy, intensive care unit occupancy, bed occupancy rates, and case fatality ratios. These indicators allowed the WHO to estimate the extent of the pandemic and the COVID-19 response. A very high case fatality ratio could mean that the country is not reporting all cases, has a weak testing capacity or strategy, is poorly managing cases, or may have different case fatality definitions. Therefore, it is important to not only report data but to also interpret it for evidence-based policy making [25].

Limited testing capacity, supplies, and infrastructure were other reasons countries did not provide accurate data [26]. For example, Somalia and Djibouti did not have RT-PCR capacity at the beginning of the pandemic; however, they managed to procure the needed equipment [27, 28]. Enhanced data reporting was significantly increased due to improved laboratory capacity in the region, both for testing and sequencing. The pandemic was an opportunity to build capacity in genome sequencing, enabling the monitoring of the circulation of SARS-CoV-2 variants. Almost 15 countries in the region have genome sequencing capacity and are able to support other countries without such capacity [16, 29].

The pandemic resulted in several innovations, including apps, telemedicine, hotlines, and e-clinics. More countries in the region used innovative solutions to improve data collection, analysis, and dissemination to build strong surveillance capacities and bridge information gaps. For example, using phone companies’ mobility data to measure whether social interventions are successfully implemented allowed for a better understanding of people’s mobility patterns [30]. Another innovation included oxygen production and supply, as some countries, such as Somalia, had no oxygen plants and no capacity to produce oxygen [31]. Since Somalia did not have electric power, especially at the subnational level, they built small solar plants and used solar power to generate oxygen [31]. Similarly, there was a lot of stigma around mental health in the region [32]. However, during the pandemic, 17 countries (Afghanistan, Bahrain, Egypt, Iran, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Pakistan, Palestine, Qatar, Saudi Arabia, Tunisia, UAE, and Yemen) incorporated mental health and psychosocial support within their emergency response plans by establishing hotlines, e-clinic consultations, and platforms to serve the remote areas [33-35].

A total of 17 EMR countries (Bahrain, Egypt, Iran, Iraq, Jordan, Kuwait, Lebanon, Oman, Pakistan, Palestine, Qatar, Saudi Arabia, Sudan, Syria Arab Republic, Tunisia, UAE, and Yemen) banned waterpipe smoking outdoors; however, this was temporary [36]. In terms of the next steps and moving forward, the region needs to take advantage of these innovations and opportunities to enhance preparedness because they are feasible if there is the right commitment and policies.

There was a need for epidemiological modeling to predict the potential impact of public health and social measures, so the WHO established a modeling support team and used data from countries that requested this assistance to help model the progression of the pandemic [37]. COVID-19 also increased interest in epidemiological studies on the ground, including seroprevalence studies to understand population immunity levels or risk factors. The WHO provides standard protocols for seroprevalence studies [38]. The WHO also worked with countries on vaccine effectiveness studies, and many countries took part in clinical trials on vaccines and therapeutics. Through this, countries built their capacities, which is helpful for preparedness [16].

The positive side of the COVID-19 pandemic was that it revealed many of the gaps and weaknesses in the public health system and how to bridge these gaps [7]. The gaps include the nonflexibility of the health systems, workforce shortages, health service fragmentation (primary care, secondary care, and public health), and designs of health facilities. Additionally, misinformation was a significant challenge during the pandemic, and the outdated “disease model” adopted by most health systems does not meet the current needs of the population [39]. Indeed, access and use of essential health services such as maternal and neonatal care and routine immunization were highly affected [40]. The WHO conducted 3 surveys [41, 42] to measure the global disruption of essential health services. The first, conducted between May 2020 and July 2020, showed that the EMR was the WHO region with the highest disruption in health services due to health center closures, stigma, and fear of transmission [41]. The WHO worked closely with countries and published guidelines to address the impact of the pandemic.
Community engagement and community trust are important and lessons learned that will benefit the EMR include the recommendations [47], but some of the key recommendations and lessons learned that will benefit the EMR include the following:

1. The importance of political leadership is what made a big difference in the countries that did better than others.
2. Community engagement and community trust are important to prevent the pandemic’s spread and major economic collapse.
3. “No one is safe until everyone is safe” is not just a slogan, because if any country does not implement prevention measures, the virus can travel to other areas of the world.
4. The return on investment in health security is immense. How long does it take to receive a return on your investment? How much do we need to invest in pandemics?

There were estimates done, and an early estimate of the cost of the pandemic was about US $11 trillion up to the middle of 2021; another US $10 trillion has been added since then. However, if you invest in preparedness, it will cost US $5 per person per year, so for a global population of 8 billion, it will only cost US $40 billion. To prompt action from policy makers, it is essential to present them with these monetary figures.

**Reshaping Public Health**

During the COVID-19 pandemic, different reasons have called on the EMR countries to reshape their public health. Many countries in the EMR are experiencing new dynamic population growth due to birth, migration, and aging populations. Additionally, the region is facing what is termed a “brain drain.” For example, Egypt has a shortage of doctors because 65% of Egyptian doctors leave the country to seek opportunities abroad. Additionally, the world is witnessing continuous technological advances in the biological, physical, and digital spheres [48].

The burden of the disease in the region should also be considered. The Institute of Health Metrics studied the number of deaths per 100,000 in 2019, per disease per country, and showed that besides the known causes of death such as communicable disease, noncommunicable disease, etc., the region has witnessed other causes of mortality due to violence, mainly in Iraq, Syria, and Yemen. Accordingly, in the region, the leading cause of death from 2000-2019 was related to collective violence and interventions [49].

Furthermore, there is no specific data on public health personnel in the region. Data are mainly available on medical health workers, and Saudi Arabia and Sudan are the only 2 countries with registered public health professionals in the Arab world. On the other hand, only a small proportion of the public health workforce (public health consultants) are qualified in public health but without structured training. Therefore, the public workforce in the region needs to be tackled in terms of skills, experience, and foundational knowledge [50].

Investments in primary health care to mitigate the risks of future pandemics and to maintain accessibility and delivery of essential health services during emergencies [40]; investment in the health workforce, including training, mobilization, and redistribution to sustain high-quality essential health services delivery; and data management and surveillance are key elements of a successful response [16]. These actions and initiatives can also expand and reinforce health system capacities, providing an opportunity within the COVID-19 response for countries to reshape their health workforce and services and improve health security for future health emergencies [7,40,51].

Integrating public health into primary health care is an essential approach to reshaping public health and achieving preparedness. Primary care is the first point of contact for the community with the health system; it must be available 24 hours a day and should provide services in a continuous, personalized, and holistic way. Unfortunately, in many settings, primary health care focuses on treating the illness rather than preventing it. Therefore, integrating public health functions into the primary health care system is highly significant to ensure disease prevention, health promotion and protection, and a proper response to threats [39].
Accordingly, 6 models were identified by the WHO technical series on primary care called “Closing the gaps between public health and primary care through integration” to attain the integration of public health into primary health care and provide a tool to help countries be prepared during threats and emergencies. This, in turn, focuses the services on the population’s needs, achieving a person-centered approach. These models can be applied either individually or in combination, depending on the flexibility of the health systems, and they are titled as follows [39]:

1. Public health services are integrated into primary care
2. Public health professionals and primary care providers are working together
3. Comprehensive and proactive benefit packages that include public health
4. Primary care services within public health settings
5. Building public health incentives in primary care
6. Multidisciplinary training of primary care staff in public health

With a focus on the first model, integrating public health professionals into primary care, where they are involved in many public health functions, can be easily achieved progressively. It is essential to understand that the primary health team is complex and should not only include family medicine and freshly graduated doctors; rather, it should include the public health workforce, such as Field Epidemiology Training Program professionals. Additionally, building the competencies of the public health workforce is crucial because, even though the region has 2 public health academies—the International Academy of Public Health and Weqaya Public Health Academy—most of the workforce is untrained [39,50].

Many aspects need to be considered when reshaping public health: (1) financial allocation and establishing an independent national body for public health; (2) investing in public health laboratories, whole genome sequencing, public health analysis such as artificial intelligence, real-world data, real-world evidence, research, and people; and (3) teaching for precision public health where people are the center of health.

A hard lesson learned from this pandemic is that countries should not be dependent on other countries to provide them with essential medicines and vaccines. It is of great importance to strengthen health systems and work toward universal health coverage and health security, as Dr Tedros Adhanom Ghebreyesus, the Director General of the WHO, says that “Health security and global health coverage are two sides of the same coin” [32]. Capacity building is the way forward, including capacity building in integrated disease surveillance and in ensuring the continuity of health services. Additionally, engaging all relevant stakeholders, accelerating vaccine rollout, prioritizing COVID-19 response, and investing in emergency preparedness and the health systems are essential.

**Conclusion**

The political instability and fragile health systems in some of the EMR countries have hampered the effectiveness and efficiency of the strategies adopted to combat the COVID-19 pandemic. The EMR IMST for COVID-19 pillars was critical to the WHO’s role in coordinating the response during the pandemic. Although multiple challenges affect the transmission of the virus in this diverse region, there were many successes, and it is of great importance to build on these successes and focus on building the human and regional capacities as a way forward. Furthermore, focusing on public health is a key factor in responding to pandemics and biological threats, with COVID-19 being a clear example. The current health system faces many gaps and challenges, which can be overcome by adopting different approaches—specifically, integrating public health into primary care as an essential approach to reshape public health in the region and be prepared against threats and emergencies.

**Recommendations and Areas of Improvement**

Several areas of improvement need to be taken into consideration at both the national and regional levels to improve the response to future threats and pandemics. Countries should develop and update a multisectoral emergency preparedness plan and enhance government and political leadership capacity toward biological threats. They must strengthen their health systems and work toward universal health coverage and security. This can be achieved by integrating public health into primary care as an essential approach to reshaping public health coverage and security. This can be achieved by integrating public health into primary care as an essential approach to reshaping public health through adopting 1 or more of the 6 models of integration identified by the WHO. Moreover, there is a need to invest in building human capacities, including epidemiologists; emergency responders; community health workers; health economists; communication specialists; and most crucially, health leaders. Countries should also work toward community engagement and community trust by assessing people’s needs and engaging them in the decision-making process because public health is about people, for people, and by people. Finally, analyzing and interpreting collected data and using it by policy makers is essential for action and decision-making.

On the other hand, different actions need to be taken at the regional level to effectively control the spread of the pandemic. As the COVID-19 pandemic demonstrated, certain countries have greater capacities than others in the region and must facilitate cooperation, solidarity, and support. High-income countries, for example, should ensure vaccine sharing, equity, and distribution with low-income countries. Moreover, countries in the region can implement twin programs where human resources can be shared across countries.

**Conflicts of Interest**

None declared.


Abbreviations

EMR: Eastern Mediterranean Region
IMST: Incident Management Support Team
RT-PCR: reverse transcription polymerase chain reaction
UAE: United Arab Emirates
WHO: World Health Organization

©Lara Kufoof, Rana Hajjeh, Mohannad Al Nsour, Randa Saad, Victoria Bélorgeot, Abdinasir Abubakar, Yousef Khader, Salman Rawaf. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 15.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Untapped Potential of Unobtrusive Observation for Studying Health Behaviors

Abstract

Improving the environment is an important upstream intervention to promote population health by influencing health behaviors such as physical activity, smoking, and social distancing. Examples of promising environmental interventions include creating high-quality green spaces, building active transport infrastructure, and implementing urban planning regulations. However, there is little robust evidence to inform policy and decision makers about what kinds of environmental interventions are effective and for which populations. In this viewpoint, we make the case that this evidence gap exists partly because health behavior research is dominated by obtrusive methods that focus on studying individual behavior and that are less suitable for understanding environmental influences. In contrast, unobtrusive observation can assess how behavior varies in different environmental contexts. It thereby provides valuable data relating to how environments affect the behavior of populations, which is often useful knowledge for effectively and equitably tackling population health challenges such as obesity and noncommunicable diseases. Yet despite a long history, unobtrusive observation methods are currently underused in health behavior research. We discuss how developing the use of video technology and automated computer vision techniques can offer a scalable solution for assessing health behaviors, facilitating a more thorough investigation of how environments influence health behaviors. We also reflect on the important ethical challenges associated with unobtrusive observation and the use of these emerging video technologies. By increasing the use of unobtrusive observation alongside other methods, we strongly believe this will improve our understanding of the influences of the environment on health behaviors.

Keywords

health behavior; environments; context; unobtrusive observation; video technology; computer vision

Introduction

It is now widely recognized that features of the environment (eg, green spaces, transport systems, and land use) can shape human behavior [1]. This has led to increasing research and policy interest in the idea that environmental change can be used as an upstream intervention to influence health behaviors, such as physical activity, diet, smoking, and alcohol consumption [2]. Despite the intuitive appeal of this idea, there is a shortage of robust evidence on the effects of environmental interventions [3,4]. A key reason for this is that studies often rely on obtrusive methods for measuring health behavior, which require direct elicitation of information from participants through measures such as self-report (eg, questionnaires), wearable devices (eg, accelerometers), and clinical indicators of behavior (eg, heart rate).

In this viewpoint, we argue that increased use of unobtrusive methods, where measurement does not involve the elicitation of information from participants, is needed to accelerate progress in understanding how environments influence health behaviors. We make the case for unobtrusive observation and discuss the opportunities and ethical challenges associated with the application of video technology and automated computer vision techniques, which could unlock the untapped potential of these underused methods.
Limitations of Obtrusive Methods for Studying Health Behaviors

Studies on human behavior are dominated by traditional obtrusive methods that focus on understanding individual behavior, often overlooking the broader environmental context. For example, typical studies examining interventions to increase physical activity involve interventions delivered to individuals (eg, in primary care) and rely on obtrusive methods to measure physical activity (eg, self-report, pedometers, and accelerometers) [5]. These studies typically focus on individual behavior irrespective of location, rather than understanding how populations behave in environments, that is, place-based behaviors. However, if we are to effectively and equitably tackle population health challenges such as obesity and noncommunicable diseases, interventions must focus on changing the environmental determinants in populations instead of trying to change each person individually [6].

Research on 3 widely studied health behaviors (alcohol, smoking, and diet) illustrates the importance of focusing on environments. Studies using unobtrusive objective data on sales of alcoholic beverages [7], cigarettes [8], and unhealthy food [9] have tested the effects of policy and environmental interventions, often through evaluating “natural experiments,” that is, real-world interventions where the researcher has no control over the design and delivery of the intervention [10]. Fiscal measures (eg, taxation), policy and legislation (eg, smoke-free policies), and environmental changes (eg, food advertising) are all examples of population-level interventions for changing health behaviors that have been found to work quickly and are cost-effective based on studies using unobtrusive measures [11,12]. These studies have explicitly focused on how changes in environments affect population outcomes (eg, overall sales), rather than examining changes in individuals. However, there remains a skeptical attitude toward these kinds of population-level studies, reflecting the belief that associations on an individual level better reflect “true” causal relationships than those on a population level.

Studies of individuals using traditional obtrusive methods have also methodological biases, which arise because they rely on eliciting information from humans, all of whom have constraints in terms of time, attention, and capabilities. Even before a study has begun, the lengthy and burdensome recruitment process typically stretches from identifying and contacting potentially eligible participants, through eligibility assessment, to obtaining informed consent [13]. This substantial burden on participants reduces response rates and increases attrition, therefore producing sampling bias. Moreover, people from already disadvantaged populations (eg, ethnic minority groups and people with low literacy levels) are more likely to be deterred at each stage, albeit unintentionally [14]. This differential recruitment and attrition threatens the generalizability and equity of research findings.

Even if researchers manage to recruit individuals who are both willing and able to participate in research, obtrusive methods are prone to measurement reactivity. Reactivity effects of research participation include change due to being assessed, having views about the desirability of different possible research requirements, and deliberately or unwittingly trying to satisfy researchers [15]. Self-report, which has long been the dominant method for measuring human behavior, is particularly vulnerable to reactivity because it relies on introspection. Therefore, asking participants to self-report their behavior can lead to response biases from memory recall, cognitive difficulties, and social desirability [16]. Although these various biases are well known, researchers often overlook the extent to which research studies are unusual contexts and that participants may react in unexpected ways to what researchers ask them to do [17]. These biases, resulting from “research participation effects,” have the potential to affect study outcomes in ways that undermine the validity and representativeness of research findings.

Making the Case for Unobtrusive Observation

Unobtrusive methods have long been recommended to avoid these issues associated with humans taking part in research. In their influential book in 1966, Webb et al [18] argued that researchers rely too heavily on traditional obtrusive measures of data collection. They advocated for greater triangulation using both obtrusive (reactive) and unobtrusive (nonreactive) methods together to provide reassurance that research is robust to the different types of bias associated with each method of measurement. Webb et al [18] described four categories of unobtrusive methods: (1) physical traces, (2) archives, (3) simple observation (observing in natural settings), and (4) contrived observation (observing in controlled settings).

The method that we focus on in this paper is simple observation; specifically, nonparticipative observation of human behavior in the context in which it naturally occurs (hereafter referred to as “unobtrusive observation”). We have been involved in over 1000 hours of unobtrusive observation in natural experimental studies of built environment interventions (eg, a new sustainable park) on physical activity and other behaviors (eg, social interactions) [19-21]. Hence, our experiences derive from a positivist approach, producing quantitative data through systematic observation—a structured method of observation using a predefined coding system.

Unobtrusive observation has historically played a crucial role in various fields of study. For example, sociologists such as Whyte [22] and Jacobs [23] have used observation methods to investigate urban life and better understand how people socially interact in public spaces. Similarly, within urban design, unobtrusive observation has been a valuable tool to provide insights for designers to improve the quality of urban landscapes. This is exemplified in Gehl and Svarre’s [24] pioneering work where they observed public spaces and human behavior, and in studies on “desire paths,” which explore informal routes created by individuals seeking shortcuts rather than adhering to designated paths [25]. Ethnographic research on cultures, communities, and social practices also relies heavily on observation methods, often involving researchers immersing themselves in the communities they observe. Additionally, in ethology (the study of animal behavior), prominent researchers such as Lorenz [26] and Chivers and Goodall [27] have
conducted extensive field observations to uncover insights into animal behavior and social structures. In health research, unobtrusive observation has been used to assess a range of health behaviors, including physical activity [28], smoking [29], suicidal behaviors [30], handwashing in clinical settings [31], and social distancing [32]. A common theme across all these observation approaches is that behavior is intricately tied to the environment, and a comprehensive understanding of behavior requires consideration of this contextual influence.

A unique strength of unobtrusive observation, in comparison to other unobtrusive methods, is its ability for fine-grained analysis of variations in health behaviors directly within the environments, or places, in which they occur. This provides us with strong insights into how people’s behavior is influenced by the microenvironments to which they are exposed. As a result, unobtrusive observation is particularly useful for evaluating the effectiveness of environmental interventions aimed at changing health behaviors. For example, Petticrew et al [29] used unobtrusive observation to evaluate the before-and-after impact of a Scottish legislative ban on smoking in public places, which allowed for the assessment of smoking behavior in great detail (eg, quantifying characteristics and behaviors of the smokers or nonsmokers, signage, and positioning of smoking materials) and in different environmental settings (eg, bars, bookmakers, and restaurants). In contrast, other unobtrusive measures, such as archival data, typically involve aggregated measures that make it more difficult for researchers to understand how people are exposed to specific environments of interest.

Observations can be conducted in many public settings, such as green spaces, public highways, shops, and bars, therefore providing real-world contexts for studying health behaviors and the impact of interventions designed to change them. Such studies can provide answers to important questions for policy and decision makers, for example: What kinds of green spaces best encourage physical activity? How can healthier food choices be promoted by changing physical microenvironments (eg, by altering the availability of unhealthy foods)? And how can smoke-free policies in public spaces influence smoking behaviors and secondhand smoke exposure? There is currently a small evidence base for these types of environmental interventions, suggestive of potentially large effects on health behaviors, but with considerable uncertainty and limited understanding of processes by which these outcomes are brought about.

Furthermore, as unobtrusive measures do not require explicit recruitment of participants, observations allow studies in a wide range of populations and settings. Therefore, unlike most traditional research that often fails to recruit participants from underserved groups (typically referred to as “hard-to-reach” groups) [33], using unobtrusive observation can produce valuable evidence in underserved populations where evidence is lacking but the need to improve health is the greatest. This is particularly important given that intervention effectiveness may differ between socioeconomically advantaged and disadvantaged populations [34].

Despite these advantages, unobtrusive observation remains an underused method for studying health behaviors, even though it has been advocated for over half a century [18]. For example, in a recent systematic review of 116 studies that had an explicit focus on how public spaces influence physical activity, leisure activity, and social activity [35], one would expect that unobtrusive observation would be most appropriate because the focus is on the link between the environment (ie, public spaces) and behavior, rather than the person. Despite this, 95 (82%) of the studies included in this review used obtrusive methods, compared to 53 (46%) studies that used unobtrusive behavior observation. More importantly, of the 95 studies that used obtrusive methods, 57 (60%) studies relied on a single outcome measure to assess behavior, mostly relying on a questionnaire. This is particularly problematic because previous research suggests that relying on methods where participants complete measures in nonbehavioral contexts (eg, at home and in laboratories) may underestimate the importance of contextual factors [36]. Therefore, although this systematic review did not compare differences in findings between obtrusive and unobtrusive methods, relying on questionnaires may lead to inaccurate inferences about the relationship between these behaviors and environmental contexts. This example highlights the importance of triangulation between different methods to reduce the risk of threats to validity based on single-measure research. For example, unobtrusive observation is stronger at quantifying place-based behaviors to examine variation between different contexts (where the place is the subject of analysis), while obtrusive measures are necessary for individual-level longitudinal analysis and assessing intrinsic factors (where the person is the subject of analysis).

So why do researchers typically rely on obtrusive methods and overlook unobtrusive observation? An important factor is that unobtrusive observation does not conform to the traditional mainstays of ethical research that prioritize participants’ right to be informed and freely choose to participate in research. A decline in the number of observational studies reported in journals in several fields has been attributed, in part, to the impact of ethical regulation (eg, [37]). Researchers should, of course, always consider the ethical issues involved in the use of unobtrusive measures, balancing wider societal benefits derived from the research against possible harm to participants. Some specific guidelines have been developed to advise on the unique ethical issues raised by the use of unobtrusive observation (eg, [38,39]). These guidelines typically advise that observational research in public settings where those observed would expect to be observed by strangers, and from which no harm could be reasonably supposed to come, does generally not require consent. Nonetheless, researchers must engage with communities to understand any concerns specific to the sociocultural context they are studying and develop contextually specific solutions to minimize the risk of negative responses when conducting unobtrusive observation. For a more detailed discussion of these ethical challenges, see Clark [40].
Another barrier to the use of unobtrusive observation is the need to deploy in-person observers across multiple study sites, which involves substantial staffing, training, observation time, and data entry—all of which limit scalability. It is therefore perhaps unsurprising that researchers often choose the more convenient and familiar option of traditional obtrusive methods, such as questionnaires and surveys, which have become even more accessible with the rapid rise of digital methods.

**Opportunities in Using Video Technology**

With advancements in video technology, observation methods are beginning to use video recordings, which could help address the scalability issues associated with in-person observations and thereby increase the uptake of observational methods. Specifically, cameras can be used to collect video recordings in public spaces, which can then be watched and assessed (“coded”) by a researcher. Using cameras removes the need to recruit, train, deploy, and supervise in-person observers. Therefore, camera-based observations can overcome issues of observer availability, fatigue, and inattention; reduce risks to researchers from working alone and at night for prolonged periods during observations; reduce measurement reactivity associated with the physical presence of observers (JS Benton et al, unpublished data, 2024); and ultimately decrease costs. Furthermore, the ability to pause, rewind, and rewatch footage can improve the reliability of coding (JS Benton et al, unpublished data, 2024) and allow for more in-depth analysis compared with “live” in-person observations.

Although rare, there are examples of camera-based observation research, such as the use of closed-circuit television (CCTV) surveillance to assess bystander behavior in public spaces [41], traffic webcams to assess physical activity [42], drones to assess park use [43], and wearable video devices to assess behavior on sidewalks or streets [44]. The level of unobtrusiveness associated with these various camera-based approaches will depend on the research context. For example, a recent study found no evidence of participant reactivity to the deployment of fixed video cameras in public spaces where there was already existing CCTV surveillance (JS Benton et al, unpublished data, 2024). However, there may be an increased risk of reactivity in public spaces where cameras might be more conspicuous, for example, due to sociocultural norms.

It is difficult to ignore the emergence of new technologies, such as internet of things devices, artificial intelligence, laser tracking, and remote electroencephalography, which are opening up new avenues of unobtrusive measurement of human behaviors [45,46]. For example, researchers are beginning to capitalize on advances in computer vision to use deep learning models (a subset of machine learning) to automatically detect and recognize behaviors within video images. Examples of diverse applications for automated human behavior recognition include analysis of pedestrian behavior and crowds (eg, monitoring social distancing) [47], detecting when a person falls in a health care facility [48], and evaluating sports performance [49]. Developing such models for assessing health behaviors could dramatically reduce the labor, time, and cost needed to collect data at scale, over extended periods, and with increased consistency across video images compared with human observers.

**Ethical Challenges in Using Video Technology**

Capitalizing on these emerging video technologies creates new risks associated with recording images of people in public spaces, rather than just observing them. Privacy, consent, and confidentiality are all important challenges, which are entwined within data protection laws that researchers must comply with when processing video recordings of people in public spaces. Using computer vision models could address issues of privacy by eliminating the need for humans to watch video recordings once the models are developed and validated. However, less is known about the broader ethical and societal implications of this approach. Therefore, further work is required to establish responsible research practices for the use (and nonuse) of these techniques.

We recently attempted to provide recommendations on how camera-based research can be conducted ethically and in line with data protection requirements [50], drawing on our experiences in the United Kingdom of conducting 3 studies using fixed video cameras to assess observable health behaviors in public spaces. Examples of good practice include engaging with local communities to codevelop privacy and cybersecurity solutions to minimize the risk of negative responses; displaying privacy notice signs and participant information sheets to increase transparency and ensure compliance with data protection legislation; having clear reporting procedures in place for any observed illegal activities; and implementing robust cybersecurity measures to prevent personal data from being intentionally or unintentionally compromised (eg, using secure data storage solutions).

However, views on what makes this type of camera-based research ethical or not can change depending on the researcher’s positionality, context, and experience. For example, visual researchers in the United Kingdom are increasingly concerned about heightened ethical scrutiny and regulation [51], whereas in the United States, exemptions under the Code of Federal Regulations 46 allow for certain research activities to bypass extensive ethical oversight. It is therefore important to acknowledge differences in ethical standards across different jurisdictions and physical and sociocultural contexts, which will inevitably evolve over time in response to societal, technological, and cultural changes.

There are also important wider societal debates about the use of cameras in research, particularly concerning CCTV use, given its ubiquity in many urban spaces around the world. While the use of CCTV in research is on the rise [52], there are differences between using CCTV footage as an observational method in research and its broader application for public safety. A recent study explored the acceptability of using CCTV for research on suicide prevention, which found that there were positive public attitudes toward this approach [53]. Further research is needed to examine acceptability in different geographical and sociocultural settings and in other areas of health research.
Conclusions

Understanding how environments influence health behaviors requires a major change in research practices to address our overreliance on obtrusive methods that primarily focus on understanding individual behavior and that tend to overlook environmental influences. Unobtrusive observation can assess how environments affect the behavior of populations; yet despite a long history, it remains an underused method in health behavior research. Capitalizing on video technology and automated computer vision techniques could provide a scalable solution to increase the uptake of these methods. However, we must find a way to ensure that the scientific and societal benefits are maximized while protecting individual rights. By increasing the use of unobtrusive observation alongside other methods, we strongly believe that this will improve our understanding of the influences of the environment on health behaviors.

Acknowledgments

JSB is funded by a Wellcome Trust ISSF Fellowship (204796). The views expressed are those of the authors and not necessarily those of Wellcome Trust.

Authors’ Contributions

JSB and DPF conceptualized the paper. JSB drafted this paper and DPF contributed to the revision of this paper. JSB and DPF read and approved the final version of this paper.

Conflicts of Interest

None declared.

References


https://publichealth.jmir.org/2024/1/e46638

JMIR Public Health Surveill 2024 | vol. 10 | e46638 | p.199

(page number not for citation purposes)


Abbreviations

CCTV: closed-circuit television

©Jack S Benton, David P French. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 21.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Data Sharing in a Decentralized Public Health System: Lessons From COVID-19 Syndromic Surveillance

Ryan C Rigby\textsuperscript{1}, MHA, MBA; Alva O Ferdinand\textsuperscript{2}, DrPH, JD; Hye-Chung Kum\textsuperscript{3}, MS, MSW, PhD; Cason Schmit\textsuperscript{1,3}, JD

\textsuperscript{1}Program in Health, Law, and Policy, Department of Health Policy and Management, Texas A&M University School of Public Health, College Station, TX, United States
\textsuperscript{2}Department of Health Policy and Management, Texas A&M University School of Public Health, College Station, TX, United States
\textsuperscript{3}Population Informatics Lab, Department of Health Policy and Management, Texas A&M University School of Public Health, College Station, TX, United States

Corresponding Author:
Cason Schmit, JD
Population Informatics Lab
Department of Health Policy and Management
Texas A&M University School of Public Health
212 Adriance Lab Rd
College Station, TX, 77845
United States
Phone: 1 9794360277
Email: schmit@tamu.edu

Abstract

The COVID-19 pandemic revealed that data sharing challenges persist across public health information systems. We examine the specific challenges in sharing syndromic surveillance data between state, local, and federal partners. These challenges are complicated by US federalism, which decentralizes public health response and creates friction between different government units. The current policies restrict federal access to state and local syndromic surveillance data without each jurisdiction’s consent. These policies frustrate legitimate federal governmental interests and are contrary to ethical guidelines for public health data sharing. Nevertheless, state and local public health agencies must continue to play a central role as there are important risks in interpreting syndromic surveillance data without understanding local contexts. Policies establishing a collaborative framework will be needed to support data sharing between federal, state, and local partners. A collaborative framework would be enhanced by a governance group with robust state and local involvement and policy guardrails to ensure the use of data is appropriate. These policy and relational challenges must be addressed to actualize a truly national public health information system.

(Keywords: syndromic surveillance; federalism; COVID-19; public health; SARS-CoV-2; COVID-19 pandemic; United States; decentralized; data sharing; digital health; ethical guidelines; risk score; technology; innovation; information system; collaborative framework; infodemiology; digital technology; health information; health data; health policy; surveillance)

Introduction

Dr Rochelle Walensky, director of the Centers for Disease Control and Prevention (CDC), has admitted that the national response to the COVID-19 pandemic was deficient in many respects, citing the use and dissemination of data as a principal problem [1]. Indeed, the US public often relied on private data sources, such as the Johns Hopkins Coronavirus Resource Center, as the “go-to data source” for COVID-19 data rather than official government sources [2]. Many of the challenges concerning the use and dissemination of data are not due to the need to collect missing data, nor due to the technical limitations in sharing the needed data. Rather, many of these challenges are due to policy barriers that impede the needed data flows between public health entities, including local, state, and federal partners [3].

While data sharing challenges exist across the public health surveillance and dissemination systems, here we examine the specific challenges faced when sharing data between state, local, and federal syndromic surveillance partners. These challenges...
have been explored from state perspectives, but we examine these challenges through a national lens and how federalism issues affect federal, state, and local partners and shape the data use agreements (DUAs) that govern the data sharing relationships [4,5].

**US Syndromic Surveillance and Federalism**

Syndromic surveillance refers to the process of sharing electronic data with health departments—often in near real time—to understand existing and emerging public health issues [6]. This process can include diverse data sources, including emergency department electronic health records (EHRs), environmental data, vital statistics, and laboratory data. The rapid nature of syndromic surveillance permits real-time situational awareness of emergent public health issues [6].

The United States began its investment in syndromic surveillance after the terrorist attacks of September 11, 2001, and a rash of anthrax-laced letters [7,8]. In 2009, syndromic surveillance was included as part of the US $20 billion EHRs incentive program created by the Health Information Technology of Economic and Clinical Health (HITECH) Act [9]. The Act incentivized hospitals to adopt, and meaningfully use certified EHR systems. The Act also accelerated syndromic surveillance nationally by incentivizing hospitals to report specific public health measures, including syndromic surveillance, to their local public health authorities [10]. Over this time, the national syndromic surveillance system, which was first developed by CDC, evolved from BioSense to BioSense 2.0 to the National Syndromic Surveillance System (NSSP), with the overall purpose of establishing a nationwide surveillance system to detect and assess potential health outbreaks throughout the United States [7]. Each of these evolutions has been impelled by technical, policy, and relational challenges between federal and state or local syndromic surveillance partners, including CDC, Council of State and Territorial Epidemiologists (CSTE), Association of State and Territorial Health Officials, and National Association of County and City Health Officials.

The community of practice, one of the defining aspects of the NSSP, aims to improve syndromic surveillance nationally through collaboration and knowledge sharing between federal, state, and local partners [11]. The community of practice offers a platform for public health professionals to collectively identify and use the optimal approach in syndromic surveillance. Through professional engagements, members collaborate to exchange knowledge, improve understanding, cultivate expertise, and address issues, to advance the practice of syndromic surveillance [11].

Currently, the NSSP stores all participating state and local syndromic surveillance data in a single repository maintained by the CDC [6]. While some federal employees have access to all syndromic surveillance to provide technical system support for state and local jurisdictions (eg, quality assurance, troubleshooting, and developing analytical tools), these data are not shared by default for federal public health uses [5]. The DUAs between state or local governments and the CDC prohibit any federal access to state and local data for public health purposes without the express consent of the jurisdictions [4,5]. Specifically, the default federal access to syndromic surveillance data for public health purposes is limited to the US Department of Health and Human Services (HHS) region level. For example, while federal NSSP personnel might observe an incidence increase within HHS Region 10 comprising Alaska, Idaho, Oregon, and Washington, they could not distinguish between (1) an isolated event in Washington; (2) unrelated, but similar, events in Washington and Alaska; or (3) related events in Washington and Oregon [5,12]. At the regional level, the federal government lacks awareness of interstate public health events. These policy data sharing limitations effectively hamstrung the federal COVID-19 response by not allowing access to vital data that could have been used to give a broader view to federal agencies in the context of a novel, rapidly moving and evolving virus.

The COVID-19 data sharing challenges cannot be understood without fully appreciating how US federalism shapes and affects the US public health system. In the United States, federalism consists of the sharing of authority between the national (federal) government and the state governments. At the onset, it is important to recognize that there is no express public health power that is explicitly outlined in the US Constitution. This means that federal public health actions must derive from one of the US Constitution’s enumerated powers. Typically, most federal public health actions are derived from the powers to tax and spend for the “general welfare” and the interstate commerce clause (ie, regulation of industries and activities that affect interstate and international commerce, including health).

The police power—the states’ power to regulate in the interest of public health, safety, and community values—is the fundamental governmental power that authorizes nearly all traditional public health actions [13]. Under the Constitution’s 10th Amendment, these powers are reserved for the states. This means that the states have the authority to set quarantines, restrict businesses, mandate isolation for those infected with a communicable disease, and impose primary public health responsibility [14].

In comparison, the federal government’s role is much more limited to providing support to state and local governments (ie, taxing and spending for the general welfare) or addressing those issues that have interstate impacts, like vaccine approvals (ie, interstate commerce). Consequently, state and local governments can be thought to have the primary public health responsibility within their jurisdictions while the federal government has less direct public health responsibilities and interests (ie, supporting state and local governments and interstate commerce).

These federalism nuances and the consequent relationships between local, state, and federal partners came under strain during the COVID-19 response. Specifically, the DUAs between the CDC and the states effectively impeded the response to the COVID-19 pandemic by restricting federal access to HHS region level syndromic surveillance data [15]. In early 2020, the White House COVID-19 Task Force obtained access to all COVID-19 NSSP data in an apparent exercise of emergency authority [4,5]. Although the federal government was able to bypass these policy...
barriers, the decision left some state and local epidemiologists feeling that the agreed-upon DUAs were “thrown out the window” [5].

While examining the US legal framework for data sharing within US federalism, Fahey [16] writes that data sharing is a rapidly expanding intergovernmental marketplace in many areas of government, including public health. She argues that in the absence of federal legislation, documents such as DUAs are the strongest legal policies [16]. The consequences and the reality of DUAs being discarded in the event of a national emergency are something that will only become more complex as data collection and data sharing increases. The power that federalism gives the states in these circumstances may become harder to define if clear policy regarding data is not codified through binding statutes.

**Improving the National Public Health Data System for Future Epidemics**

In the following sections, we derive important lessons learned about public health data sharing within the US public health system, derived from the findings from a 2021 study by CSTE and the NSSP [4,5]. The study involved several work group calls with 20-30 state and local epidemiologists in leadership or decision-making roles and 8 key informant interviews. In addition, the study cited 8 randomly selected state and local epidemiologists from the list of NSSP site administrators to provide feedback on specific policy options [5].

**Align US Public Health Data Sharing Policies With Ethical Guidelines**

An increasing number of public health ethicists assert that there are ethical obligations to share public health surveillance data in certain circumstances [17-19]. For example, the World Health Organization’s (WHO) ethical guidelines expressly state [18]:

> [w]ith appropriate safeguards and justification, those responsible for public health surveillance have an obligation to share data with other national and international public health agencies.

Surveillance data can be legitimately disseminated to enable public health response, improve the efficiency and effectiveness of public health activities, and inform resource allocation and other support [20].

By limiting federal access to state and local NSSP data by default, the DUAs between the CDC and the states create significant impediments to legitimate public health functions under the enumerated powers to regulate interstate commerce and to spend for the general welfare [13]. In the context of a rapidly spreading public health threat such as the COVID-19 pandemic, the burden on the federal government to seek consent from all affected jurisdictions, each of which are also burdened by local public health actions, is antithetical to the calls for increased public health data sharing from ethicists. These DUAs should be revised to eliminate data sharing barriers when there is a documented and communicated public health need and there are appropriate policy guardrails to ensure only appropriate public health uses them. Importantly, proper guardrails can also eliminate other data sharing barriers [21].

**Using Granular Data to Respond to Interstate Threats and Allocate Federal Resources Equitably**

The NSSP DUAs that restrict federal access to state and local syndromic surveillance data impede several legitimate federal interests and national public health objectives. The DUAs permit routine federal access only to HHS region level aggregations [5]. In order to have a national view of emerging public health threats, access to data at the state level or at a more granular level is required.

A national view of public health threats is essential for the federal government to fulfill its legitimate governmental role in national public health responses [5]. The federal government plays a critical role in providing support and allocating resources to state and local jurisdictions. However, allocating these resources equitably in response to national public health events requires broad situational awareness of the burdens faced by all jurisdictions. This type of assessment cannot occur when the federal government lacks the data at the required level of granularity. This proved to be one of the major challenges during the COVID-19 pandemic [22].

Moreover, several state and local epidemiologists expressed openness to increased federal access to state and local NSSP data to enhance their public health actions. For example, a commonly expressed benefit among key informants was having extra “eyes” on data to provide greater detection capacity. One epidemiologist stated, “I do think there’s also a lot of opportunity right now, [but] there’s not enough capacity...to look at local data” [5]. Other articulated anticipated benefits were more coordination between agencies and the generation of regular reports or visualizations based on agreed-upon queries of state and local NSSP data [5]. Additionally, enabling greater access to the CDC could permit the creation of national training programs to train and support a growing number of state and local epidemiologists in syndromic surveillance methods and techniques.

**State and Local Agencies Remain Essential Partners in Public Health Surveillance Activities**

Regardless of legitimate federal interests, state and local governments still need to play a central role in syndromic surveillance for several reasons. First and foremost, state and local governments retain the primary public health responsibility for the communities within their jurisdictions. As a result, the duty to use syndromic surveillance data to improve population health in their jurisdictions rests squarely with state and local jurisdictions. This fundamental responsibility implicitly carries an ethical duty to safeguard this sensitive information and to protect the confidentiality of their constituents. This duty has additional significance to the relationships with the health care providers because many health care providers that contribute...
their data do so voluntarily (ie, few states mandate syndromic surveillance). Moreover, state and local governments have a legitimate interest in ensuring the responsible use of these sensitive data.

These fundamental responsibilities and interests do not end when state and local governments share their data to enable federal agencies to fulfill their legitimate public health interests. Federal partners must keep their state and local partners abreast of their use of syndromic surveillance data and any dissemination of findings. There are important risks in interpreting syndromic surveillance data without understanding local contexts. As prediagnostic EHR data are automatically transmitted in near real time, syndromic surveillance data are messy by design. There can be local variations in submission intervals or health facilities’ data entry conventions that can lead to misleading artifacts in data analyses for the unaware. Additionally, local events—such as a festival drawing out-of-state visitors—could result in expected localized spikes in emergency room or urgent care visits that could be misinterpreted as an emerging public health concern if analysts are unaware of the local context [5]. If federal agencies make decisions from syndromic surveillance data without fully understanding the local context, it is possible that federal actions could interfere with state and local public health activities or create unnecessary communication burdens (ie, states forced to explain or reconcile federal data releases).

State and local health agencies also have important, and sometimes fragile relationships with health care facilities that often voluntarily contribute syndromic surveillance data. Throughout the CSTE study, state and local epidemiologists expressed concern that increasing federal use of granular syndromic surveillance data could endanger these facility relationships if federal communications stopped once the data are received from local agencies without further clarification before any decisions or actions based on these data analyses. Disclosure of a facility’s syndromic data also exposes that facility to certain risks. As 1 key informant said [5]:

\[
\text{we have an understanding with the facilities that contribute data that we're not just going to release data from a single hospital to the public. So, sometimes it seems like federal users are not as sensitive to that.}
\]

Facilities in rural areas are particularly vulnerable to these types of risks as disclosures of granular syndromic surveillance data in rural areas may inadvertently expose the sole rural provider in that area. One informant stated [5]:

\[
\text{I think that the [syndromic] data is just a little bit more fraught than other data, where issues with facility level disclosures...speaking poorly...about a community without knowing that you are, and that...cultural awareness...that might not be apparent from a federal level...I think that states can bring a lot of value too.}
\]

While state and local epidemiologists saw benefits to increased federal access to state and local syndromic surveillance data, access to those data comes with risks that are exacerbated without collaboration and cooperation between federal, state, and local partners [5]. For example, federal interactions with health care providers that bypass state and local partners could threaten local public health relationships with providers that often voluntarily contribute their data to syndromic surveillance programs. If so, these federal actions would also interfere with state and local governments’ fundamental public health mission.

**Intergovernmental Relationships are Critically Important to Data Sharing**

Due to the decentralized public health structure inherent to US federalism, strong interagency and intergovernmental relationships are essential for any national public health response. Data sharing between agencies is also critically important in these responses. Indeed, fractures in relationships can be potently data sharing barriers [23]. Relationships are established through time spent communicating and working with each other. Notably, these relationships require trust between partners, and that trust was shaken during the COVID-19 response [5]. One nonfederal key informant noted [5]:

\[
\text{There's been such a lack of trust that has been reinforced during this response. I think it's actually going to be harder rather than easier [to permit greater federal access to state or local data]. I say that anyway because I think the NSSP program itself, in its current form, and I think it's probably important that this gets documented, has been an amazing steward of the data, but the system around it has become less trustworthy and I think the system around it and the system, the way that the response has worked with the states, is now going to impact the program's ability to do its best work. So, in today's world, CDC has become less and less willing to really talk to states in pre-decisional ways and help states understand this data is driving this decision, and there's been a much larger tendency for CDC to make decisions and then just inform states about it in this response. And so, I think pre-COVID, it actually would have been easier rather than harder to implement some of these changes right now, in a way that the states felt good about.}
\]

Enabling a truly national public health information system within the context of US federalism requires federal, state, and local partners to forge stronger relationships that will enable greater use of public health data. Done well, collaborating around the use of public health data has the potential to increase communication, common understanding, and teamwork toward a common goal that could increase trust. Importantly, however, public health politicization likely creates additional data sharing barriers, as some governments equated the COVID-19 pandemic with successful or unsuccessful governance during a global crisis [24].
Establish a State and Local Syndromic Surveillance Governance Group

Several state and local epidemiologists suggested openness to creating an NSSP governance group with state and local constituents [5]. Such a governance group should have some decision-making capacities that are currently absent from the community of practice, which functions more as a sounding board or advisory council to the CDC. A governance group with some decision-making authority would create an important check on the expanded federal access to the state and local syndromic surveillance data.

A governance group could have several different benefits for national syndromic surveillance efforts. First, a governance group would be well-positioned to establish a collaborative framework and norms between the federal, state, and local partners. Second, a governance group could establish protocols to help improve transparent communications and collaboration with federal partners regarding ongoing or proposed uses of state and local syndromic data. Third, a governance group could help flag important issues for other state and local partners to monitor, such as proposed federal publications, communications, or policy changes. Finally, a governance group could be empowered to facilitate emergency access to appropriate state and local syndromic surveillance data, greatly reducing the current individual jurisdictional consent burden required in a public health emergency such as the COVID-19 pandemic [5].

The DUAs and other applicable laws should be revised to enable the governance group to function nimbly as public health circumstances necessitate. For example, DUAs could bind state, local, and federal partners to decisions by the governance group to permit expanded emergency access to state and local syndromic surveillance data (and restrict access when emergencies end). This would build flexibility into the DUAs to anticipate and accommodate emergency situations reducing transactional friction while all parties are managing a crisis.

Develop a Collaborative Framework Between Public Health Surveillance Partners

The US federalism created a decentralized public health system, so policies and agreements establishing a collaborative framework are needed to create a national public health information system, while still recognizing state and local authorities as the primary public health authority. Indeed, the CSTE study identified “improved cross-jurisdiction collaboration efforts” as an important benefit of federal access [5]. This framework would have to carefully establish expectations for federal, state, and local partners and ensure that the legitimate public health interests of all parties are respected.

The CSTE study identified several critical issues that should be addressed in this collaborative framework [5]. Perhaps the most important issue is establishing protocols for communications between syndromic surveillance partners. Currently, communications are not standardized in mode or content. Furthermore, it is not always clear to state and local health departments when the federal government expects responses to their communications (eg, an “FYI” communication vs an investigation inquiry; see Table 1).

Table 1. Example of a tiered approach to federal National Syndromic Surveillance System communication response expectations.

<table>
<thead>
<tr>
<th>Tier</th>
<th>Public health threat</th>
<th>Expected response from states</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier 3</td>
<td>Low or moderate but only affecting targeted jurisdiction</td>
<td>None</td>
</tr>
<tr>
<td>Tier 2</td>
<td>Moderate but interjurisdictional in nature or high but only affecting targeted jurisdiction</td>
<td>Acknowledge receipt</td>
</tr>
<tr>
<td>Tier 1</td>
<td>High and interjurisdictional in nature</td>
<td>Response is expected</td>
</tr>
</tbody>
</table>

Given the division of public health responsibilities within the US federalism, these communication ambiguities are additional strains on thin public health resources [5].

A collaborative framework must also (1) establish processes and support for state and local involvement in data analysis and methodological development, (2) provide appropriate acknowledgment, (3) standardize data requests, and (4) restrict federal syndromic surveillance communications with participating health care facilities without the consent of relevant state or local public health partners.

Establish Appropriate Federal Policy Guardrails for Syndromic Data That Are Shared With Public Health

While many state and local epidemiologists who participated in the CSTE study indicated benefits to greater data sharing with federal partners, many called for strong policy guardrails to ensure appropriate federal uses [5]. To be clear, there are substantial restrictions on syndromic surveillance data currently, but these restrictions prevent legitimate public health data use. If these existing restrictions are loosened to allow legitimate public health uses, then new guardrails need to be established to prevent data misuse and manage risks while still enabling suitable data sharing between agencies. Critically, these sentiments were shared by federal participants in the CSTE study [5]. State and local epidemiologists in the CSTE study indicated support for policies that would (1) establish audit and documentation processes, (2) implement standards and processes to remove access from federal users, and (3) protect sensitive data from public disclosure—such as Freedom of Information Act requests [5]. Additionally, there should be strict limits on sharing NSSP data for nonpublic health purposes. For example, the WHO ethical guidelines strongly advocate against sharing data with agencies that are likely going to take law enforcement action against individuals [19]. Similarly, the facilities that voluntarily contribute their data for the betterment of their...
communities should not have to fear that their syndromic data will be used against them. The DUAs that address these guardrails would foster stronger relationships of trust in state and local jurisdictions. The existing DUA did not address these issues considering the proposed expanded federal access to state and local data [15].

As data sharing continues to grow between agencies the policies that govern data use need to have well-defined guardrails. In the absence of statutory protections having these guardrails defined in DUAs becomes even more important. This will ensure data are being used in the manner it was intended.

Significance for Broader Public Health Data Sharing

Data sharing is a persistent challenge for public health agencies [23,25-27]. Although this analysis is limited to the challenges in addressing broader data access to US syndromic surveillance data, it provides insights into data sharing in other contexts. For instance, Aamer Ikram, the executive director of Pakistan’s National Institute of Health recently spoke of the need to support coordinated and integrated public health surveillance, noting that the lessons of the COVID-19 pandemic—good and bad—“must be immediately translated into strategies and policies” [28]. In a 2022 study, The International Association of National Public Health Institutes (IANPHA) identified several jurisdictions where decentralized public health surveillance has complicated the data sharing, coordination, and development of integrated public health data systems [29]. For instance, Mozambique, Canada, and Pakistan all face challenges with local, provincial, and national public health partners that complicate efforts to create integrated surveillance systems [30,31]. The IANPHA report finds that optimal data flow requires systems that permit local and provincial data to flow into compatible national infrastructures supported by formalized data sharing agreements [29]. Accordingly, lessons from this analysis—that deal with facilitating similar data sharing between US local, state, and federal public health partners—could be useful in other international settings.

Limitations

Our analysis derives from insights from national, state, and local informants based on their experiences in syndromic surveillance practice before and during the pandemic. By the CSTF study’s design, these perspectives skew in favor of the US state and local epidemiologist informants. Moreover, many of the data in the CSTF study are qualitative, which have inherent limitations (eg, representativeness). In our analysis, we sought to consider additional national and systemic considerations, but these additional inputs likely do not eliminate the state or local perspective biases implicit in the CSTF study data used in our analysis. Additionally, there could be additional legal, political, and practical considerations that were not observed within the scope of our analysis.

Conclusions

The history of US syndromic surveillance is pendular. Every iteration of US syndromic surveillance (ie, Biosense 1.0, BioSense 2.0, and NSSP) was impelled by existing challenges and concerns [7]. However, imperfect responses to those challenges created new issues [7]. For example, to support collaboration between states, BioSense 2.0 shielded syndromic surveillance data from CDC access by using servers operated by the Association of State Health Officials [32,33]. However, this prevented the CDC from assisting with quality assurance and technical assistance [7]. The development of NSSP addressed this issue by having CDC systems once again host syndromic surveillance data but governed by DUAs that established a default of nonsharing between federal, state, and local partners [4,7]. The federal challenges to accessing state and local syndromic surveillance data suggest that the iteration of US syndromic surveillance data will have to wrestle with the policy barriers to intergovernmental public health data sharing.

Broadly, the harsh realities exposed by the COVID-19 pandemic impelled new efforts to improve the national public health data infrastructure. Chief among these efforts is the “Data Modernization Initiative” [34,35]. The CDC’s strategic plan outlines 5 key priorities that are, building the right foundation, accelerating data into action, developing a state-of-the-art workforce, supporting and extending partnerships, and managing change and governance to support new ways of thinking and working [35]. These are important and laudatory goals; however, it is not clear that they will address the challenges that are inherent in the US public health system in the context of federalism.

Future public health challenges such as the COVID-19 pandemic will force the distinct components to once again provide a national response in a decentralized system. State and local governments—having the primary public health responsibility within their jurisdictions—will once again be tasked with the responsibility of collecting data. Federal partners—requiring a national perspective—will once again seek access to these data. Absent a new collaborative framework and data sharing policies, federal agencies seeking access to the needed data will be forced to navigate the governmental bureaucracies of 50 states and hundreds of localities one DUA negotiation at a time. This creates an enormous transactional burden on a public health imperative [19].

Notably, there are substantial efforts to fix these problems. The Coronavirus Aid, Relief, and Economic Security (CARES) Act awarded US $500 million to accelerate data modernization. Part of that funding was intended to upgrade surveillance systems throughout the United States [34,35]. This funding follows the HITECH Act which heavily incentivized public health reporting through the US $35 billion meaningful use program [36]. However, it is here that we must note that these challenges cannot be fixed with new technology or funding alone.

The NSSP is a highly sophisticated public health information system in comparison to other surveillance systems. Nearly all reporting is automated. Reports are transmitted in near real time. Nearly all jurisdictional data are sent to the same data repository.
The data sharing challenges are not technical or practical. The data sharing challenges derive solely from relationships between government agencies and policies. These are the challenges that must be addressed to actualize a modern public health information system benefiting the nation as a whole.

Conflicts of Interest
None declared.

References
15. Centers for Disease Control and Prevention. National Syndromic Surveillance Program data sharing and use agreement. CDC. 2018. URL: [https://docs.google.com/document/d/1JB9V0l6Pv1TXrhaOXYv7EBaI9XR-n/edit][accessed 2022-02-10]


Abbreviations

- **CARES**: Coronavirus Aid, Relief, and Economic Security
- **CDC**: Centers for Disease Control and Prevention
- **CSTE**: Council of State and Territorial Epidemiologists
- **DUA**: data use agreement
- **EHR**: electronic health record
- **HHS**: US Department of Health and Human Services
- **HITECH**: Health Information Technology of Economic and Clinical Health
- **IANPHA**: International Association of National Public Health Institutes
- **NSSP**: National Syndromic Surveillance System
- **WHO**: World Health Organization
Responding to the Return of Influenza in the United States by Applying Centers for Disease Control and Prevention Surveillance, Analysis, and Modeling to Inform Understanding of Seasonal Influenza

Rebecca K Borchering, PhD; Matthew Biggerstaff, ScD; Lynnette Brammer, MPH; Alicia Budd, MPH; Shikha Garg, MD; Alicia M Fry, MD; A Danielle Iuliano, PhD; Carrie Reed, DSc

1National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States
2Fulton County Board of Health, Atlanta, GA, United States

Corresponding Author:
Rebecca K Borchering, PhD
National Center for Immunization and Respiratory Diseases
Centers for Disease Control and Prevention
1600 Clifton Rd
NE MS H24-7
Atlanta, GA, 30329
United States
Phone: 1 4046395214
Email: xhq2@cdc.gov

Abstract
We reviewed the tools that have been developed to characterize and communicate seasonal influenza activity in the United States. Here we focus on systematic surveillance and applied analytics, including seasonal burden and disease severity estimation, short-term forecasting, and longer-term modeling efforts. For each set of activities, we describe the challenges and opportunities that have arisen because of the COVID-19 pandemic. In conclusion, we highlight how collaboration and communication have been and will continue to be key components of reliable and actionable influenza monitoring, forecasting, and modeling activities.

(JMIR Public Health Surveill 2024;10:e54340) doi:10.2196/54340

KEYWORDS
disease burden; modeling; seasonal influenza; surveillance

Context
The COVID-19 pandemic disrupted the typical timing, intensity, and duration of influenza virus activity in the United States and many parts of the world [1,2]. These disruptions are likely due in part to the adoption of nonpharmaceutical interventions, such as mask wearing and physical distancing, which were deployed to prevent SARS-CoV-2 transmission and likely also reduced transmission of influenza viruses and other respiratory viruses [1]. Reductions in the use of nonpharmaceutical interventions in the United States during the 2022-2023 influenza season compared to what was observed during the 2020-2021 and 2021-2022 seasons may have contributed to a return of influenza virus activity more comparable to levels observed during some prepandemic seasons, which had annual estimates of 9 million-41 million illnesses, 140,000-710,000 hospitalizations, and 12,000-52,000 deaths between 2010 and 2020 [3]. The interim estimated burden of influenza for the 2022-2023 influenza season indicated that between 27 and 54 million illnesses, 300,000 and 650,000 hospitalizations, and 19,000 and 58,000 deaths occurred that season [4]. The 2022 and ongoing 2023 influenza seasons observed in the Southern Hemisphere winter (approximately May to September) were characterized by the circulation of multiple seasonal influenza viruses (ie, influenza A/H3, A/H1, and B Victoria) [5]; atypical timing patterns, including early and late seasons; or increased intensity compared to recent seasons during the COVID-19 pandemic. However, changes in care-seeking behaviors, testing practices, and general reporting practices due to the pandemic complicate the interpretation of influenza surveillance data.

The experience of the Southern Hemisphere in 2022 and 2023 [5] and the atypically early and rapid start to the 2022-2023
influenza season in the United States [6-8] serve as reminders that seasonal influenza remains a major public health threat. Influenza burden during the 2022-23 season was elevated compared to activity observed in the United States during the first 2 years of the COVID-19 pandemic [4,9], and early and high rates of influenza-associated hospitalizations were observed, particularly in children [7,8,10]. Over the past decade, the Centers for Disease Control and Prevention’s (CDC’s) Influenza Division has worked internally and with external collaborators to develop, validate, and improve forecasting, mathematical modeling, and analytical tools that help translate US influenza surveillance data into action by forecasting short-term influenza activity and characterizing the burden and severity within and across influenza seasons. With the decreased influenza activity observed in the 2020-2021 season, atypical activity and timing in the US 2021-2022 and 2022-2023 seasons, and the Southern Hemisphere’s 2022 and 2023 seasons, continued circulation of SARS-CoV-2, and the return of other respiratory pathogens (eg, respiratory syncytial virus [RSV]) [1,11], we summarize these activities to highlight how these tools can help inform influenza prevention and control efforts during future seasons in the postpandemic respiratory virus landscape.

Data From Surveillance Systems Provide the Foundation for Influenza Situational Awareness and Risk Communication

Multiple surveillance systems provide complementary information to create a comprehensive picture of influenza activity in the United States [6,12]. These information streams are fundamental for situational awareness, risk communication, and subsequent modeling and analysis. Types of surveillance data collated and distributed by the CDC include influenza hospitalizations (hospitalization rates by age, sex, and race or ethnicity, as well as clinical characteristics and outcomes) [9,13]; influenza-related mortality (the National Center for Health Statistics Mortality Surveillance Data and pediatric influenza-associated mortality); virologic (eg, percent positivity, influenza virus type or subtype, and genomic characterization of influenza viruses, which can be used to assess the match of circulating strains to vaccine components); and syndromic outpatient illness [12]. Maintaining these data sources requires coordination within the CDC as well as with international organizations, vital statistics offices, state, and local public health partners in both clinical and laboratory settings, hospitals, and health care providers.

Weekly FluView [6] reports provide systematic data updates and interpretation from the many influenza surveillance systems. These reports, related data updates, and visualizations provide a valuable resource for the CDC, external partners, and the public to understand and interpret influenza activity. Regular and rapid dissemination of surveillance data that is easily accessible and interpretable plays an important role in risk communication and promotion of prevention and mitigation measures, such as vaccination, testing, antiviral use, and nonpharmaceutical interventions.

Given the cocirculation of multiple respiratory viruses with the potential to cause severe illness, interpreting influenza virus data alongside complementary surveillance data for other respiratory viruses provides additional context to the added burden of influenza on the population and the health care system. Symptom-based surveillance for influenza has traditionally been complicated, due in part to overlapping symptomologies. Accordingly, data from the US outpatient Influenza-like Illness Surveillance Network, where outpatient visits are monitored for influenza-like illness defined as fever plus cough or sore throat, can include reports due to infection with cocirculating respiratory pathogens that may be clinically indistinguishable from influenza, including COVID-19 and RSV. In addition, changes in health care–seeking behavior due to the COVID-19 pandemic, both in terms of provider choice and timing relative to illness onset and severity, will be important to consider when assessing surveillance data collected from health care settings. Additionally, the expansion of remote or telehealth options will complicate surveillance, particularly for laboratory-based surveillance. The Influenza Division will continue to leverage the strengths of different surveillance systems to provide a comprehensive understanding of seasonal influenza virus activity. In addition, in order to better understand the burden of influenza in relation to other respiratory viruses like RSV and SARS-CoV-2, the Influenza Division has cosupported the development of multipathogen respiratory virus surveillance networks and developed publicly facing summaries that compare acute respiratory illness emergency department visits for laboratory-confirmed influenza, RSV, and COVID-19 [14] and weekly in-season rates of laboratory-confirmed influenza, COVID-19, and RSV-associated hospitalizations by age, sex, and race or ethnicity [15]. Influenza surveillance and communication efforts will continue to adapt in the uncertain and changing landscape of continued respiratory virus cocirculation.

Estimating Influenza Disease Incidence Using Analytic Methods That Account for the Underdetection of Influenza Provides a More Complete Picture of the Burden and Severity of the Influenza Season

Not all people seek medical care for respiratory illnesses, and not all of those who do will be tested to confirm infection with a specific pathogen. Even when an individual seeks medical care and is tested, this may not occur during the time window of their illness when a collected sample has the potential to test positive (ie, the pathogen or antibodies may no longer be detectable). A variety of viral and bacterial respiratory illnesses are clinically indistinguishable from one another, complicating diagnosis based on symptoms or clinical findings alone. Laboratory testing for specific pathogens can be more or less common in certain care settings, for different age groups, and at differing levels of disease severity. To account for the incomplete detection of influenza virus infection from clinical testing in health care settings, models used to estimate the disease burden incorporate information on health care usage and laboratory testing for respiratory viruses [16]. Each year,
the CDC uses mathematical and statistical models to estimate the broader disease burden from influenza in the US population; since 2018, this has been estimated weekly during the influenza season [17]. In the 2022-2023 season, the first preliminary in-season weekly influenza burden estimates were released in October [4], earlier than any other season due to the atypically high levels of influenza activity observed at that time of year.

The ongoing effects of the COVID-19 pandemic on health care usage and testing for respiratory viruses [18], including influenza, are unclear and will likely continue to evolve over time. Major changes to health care–seeking behaviors, including where patients are seeking care and the use of clinical versus at-home diagnostic testing, could influence trends observed in long-standing surveillance systems, impact ongoing efforts to forecast influenza activity, and introduce new biases into epidemiological and clinical studies that rely on the results of clinical testing. The CDC will continue to make estimates using the best available data, adapt to challenges presented by atypical seasonality, and assess developments in care-seeking behaviors and influenza testing [19]. Communicating in-season and end-of-season burden estimates through the CDC website and FluView reports will help promote public awareness and add to our understanding of influenza burden alongside COVID-19 burden. These ongoing activities will ensure that influenza burden estimation continues to be improved, using the best available statistical methods and data sources.

**Collaborative Forecasting Efforts Hosted by the CDC Use Surveillance Data to Look Ahead and Provide Quantitative Forecasts for Short-Term Trends in Influenza Activity**

Although these traditional surveillance systems and disease burden estimates are comprehensive, they measure influenza activity after it has occurred, which limits their utility to anticipate future trends and inform risk assessment, resource allocation, and health care preparedness. To address these limitations, the CDC has been collaborating with and supporting public health officials and researchers from academia, industry, and the government in open forecasting challenges since the 2013-2014 season [20]. This process has ensured that forecasting targets are relevant to public health, forecast data are openly available and communicated; and the evaluation of forecast performance across different targets, seasons, geographic locations, and methods is transparent [21-23]. This foundation helped lead to the rapid establishment of the COVID-19 forecast hub in early 2020 [24-26].

Building on 10 years of experience and working in collaboration with academic and industry partners, including 2 Centers of Excellence in Influenza Forecasting (the University of Massachusetts and Carnegie Mellon University), the CDC has regularly communicated short-term forecasts at the national, state, and territory level for the 2023-2024 influenza season for the weekly number of patients hospitalized with laboratory-confirmed influenza based on data from the National Healthcare Safety Network (NHSN) hospital admission system (previously known as the US Department of Health & Human Services (HHS) system HHS-Protect) [13,27]. This hospitalization-based forecast target was adopted during the 2021-2022 season when reporting of laboratory-confirmed influenza hospital admissions became mandatory in February 2022 at the state- and hospital facility- level and is different from the forecast target used before the COVID-19 pandemic, which was based on outpatient visits for influenza-like illness [20]. This shift was made because of changes in outpatient care seeking associated with the pandemic, the continued cocirculation of SARS-CoV-2 and RSV, which complicated the interpretation of influenza-like illness data, and the new availability of laboratory-confirmed influenza hospital admission data through NHSN for all 50 states. This centralized source of virus-specific hospitalization data may provide actionable and reliable information moving forward [28,29]. Experience from using the NHSN hospital admission indicator for influenza forecasting during the 2021-2022 and 2022-2023 influenza seasons [30], as well as during other forecasting and modeling efforts during the COVID-19 pandemic, indicates that the NHSN hospital admission data [13], which became available during the COVID-19 pandemic, are a robust source of timely information about the impact of influenza in US hospitals. Reported data include daily state-level counts of total hospitalized patients, previous day’s admissions, and total hospitalized intensive care unit patients each with a laboratory-confirmed influenza virus infection [13]. Insights from past forecasting efforts, including for the COVID-19 pandemic and past influenza seasons, can provide users of forecasts with valuable takeaways for interpreting influenza forecasts with appropriate levels of caution. First, forecasts are less reliable during periods of rapid change (eg, during periods of sharp increases in incidence or when activity is peaking). During these periods, forecasts are often unable to reliably predict changing trends, and interim evaluation of early season forecasts received during the periods of the 2022-2023 season when influenza activity was increasing rapidly or peaking indicated that more reported hospitalizations than expected fell outside of the forecast prediction intervals, indicating a lack of forecast reliability. Improvements in forecasting methodology are needed to ensure that periods of rapid change can be forecasted reliably. Due to the unreliability of forecasts observed in these periods, the Influenza Division did not start public communication of these forecasts until observed performance became more reliable (January 2023). To begin addressing this limitation, the FluSight challenge [27] piloted a set of new forecast targets for probabilities of observing categorical changes in hospital admission trends, such as increases and decreases above certain thresholds, at state, territory, and national levels and will continue the use of these targets in the 2023-2024 season. Second, ensemble models typically outperform individual model forecasts, making them a more robust choice for forecasts with dependable accurate performance [22-24,26]. Finally, forecast performance degrades as they predict farther into the future, as evidenced by the higher performance of 1-week ahead forecasts than 2-, 3-, and 4-week ahead forecasts [22], and the CDC may also pause communication of these longer-horizon forecasts in periods of rapid change. In addition to the challenges associated with rapid increases, atypical
seasonality, and new data sources for forecasting present additional challenges for forecasting influenza during and following the COVID-19 pandemic since many forecasting models rely on historical data to inform predictions of seasonal trends. Ongoing discussions with academic and industry partners that complement collaborations with the Council for State and Territorial Epidemiologists, including the pilot activities mentioned above, will help ensure that forecasting and modeling results are shared, discussed, and interpreted in a timely manner.

Additional Modeling Approaches Are Needed to Increase Our Understanding of the Potential Severity and Timing of the Upcoming Influenza Seasons

Inherent uncertainties in epidemiological conditions and the effects of interventions over longer time horizons require different modeling frameworks than forecasting alone. Therefore, the CDC started collaborating with the Scenario Modeling Hub to use a multiple-model approach to consider potential courses of the 2022-2023 influenza season, including 4 scenarios covering higher and lower rates of vaccine impact and different levels of preexisting immunity [31-33]. The simultaneous consideration of multiple scenarios (sets of assumptions) allows modelers to consider what could happen under a variety of different circumstances. Results from the first round of the influenza Scenario Modeling Hub suggested in September 2022 that hospitalizations in the 2022-2023 influenza season would be higher than in recent seasons [32]. Early in the 2022-2023 season, projections for scenarios with lower population-level immunity demonstrated the plausibility of peak hospitalizations comparable to those of larger prepandemic influenza seasons, while projections for scenarios with higher vaccine effectiveness and coverage were accompanied by substantial reductions in peak size and cumulative burden of hospitalizations due to influenza [32]. Subsequent rounds of the Influenza Scenario Modeling Hub undertaken in November and December 2022 reinforced these findings that projected peak hospitalizations could be comparable to those of larger prepandemic influenza seasons in the most plausible scenarios [32]. As demonstrated by several completed rounds of the COVID-19 Scenario Modeling Hub [34], these longer-term projections provide a useful information source for intervention planning, preparedness, and decision-making [35-38]. By continuing to engage with the Scenario Modeling Hub in weekly meetings, the CDC will be able to provide insight on which questions are of most interest for public health practice and preparedness during the 2023-2024 influenza season to help guide scenario development as well as lend expertise on data sets that may enhance modeling efforts. Results from the Influenza Scenario Modeling Hub may be used to inform potential peak timing and intensity. Additionally, influenza modelers within the CDC will continue to engage with the broader infectious disease modeling community. Information on changing influenza epidemiology and surveillance systems will be regularly communicated across forecasting groups, modeling groups, and data curators. This will be achieved in part through coordinated efforts with partners in the Models of Infectious Disease Agent Study Network [39].

Conclusion

The COVID-19 pandemic has changed many things that could influence trends observed in long-standing influenza surveillance systems. These changes have impacted ongoing efforts to describe influenza disease burden and forecast influenza activity and introduced new biases into epidemiologic and clinical studies that rely on the results of clinical testing. However, the pandemic has also provided opportunities for innovation that could better inform and support influenza efforts moving forward, including novel data sources such as the NHSN hospital admission data set, the ability to compare the impact of multiple respiratory pathogens in different clinical settings, and increased engagement with the public and external partners. Prepandemic approaches used to inform influenza prevention and control efforts, such as surveillance, burden estimation, and modeling, will continue to be reevaluated to ensure their reliability. State-level data may be used in the future to provide more spatially resolved estimates of burden beyond the national burden estimates that are currently provided. The US CDC Influenza Division remains committed to these key data and analytic activities, including the simultaneous monitoring of syndromic influenza surveillance with complimentary data sources, as well as providing a central platform for collaboration and communication across disciplines and institutions to ensure that information to inform influenza prevention, situational awareness, and control remains robust and readily available.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Conflicts of Interest

None declared.

References


32. Flu Scenario Modeling Hub. URL: https://fluscenario modelinghub.org/ [accessed 2023-08-11]


34. COVID-19 Scenario Modeling Hub GitHub. URL: https://github.com/midas-network/covid19-scenario-modeling-hub [accessed 2022-12-12]


39. Models of Infectious Disease Agent Study (MIDAS). URL: https://midasnetwork.us [accessed 2022-12-12]

Abbreviations

CDC: Centers for Disease Control and Prevention
HHS: Health & Human Services
NHSN: National Healthcare Safety Network
RSV: respiratory syncytial virus

Edited by A Mavragani; submitted 06.11.23; peer-reviewed by E Lau, J Olsen; comments to author 06.02.24; revised version received 12.02.24; accepted 13.02.24; published 08.04.24.

Please cite as:
Borchering RK, Biggerstaff M, Brammer L, Budd A, Garg S, Fry AM, Iuliano AD, Reed C
Responding to the Return of Influenza in the United States by Applying Centers for Disease Control and Prevention Surveillance, Analysis, and Modeling to Inform Understanding of Seasonal Influenza
JMIR Public Health Surveill 2024;10:e54340
URL: https://publichealth.jmir.org/2024/1/e54340
doi:10.2196/54340
PMID:38587882
From the Public Health Perspective: a Scalable Model for Improving Epidemiological Testing Efficacy in Low- and Middle-Income Areas

Xuefeng Huang
Qian-Yi Kong
Xiaowen Wan
Yating Huang
Rongrong Wang
Xiaoxue Wang
Yingying Li
Yuqing Wu
Chongyuan Guan
Junyang Wang
Yuanyuan Zhang
1 School of Public Health, Dalian Medical University, Dalian, China
2 School of Economics and Management, University of Science and Technology Beijing, Peking, China
3 School of Economics and Management, Jiangxi University of Traditional Chinese Medicine, Nanchang, China

Corresponding Author:
Yuanyuan Zhang, MD
School of Public Health
Dalian Medical University
No 9, West Section of Lvshun South Road
Dalian, 116041
China
Phone: 86 18641106272
Email: zhangyuan@dmu.edu.cn

Abstract
The globe is an organically linked whole, and in the pandemic era, COVID-19 has brought heavy public safety threats and economic costs to humanity as almost all countries began to pay more attention to taking steps to minimize the risk of harm to society from sudden-onset diseases. It is worth noting that in some low- and middle-income areas, where the environment for epidemic detection is complex, the causative and comorbid factors are numerous, and where public health resources are scarce. It is often more difficult than in other areas to obtain timely and effective detection and control in the event of widespread virus transmission, which, in turn, is a constant threat to local and global public health security. Pandemics are preventable through effective disease surveillance systems, with nonpharmacological interventions (NPIs) as the mainstay of the control system, effectively controlling the spread of epidemics and preventing larger outbreaks. However, current state-of-the-art NPIs are not applicable in low- and middle-income areas and tend to be decentralized and costly. Based on a 3-year case study of SARS-CoV-2 preventive detection in low-income areas in south-central China, we explored a strategic model for enhancing disease detection efficacy in low- and middle-income areas. This model can improve the local epidemic detection efficiency, ensure the health care needs of more people, reduce the public health costs in low- and middle-income areas in a coordinated manner, and ensure and strengthen local public health security sustainably.

KEYWORDS
low- and middle-income country; LMIC; pandemic; epidemiological surveillance; universal public health; nonpharmacological interventions; public health; callable model; efficacy; COVID-19; public safety threats; effectiveness; China; detection; epidemic

Introduction
Globally, advances in epidemiological, behavioral, and pharmacological interventions have led to a decline in the incidence of sudden public health events. However, in some low- and middle-income countries (LMICs), large-scale outbreaks of various communicable or noncommunicable diseases have not been proportionately contained, and the burden on public health continues to rise [1].

Current interventions for epidemiological prevention are suboptimal and tend to be high-cost and decentralized. A typical recent manifestation of this is the fact that in the worldwide discussion of how to win the COVID-19 war, the rush to action has varied greatly from country to country, and even within
countries. The US government advises against parties of more than 10 people, but San Francisco, California, has ordered everyone to stay home. Italy, France, and Spain have sealed off their populations almost completely, with police or military patrolling the streets in some places. Even as major cutting-edge magazines continue to report on the health threat of COVID-19, pubs remain open in Britain, Germany’s schools close, and Sweden’s schools remain open to young children. The patchwork reflects the different stages of the epidemic, as well as differences in resources, cultures, governments, and laws [2]. However, there has long been controversy about what works best, and how to balance the necessary with the reasonable, which some of the strategies adopted by these countries are missing elsewhere: widespread testing to detect cases, mass tracking of their contacts to test or quarantine them, and encouraging or forcing infected people to quarantine themselves. It is discrete and not cost-effective. Conversely, there is a large gradient in the timing of the blockades as income statuses change, with LMICs initiating disincentives earlier than ultra–high-income or high-income countries [3]. In the case of India, one of the first countries to observe the health impacts of neocorona viruses and to adopt a containment strategy, the establishment of a comprehensive blockade strategy of varying degrees of severity was declared in 2020, but in April 2021, the second wave of the COVID-19 pandemic in the country inexorably shifted from the west and north to the east and south [4]. A study investigating the scale of the infection showed an increase in seroprevalence in India from just over 20% in January 2021 to 67% in June-July 2021 [5]. Initially, local scientists attributed the increase to a variant of the virus, but after several months of comparative studies, it was found that the variant was not more transmissible [6]. Although Indian public health researchers have stressed the need for continued vigilance and standardized genomic surveillance, there are no restrictions on major events in key states, especially elections or religious gatherings, where returnee positivity rates are very high. This has upset the balance of prevention systems in different states and cities, and the interaction of old and new problems has created huge economic gaps in the health systems of different regions. The concomitant shortage of resources due to drug abuse and declining drug resistance, as well as more cases of multiple comorbidities, seems to be a common problem in many low- and middle-income areas [7]. Another issue of concern is vaccination. Most countries use electronic registries, and despite the principle of harmonization, vaccination during pandemics has long been limited by availability. Moreover, unlike the excellent tracking of vaccination in high-income regions, there has been a major gap in the ability of low- and middle-income regions to estimate the magnitude of morbidity and mortality during pandemics accurately. Past mortality analyses suggest that socioeconomic deprivation during pandemics leads to more excess deaths [8], and this is only a preliminary assessment from regions with stronger health and data systems; some transitional regions or poorly informationalized tribes may be worse off than realistically predicted. The pandemic in India and the current state of vaccine distribution worldwide point the way to the future of pandemic prevention in LMICs—early signal detection, systematic identification and analysis of health systems and their supply chain preparedness, and protection of the ability to deal with not only emergencies but also other disease-care services, as well as the ability to collect, collate, integrate, analyze, and interpret in real time. The ability for data to be recycled so that resources can be appropriately allocated when needed and health systems can better aid in understanding the causes of both poor and favorable outcomes. In 2022, A research survey further reiterated that epidemiological surveillance tends to prioritize urban and adjacent rural communities, but it simply does not reach low-income or remote rural areas [9]. To address the current gaps in pre-epidemic surveillance and prevention, Mobarak et al [10] suggest creating a relatively complete and locally relevant general detection system, including individual and population-wide engagement strategies, in areas where the basic public health system lags relatively behind that of other higher-income areas, which would address some of the cost commonalities and resolve some social conflicts.

Nonpharmacological interventions (NPIs), as the most important part of the current epidemiological prevention and control system, are aimed at raising awareness of risk reduction activities at the individual level (eg, proactive wearing of face masks and promotion of social distancing, among others) and the group level through a range of public health strategies, as well as common actions assumed in response to public health outbreaks (eg, group testing and quarantine, among others). This decelerates the spread of disease and prevents larger outbreaks to ensure that hospitalization and mortality rates are manageable [11].

In fact, before the SARS-CoV-2 outbreak, many countries had made steady progress toward achieving Sustainable Development Goal target 3.4 (reducing premature mortality from noncontagious diseases by one-third between 2015 and 2030) [12]. Of note, influenza pandemics impede these successes, disrupt health systems, divert resources, and draw attention away from them. Therefore, an important fact about COVID-19 is that anticipating and being alert to viral risks in advance and building a strong public health system with supervision, detection, and timely responses is essential and will greatly benefit the rational allocation of resources to the district health system and provide it with a path for the future. However, in designing and operating a surveillance intervention system, consideration needs to be given to the interrelated structural factors that support or impede preventive actions and behaviors of individuals or groups. These structural factors can be variable or inherent to social patterns, such as socioeconomic and cultural conditions, political will (eg, cooperation, commitment, and funding), governmental priorities or policies, and conditions related to environmental factors (eg, reducing industrial pollution). Addressing these structural factors can improve the preventive effectiveness of public health and protect the health interests of more people, thereby achieving sustainable economic and health benefits [13].
**The Significance of Structured Disease Detection Systems in Low- and Middle-Income Areas**

About half of the people in LMICs live in rural or remote areas [14]. Livestock raised in remote rural areas accounts for about 25% of the agricultural gross domestic product of LMICs and supports the livelihoods of more than 600 million smallholder farmers [15]. Complex environmental pathogenicity and high prevalence of zoonotic factors are a major problem for not only local but also global public health. In fact, previous pandemics have fueled a revolution in global disease prevention, including one in LMICs. It is emerging as a global consensus that vaccinating populations in remote areas will lead to a reduction in the risk of new viral mutations worldwide [16]. In many countries, domestic distribution capacity can be established quickly, affordably, and cost-effectively. LMICs have accumulated extensive experience with mass immunization campaigns, which have resulted in high levels of vaccine coverage for children and older adults [17]. Several countries, including Ghana, Liberia, India, Pakistan, and Sierra Leone, have begun experimenting with the concept of “mobile vaccination teams,” which bring batches of vaccines to the places where people live to make vaccinations more accessible [18,19]. These can involve nurses, supported by community mobilizers, visiting remote villages with vaccines to raise the attention of local residents and leaders, and calling for effective vaccination. This model has also been used successfully in the past: the “outreach clinics” in the Gambia, which provide immunization services to difficult-to-reach subpopulations (eg, homeless individuals, stigmatized groups, and migrant groups, among others), are considered an important factor in the elimination of measles [20]. This is mutually beneficial to the systematic monitoring scalable model we proposed. An efficient disease detection system can provide precise regional control and tools for local public health management, more effectively decelerate the spread of diseases in complex pathogenic environments, alleviate social conflicts in NPIs, enhance the vitality of public health, and provide a more precise guarantee of public health security and economic development in low- and middle-income areas. However, policy decisions are usually time- and resource-intensive and require a multidisciplinary team to implement. While policy support mechanisms tend to be effective in high-income countries, they are often inadequately monitored in many LMICs. Inadequate monitoring is largely due to budgetary constraints, which limit the ability to plan and coordinate. The use of affordable digital technologies can facilitate real-time monitoring and feedback from individuals, service providers, and policy makers. On the other hand, the theory of “synergy” has always played an important role in social science research and is often combined with governance theory by management scholars to create a more effective theory of “collaborative governance.” Thus, to address structural and other barriers to disease prevention and control, there is a need for a timely and more flexible, public, and dynamic system in place, which emphasizes the use of the total of the many ways in which individuals and various public or private institutions manage their commonalities [21], and which promotes public policy support mechanisms to guide policy decision-making in disease prevention.

**What Should We Do to Achieve Efficient Detection?**

**Overview**

As we discussed earlier, there has long been controversy about what works best and how to balance necessity with rationality. A complex pathogenic environment remains, but the past history of overscaled, short-lived surge operations has been unsystematic and wasteful. It brings not only lower-middle-income areas economic disasters during the pandemic but also the same or even greater vulnerability of local public health systems to a new round of public health challenges. So, we have transferred and quantified the traditional thinking on disease prevention and control to develop a disease prevention–sharing model with the government as the protagonist and the participation of the population, the health care system, and the private sector, and we have provided a service architecture diagram about epidemiological surveillance in low- and middle-income areas (Figure 1). Adopting the 4 dimensions of human, financial, material, and labor as the object of assessment and apportionment, this or the next public health detection event comprises 6 recyclable phases, including preparation, assessment, action, tracking, reassessment, and improvement, as well as an additional phase, with the first 6 phases operating in a closed loop. The efficacy assessment of each of these stages is based on our analysis of epidemiological data from selected low- and middle-income areas of south-central China, a comprehensive and targeted intervention design that is effective in the management of infectious disease illnesses such as pneumococcal pneumonia. Furthermore, this approach also can be considered for the management of noncommunicable or other epidemics such as zoonotic diseases.
Phase 1: Every Part Is Ready

The primary aspect of improving disease detection efficacy in low- and middle-income areas is to strengthen the response mechanisms of various stakeholders and build a bundle with efficient linkages. Throughout the COVID-19 pandemic, everyone has been able to be a transmitter and everyone has been able to be a defender. In the face of the public enemy of humanity, all controversies and prejudices were temporarily set aside [22]. We have divided all participants into 4 categories: the first is the government or local governing body, the second is the public, the third is the medical organizations, and the fourth is the private organizations. Generally speaking, the connection between regional governments and regional medical institutions is very close, and most public health organizations always respond to the first call of regional governments to solve public health matters; private institutions also bear certain social responsibilities. Simultaneously, in their daily health business dealings, they cooperate with public medical units to lay a large stage of the contact mechanism as the preparation stage of the second response echelon. The response of the population is the slowest and most important presence in the whole detection system. The development of new media communication has greatly saved the time and material costs of health information dissemination, but given that a large portion of the population lives in remote rural areas, there is still a gap in the ability of...
these populations to perceive the message due to the variability of communication networks. To solve such problems, the Chinese government has set up small government-like management teams in each village, which are elected by the residents themselves. They tend to have higher education levels and political literacy, and a high degree of familiarity with the environment in which they are located, and the residents’ authority over them is even higher than that of other expatriates, so they can be more accurate and effective in delivering information and preventing risks in their localities. In addition, and more importantly, it is often much less costly to train them than it is to reach the general population. We may wish to use such a mechanism to achieve a more efficient detection and cost-saving pathway, combined with the effectiveness of new media communications to form a third echelon of the crowd response during the preparedness phase.

Thus, we can have a more linear public health emergency response pathway, and when the government issues instructions for the preparedness phase, the first and second response echelons can have a clearer preparation plan for material preparation and personnel control according to the different levels of demand of the third response echelon, which provides an introductory for later detection effectiveness tracking and cost analysis.

**Phase 2: Load Balancing of All Available Resources**

Recommendations have been offered for nongovernmental organizations, governments, donors, and future research including studying the organizational effectiveness and sustainability of these partnerships to deliver effective and efficient health outcomes to recommend universal best practices in health care [23]. In disease detection systems in low- and middle-income areas, the usage and translational benefits of limited public health resources have a direct impact on the continuation of detection efforts. Load balancing is a quantitative assessment and optimization of resource interchange based on all available human, financial, material, and physical resources that the local area has. The 3-level response echelon should combine the actual local public health resources with the epidemic burden, analyze the public health cost of safeguarding the most realistic population’s basic health care needs at the emergent stage, and then perform resource interchange. Specifically, we start from 4 dimensions, including human, financial, material, and productivity dimensions, based on off-epidemic control, and the load capacity of each dimension in different regions needs to be precise to unit order of magnitude. For example, the number of people served by medical centers at all levels, the number of local funds invested in public health, the maximum number of testing reagents that can be provided, the maximum number of patients that can be admitted locally, etc, to complement each other’s strengths, visually help make up for the public health shortcomings of different regions in terms of funds or talents, optimize short-term benefits, and continuously mobilize other potential resources to maintain a more regular and sustainable development strategy.

In the previous phase of preparation, we emphasized the importance of response mechanisms, where the participation of all stakeholders enhances the carrying capacity of regional pre–disease detection efforts. The quantification of load balancing can effectively stimulate the participation of stakeholders to meet the simplest health rights needs of the widest local population in the pandemic era. On the one hand, it is an effective way to stimulate the work capacity of government and other health personnel, and on the other hand, it is a way to give direction and assistance to other stakeholders who are willing to take more social responsibility, thus maximizing the potential local public health capacity. The government as the core of the work can establish a public mechanism after quantifying the load capacity, and a joint class of government personnel and media to disclose the progress and cost budget of disease detection and prevention and control work at each stage to the population so that the transformation among human, financial, material, and productivity can have more flexible possibilities due to the participation of more aspects.

**Phase 3: Detecting General Differences in Actions May Be the Key to Cost Savings**

The collaboration of stakeholders who have gone through the first 2 phases provides a clear framework and a more convenient environment for our next testing initiatives. We also need to take into account the general differences in population and environment when conducting disease detection in low- and middle-income areas. On the one hand, most people in low- and middle-income areas live in rural or remote areas and depend on the agricultural economy for their livelihoods, with low external mobility and relatively high levels of self-sufficiency. It is worth noting that these groups tend to have far more frequent contact with the natural environment (the novel coronavirus was originally identified as being carried by bats [24]). On the other hand, the degree of aging has always been at a high level in the rural population, so even though in the first phase of our preparation of the links to improve the attitude of the population, especially among older groups to cooperate and respond to the current disease, most of them are only motivated by the importance of their health care, while other disease-causing factors around them, such as food, domestic animals, work environment, etc, have a lower deficit of prevention awareness. General differences in population characteristics and the complexity of disease environments that are difficult to monitor have been important reasons underlying the difficulty of timely and effective control of outbreaks in low- and middle-income areas, and even higher disease mortality rates. Such a situation requires us to prioritize environmental and “special population” factors in the detection system, and governments, public health organizations, and other private institutions need to adopt a more appropriate multifaceted cooperation mechanism to achieve population and environmental biases of resources, effectively and precisely cut off the transmission of diseases, protect special populations, and thus save long-term testing costs.

Based on this, to achieve the efficiency and sustainability of the testing initiative, we suggest that the government could take the lead in responding to the third-tier government-like personnel and public health organizations in cooperation with private institutions to establish a special disease testing service site in each tribe of the low- and middle-income areas by
measuring the geographic population and demand differences. The site could be staffed by local government workers with village doctors, and the testing laboratory and testing reagents and other equipment could be outsourced by a third-party organization. Adhering to the purpose of prioritizing the health testing needs of special populations and maintaining the living environment, we can choose to prioritize the training of these government workers and village doctors on disease prevention knowledge about collecting the disease testing needs of local special populations and the environmental risks of suspected transmission routes that need to be urgently cut off, so that when risk factors are found in the NPI work process, they can be fed back promptly to the service station. The medical staff of the service station can then directly implement testing and intervention, thus changing the local passive testing prevention system.

**Phase 4: All Traces of Traced**

The main reason why, in the previous phase, each task needs to be divided so carefully is to achieve full pathway effectiveness tracking in this phase, thus reducing internal suspicion to a greater extent and achieving a relatively stable phase cycle. In low- and middle-income areas, a series of problems such as unprofessional medical personnel, unclear division of functions between government and other agencies, and simplistic and one-size-fits-all disease prevention processes have been running through the whole stage of disease outbreaks or suboutbreaks. This is resulting in the emergence and spread of sudden epidemics in low- and middle-income areas, once they are prone to difficult preventive measures and slow effectiveness of control means, and such recurrence has led many low- and middle-income regional governments to find it difficult to cope with the cost overlap, and this even causes them to choose to adopt the sluggish strategy of abandoning detection and relaxing preventive control.

**Phase 5: Another Small Revolution in Costs**

At this stage, we have a general framework for disease detection, and some cost-trimming is required. The government, as the main body of action, needs to, on the one hand, give new load ratios based on the standard costs, actual costs, and sunk costs of the testing process as fed back by the hospitals in conjunction with the initial assessment, and then develop appropriate financial subsidy policies and epidemic prevention regulations to adjust the targets and outcome outputs of the testing work of medical institutions. On the other hand, it is necessary to promptly assess the cost-effectiveness of the testing services provided by third-party institutions throughout the testing service phase and determine the effect of social responsibility, to delineate new testing service intervals.

**Phase 6: Time to Make Some Changes**

The first 5 phases provide the basis for creating meaningful improvements in the improvement phase. During the outbreak period, we have had the experience of the first round, the population detection strategy that can be adjusted to areas with high susceptibility or population mobility. With limited detectors, a deployment system that can be adapted to assign additional personnel to besiege high-incidence areas, the ability to abate the detection force of the extensive population, and prevent and control with precision, the government can redivide the areas and classify them to guide the medical institutions’ service bias. The work of environmental testing can also be based on the susceptibility of different objects to emphasize the transformation of responsibilities between third-party testing agencies and public health agencies, targeting biological testing with a higher effectiveness bias than nonbiological testing to prevent the emergence of new forms of transmission such as zoonotic diseases that cause more serious economic losses. In the small outbreak period or stabilization period, all aspects of the detection force need to return to the 3 stages of tracking link posts, narrowing the envelope, strengthening the regularity of detection laws, reducing the frequency, shrinking the detection object at the same time, according to the cost criteria assessed, gradually narrowing the envelope of disease detection, increasing the service station of other disease surveillance functions. Thus, the cost of testing is continuously reduced and economic pressure is released.

**An Additional Common Claim for Enhanced Disease Detection in Low- and Middle-Income Areas**

Maintaining a strong database on disease surveillance and building a professional public health team to safeguard the health rights of a broader population and preserve an otherwise fragile economic environment has been a rather genuine need in low- and middle-income areas [25]. This will require a long-term joint effort by several committed and highly professional scholars in epidemiology, accounting, social sciences, and other related fields, together with low- and middle-income areas [26]. Despite the methodical efforts of disease detection and prevention in low- and middle-income areas, they are still passively waiting and receiving policy guidelines or emulation from higher levels. Much of the collection, organization, storage, and analysis of self-reported data is based on templates from other developed regions, and the cost of the vicious cycle of such a model can be devastating to some LMICs [27].

**Conclusions**

Here we describe an integrated and comprehensive approach that covers structural, social, and personal strategies to optimize the epidemic surveillance system in low- and middle-income areas. This model can improve the local epidemic detection efficiency, ensure the health care needs of more people, reduce the public health costs in low- and middle-income areas in a coordinated manner, and ensure and strengthen local public health security sustainably. Although there are certain differences in the policies and environment of disease detection in each region, our framework is more intuitive and easier to understand. We hope to provide clear behavior paths and tracking paths for managers, public health workers, and participants in other collaborative events in low- and middle-income areas, and improve the possibility of all measures that can maintain local public health security.
Acknowledgments

This project was supported by the “Research on the Opening up of Higher Education to the Outside World” project of the Introduction of Foreign Intelligence Work Branch of the Chinese Association of Higher Education (21ZSYZXB16). We thank the Centers for Disease Control and Prevention (CDC) of Yuyang County, Yingtan City, Jiangxi Province for their support of our research.

Authors’ Contributions

XH designed the scheme, completed the survey, and drafted the manuscript. YZ is responsible for the legal, academic, and ethical norms of the paper. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

LMIC: low- and middle-income country
NPI: nonpharmacological intervention
Development of a Consolidated Health Facility Masterlist Using Data From Polio Electronic Surveillance in the World Health Organization African Region

Marie Aimee Babona Nshuti\(^1\), MSc; Kebba Touray\(^1\), MSc; Ticha Johnson Muluh\(^1\), MD; Godwin Akpan Ubong\(^1\), PhD; Reuben Opara Ngofo\(^1\), MSc; Bello Isa Mohammed\(^1\), PhD; Ishimwe Roselyne\(^1\), MSc; David Oviaesu\(^1\), MIS; Evans Mawa Oliver Bakata\(^1\), MSc; Fiona Lau\(^1\), PhD; John Kipterer\(^1\), MPH; Hugh Henry W Green\(^1\), MBA; Vincent Seaman\(^2\), PhD; Jamal A Ahmed\(^1\), MD; Modjirom Ndoutabe\(^1\), MD

\(^1\)World Health Organization Regional Office for Africa, Brazzaville, Congo
\(^2\)Bill & Melinda Gates Foundation, Seattle, WA, United States

Corresponding Author: Marie Aimee Babona Nshuti, MSc
World Health Organization Regional Office for Africa
Cite du Djoûé, Bacongo, PO Box 06
Brazzaville
Congo
Phone: 242 069248040
Email: shutaime03@gmail.com

Abstract
Geospatial data reporting from surveillance and immunization efforts is a key aspect of the World Health Organization (WHO) Global Polio Eradication Initiative in Africa. These activities are coordinated through the WHO Regional Office for Africa Geographic Information Systems Centre. To ensure the accuracy of field-collected data, the WHO Regional Office for Africa Geographic Information Systems Centre has developed mobile phone apps such as electronic surveillance (eSURV) and integrated supportive supervision (ISS) geospatial data collection programs. While eSURV and ISS have played a vital role in efforts to eradicate polio and control other communicable diseases in Africa, disease surveillance efforts have been hampered by incomplete and inaccurate listings of health care sites throughout the continent. To address this shortcoming, data compiled from eSURV and ISS are being used to develop, update, and validate a Health Facility master list for the WHO African region that contains comprehensive listings of the names, locations, and types of health facilities in each member state. The WHO and Ministry of Health field officers are responsible for documenting and transmitting the relevant geospatial location information regarding health facilities and traditional medicine sites using the eSURV and ISS form; this information is then used to update the Health Facility master list and is also made available to national ministries of health to update their respective health facility lists. This consolidation of health facility information into a single registry is expected to improve disease surveillance and facilitate epidemiologic research for the Global Polio Eradication Initiative, as well as aid public health efforts directed at other diseases across the African continent. This review examines active surveillance using eSURV at the district, country, and regional levels, highlighting its role in supporting polio surveillance and immunization efforts, as well as its potential to serve as a fundamental basis for broader public health initiatives and research throughout Africa.

(JMIR Public Health Surveill 2024;10:e54250) doi:10.2196/54250

KEYWORDS
African region; electronic surveillance; geographic information systems; Global Polio Eradication Initiative; integrated supportive supervision; polio

Introduction
After a decades-long, global effort, wild-type polioviruses (WPV) 2 and 3—the cause of poliomyelitis (polio)—were eliminated worldwide in 2015 and 2019, respectively [1]. However, WPV1 remains endemic in Afghanistan and Pakistan, where mountainous terrain, armed conflict, and societal and political factors have impeded polio immunization and surveillance [1,2]. More than 95% of polio cases are asymptomatic or associated with only mild, cold-like symptoms.
Mature, however, ~1 in 200 infected individuals develop acute flaccid paralysis (AFP) and paralytic polio, which have case fatality rates of 2% to 5% among children and up to 15% to 30% among adolescents and adults [3,4].

Infected persons, including asymptomatic individuals, may shed the polio virus in nasopharyngeal secretions or stool for several days to weeks, increasing the risk of transmission of this incurable and potentially fatal disease [3]. A Pakistani lineage of WPV1 was recently found in Mozambique and Malawi, demonstrating that polio risk extends far beyond the communities where it is endemic [5]. In addition, circulating vaccine-derived poliovirus (cVDPV) may emerge in populations with low immunization rates due to the genetic instability of the live-attenuated viruses included in the oral polio vaccine [6,7]. These mutated, neurovirulent strains produce paralysis much like WPV-associated polio [7]. They also account for the vast majority of polio cases [8].

To stop the spread of polio, the World Health Organization (WHO) Global Polio Eradication Initiative (GPEI) advocates intensive efforts to identify AFP cases in children aged <15 years (ie, disease surveillance)—both routine immunization (coverage >95%) and immunization campaigns to boost vaccination coverage to >95% and enhance herd immunity—to eliminate the potential for cVDPV transmission, and the deployment of novel type 2 oral polio vaccine, which reduces transmission of cVDPV2, the most common cause of polio [2,9,10]. The complementary strategies of surveillance and immunization led to the African continent being declared indigenous WPV-free in 2020, and since GPEI efforts began in 1988, nearly 1 million polio deaths were averted [11,12]. However, disruptions in GPEI activities during the COVID-19 pandemic contributed to a resurgence in African polio cases [2]. Since 2020, >1000 polio cases have been documented, including >600 reported in Africa in 2022, which highlights the urgent and ongoing need for intensive disease surveillance and immunization to prevent the further spread of polio [8,12-14].

Geospatial data reporting from surveillance and immunization efforts is a key aspect of WHO GPEI polio-eradication efforts in Africa, which are coordinated through the WHO Regional Office for Africa (AFRO) Geographic Information Systems (GIS) Centre. To ensure the accuracy of field-collected data, the AFRO GIS Centre has developed mobile phone apps such as electronic surveillance (eSURV) and integrated supportive supervision (ISS) geospatial data collection programs. These forms, apart from reporting on unreported AFP cases, are currently being used to create a consolidated Health Facility Masterlist for the WHO AFRO, which is a repository database containing validated baseline data for all health facilities in the African region. The Health Facility Masterlist contains comprehensive listings of the names, locations, and types of health facilities in each WHO AFRO member state. Field officers are responsible for documenting and transmitting geospatial location information regarding health facilities and traditional medicine (TM) sites using the eSURV form during their visits to active surveillance sites. The functionality of these facilities, as described in the Health Facility Masterlist, goes beyond mere listing. It involves the integration of health facility data and traditional health sites to enhance disease surveillance capabilities, particularly for polio. The Masterlist is intended to serve as a dynamic resource for health workers, enabling better planning and targeting of health interventions, improving response times to disease outbreaks, and facilitating more effective deployment of resources. The consolidated Masterlist should support at the operational level health strategies and interventions by ensuring surveillance is done properly at health facilities, and by providing a comprehensive, accessible, and up-to-date database of health facilities across the WHO African region. This review delves into the use of current and comprehensive health facility lists at the district, country, and regional levels, highlighting their role in supporting polio surveillance and immunization efforts, as well as their potential to serve as a fundamental basis for broader public health initiatives and research endeavors across Africa.

**Master Facility List Program in Africa**

A master facility list (MFL) is a complete, up-to-date, authoritative listing of health care facilities in any given country. Each MFL includes the information needed to accurately identify each facility, including the name, unique identifier, location, facility type, ownership, operational status, contact information for key personnel, types of services offered, and number of beds. The MFL should be validated, continuously updated, and accessible to all authorized stakeholders [15].

To date, MFLs have been updated manually. This process typically involves the country team sharing new facility information as it becomes available or through the use of the “Other” option. When a facility is not found in the existing list and is subsequently added as an “Other” option, it necessitates a manual update of the list. The upcoming version of eSURV should, however, bring significant improvements. It will incorporate a health facility registry, which will streamline the process of adding new facilities and updating existing facility forms.

Several countries have successfully compiled and made their MFLs available online, marking significant progress in health care information accessibility over the past 5 years. Notable examples include Zambia [16], Botswana [17], and Kenya [18]. There are also broader initiatives such as the WHO AFRO Integrated African Health Observatory MFL database and the Geolocated Health Facilities Data Initiative, aimed at enhancing health care planning and delivery with a more comprehensive MFL.

Despite these advancements, challenges remain in ensuring the continuous updating and maintenance of these lists. While eSURV visits are conducted regularly at health facilities—including traditional sites such as those of traditional practitioners—to verify and augment existing MFLs with new or missing facilities, the objective of sustaining up-to-date MFLs relies heavily on the diligent efforts of field officers conducting active surveillance. This ongoing process is crucial for maintaining the accuracy and utility of MFLs over time.
MFL Resource Package and Training

Before the availability of electronic databases, African MFLs were beset by challenges common to all paper-based data management systems, including out-of-date information, omissions, inconsistencies, and inaccuracies [19]. In addition, updating or correcting paper-based files and sharing the information with stakeholders such as other government ministries, public health agencies, and charitable organizations could be slow and resource-intensive. To address these shortcomings, the WHO and the United States Agency for International Development developed a set of guidelines for the implementation of a software-based facility registry service that makes the MFL accessible to stakeholders such as insurance companies; charitable groups undertaking health interventions; researchers assessing health system performance; and personnel involved in health management information systems, disease surveillance, and supply chain management [15,20]. A training program was also developed to help country teams establish or improve their MFLs. The target audience for this program includes ministry officials, facility list managers, data curators, national and regional health ministry personnel, and staff from nongovernmental organizations. Participants are trained in the purpose and contents of an MFL, MFL assessment, governance, data content, incorporation of geographic coordinates, establishment of the data set and registry service, ongoing maintenance, and sharing of the MFL, with a final session devoted to crafting an action plan for MFL creation and implementation.

TM Sites in MFLs

An important component of modern MFL development is the inclusion of TM sites in the database. Traditional healers serve as first-line health care providers for most of the population in Africa, where there is only 1 medical physician for every 40,000 people, compared with a 1:500 ratio of traditional healers to African residents [21-23]. Although some TM practices are associated with misdiagnoses leading to missed cases and ineffective treatment, traditional healers can have a strong, positive impact on the health and wellness of the people they serve when their products and practice are properly regulated and integrated into health systems [22]. These practitioners are also an important resource for disease surveillance and immunization efforts and may be trained to conduct community disease surveillance in areas where there is a shortage of medical facilities and formally trained health care providers.

eSURV and ISS in Practice

About eSURV

eSURV was developed by the WHO AFRO GIS Centre as an electronic health solution to enable surveillance agents to use mobile phones to document active case searches for polio and other infectious disease cases in health facilities and the community. The app comprises an electronic form based on a checklist that records the geographic location of health facilities visited by data collectors based on the global positioning system software on the user’s phone. The tool is used in 46 WHO AFRO member states, where case search activity is automatically recorded on country-specific modules on the GIS servers. The accumulated data are used to identify outbreaks of not only polio, but also other priority diseases such as COVID-19 [24,25]. As a result of data being continuously transmitted to GIS servers in near real time, areas with little or no surveillance activity can be promptly identified and corrective actions can be taken.

About ISS

Supportive supervision fosters high-quality disease surveillance and immunization best practices through systematic visits to priority sites (health facilities and surveillance sites, including TM practitioner sites) by supervisory WHO and national health ministry personnel, where they conduct monitoring, evaluation, and on-the-job training of health workers. To document these activities, the supervisors use ISS, a mobile phone app consisting of an integrated electronic checklist used for the supervision of active AFP case finding and routine polio immunization [24]. Approximately 40% of checklist questions focus on case surveillance for AFP and other vaccine-preventable diseases, whereas the other 60% assess routine immunization practices. Through ISS, supervisors promote understanding and recognition of AFP among frontline workers and help them keep immunization records up to date [26]. In addition, ISS facilitates the detection of previously missed AFP cases (Figure 1 and Table 1).

To ensure uniformity and standardization, the ISS tool includes 4 question categories the supervisor can ask frontline workers. The ISS checklist includes questions about both AFP cases and polio vaccinations. The app also includes the eSURV checklist, consisting of questions only about active AFP cases. A third category covers only routine polio immunizations and a fourth focuses on COVID-19 case surveillance (Figure 2 and Table 2).
**Figure 1.** Number of districts with unreported cases of AFP in 2021 as revealed through the use of integrated supportive supervision. AFP: acute flaccid paralysis.

**Table 1.** Number of districts with unreported cases of acute flaccid paralysis in 2021 as revealed through the use of integrated supportive supervision.

<table>
<thead>
<tr>
<th>District</th>
<th>Cases, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cameroon</td>
<td>350</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>198</td>
</tr>
<tr>
<td>Congo</td>
<td>103</td>
</tr>
<tr>
<td>Kenya</td>
<td>102</td>
</tr>
<tr>
<td>South Sudan</td>
<td>102</td>
</tr>
<tr>
<td>Democratic Republic of Congo</td>
<td>94</td>
</tr>
<tr>
<td>Mali</td>
<td>70</td>
</tr>
<tr>
<td>Uganda</td>
<td>64</td>
</tr>
<tr>
<td>Côte d’Ivoire</td>
<td>46</td>
</tr>
<tr>
<td>Malawi</td>
<td>42</td>
</tr>
<tr>
<td>Senegal</td>
<td>30</td>
</tr>
<tr>
<td>Togo</td>
<td>30</td>
</tr>
<tr>
<td>Zambia</td>
<td>28</td>
</tr>
<tr>
<td>Niger</td>
<td>18</td>
</tr>
<tr>
<td>Madagascar</td>
<td>17</td>
</tr>
<tr>
<td>Gabon</td>
<td>11</td>
</tr>
<tr>
<td>Chad</td>
<td>10</td>
</tr>
<tr>
<td>Ghana</td>
<td>10</td>
</tr>
<tr>
<td>Mauritania</td>
<td>9</td>
</tr>
<tr>
<td>Guinea</td>
<td>8</td>
</tr>
<tr>
<td>Namibia</td>
<td>8</td>
</tr>
<tr>
<td>South Africa</td>
<td>8</td>
</tr>
<tr>
<td>Comoros</td>
<td>7</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>5</td>
</tr>
<tr>
<td>Benin</td>
<td>4</td>
</tr>
<tr>
<td>Liberia</td>
<td>4</td>
</tr>
</tbody>
</table>
Table 2. Frequency of type of supervision in Cameroon and Mali from July 2020 through June 2021.

<table>
<thead>
<tr>
<th>Type of supervision</th>
<th>Cameroon cases, n(^a)</th>
<th>Mali cases, n(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated supportive supervision (surveillance and routine immunization)</td>
<td>13,152</td>
<td>3099</td>
</tr>
<tr>
<td>Surveillance visit (eSURV(^c)—active case search)</td>
<td>1,171</td>
<td>4811</td>
</tr>
<tr>
<td>Routine immunization (supportive supervision)</td>
<td>545</td>
<td>213</td>
</tr>
<tr>
<td>Remote eSURV</td>
<td>125</td>
<td>34</td>
</tr>
<tr>
<td>COVID-19 surveillance (only)</td>
<td>335</td>
<td>60</td>
</tr>
</tbody>
</table>

\(^a\)Displaying 22,328 of 63,343 records.

\(^b\)Displaying 8217 of 20,639 records.

\(^c\)eSURV: electronic surveillance.

**Frequency of eSURV and ISS Visits to Health Facilities**

The WHO recommends regular surveillance and supervisory visits to health facilities, with the frequency of visits being dependent on the priority level of the site (Table 3). The guidance suggests weekly visits to high-priority sites, including 3 eSURV and 1 ISS visit/mo, a visit every 2 weeks (1 eSURV and 1 ISS visit/mo) to medium-priority sites, and 1 ISS visit/mo to low-priority sites.

In the past, the numbers of actual eSURV and ISS visits have been reported to be lower than recommended, even in countries where ISS visits are required for regional certification [26]. Adherence to visit frequency recommendations appears, however, to be improving. Between 2020 and 2021, active eSURV visits increased by ~25,000, which permitted the identification of ~4000 missed AFP cases (Figure 3).

Table 3. Active surveillance site priority level and recommended visit frequency.

<table>
<thead>
<tr>
<th>Site priority</th>
<th>Target/wk</th>
<th>Target/mo</th>
<th>Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wk 1</td>
<td>Wk 2</td>
<td>Wk 3</td>
</tr>
<tr>
<td>Highest: structure or person located in a high-risk area (eg, refugee or IDP(^a) camp)</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>High: structure or person from which an AFP(^b) case would very likely seek care</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Medium: structure or person from which an AFP case would likely seek care</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Low: structure or person from which an AFP case would perhaps seek care</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\)IDP: internally displaced person.

\(^b\)AFP: acute flaccid paralysis.

\(^c\)Not available.
**Geospatial Data Entry**

When a field officer visits a health facility, the geolocation data are automatically transmitted to that country’s eSURV and ISS database. Field-collected data are synchronized on the server, and the dashboard is viewable by WHO and country health ministry personnel.

One of the challenges is that a single site may have multiple names in the database, resulting in site identification mismatches or multiple sites recorded for the same geolocation (Figure 4). To address this problem, eSURV includes data entry constraints, with a dropdown menu of the country’s MFL from the health observatory facility list or, directly, an inventory of the health facility list from the country team, which helps prevent data entry and server synchronization errors [19]. To allow for the entry of previously unrecorded sites or those missing from the list, the program includes an “Other” option, whereby data collectors can enter new data.
**Contribution of eSURV and ISS to Polio Eradication Efforts**

The eSURV and ISS apps provide real-time georeferenced surveillance and immunization data that can be accessed by polio program staff and used to audit data and evaluate data collector performance and accountability. In the past, paper-based systems were found to be not only slow, but also susceptible to data fabrication and omissions [19]. In contrast, real-time data transmission with mobile phone apps increases surveillance sensitivity and transparent data management increases WHO GPEI confidence in program results. The Open Data Kit (ODK) app currently faces some compatibility issues with iOS users. A few mechanisms have, however, been implemented to facilitate access for iPhone users. An example of this is the introduction of GIC Collect, which is compatible with iOS, thereby addressing some of these constraints. GIC Collect works as a port for the ODK for use with an iPhone.

The data provided through eSURV and ISS contributed to the attainment of WPV-free certification for the WHO African region [24]. For example, these tools were essential in AFP surveillance and immunization of >40,000 children of nomads living in 62 high-risk districts in and around the remote Lake Chad region (Figure 5) [25]. The eSURV and ISS data facilitated the identification and location of at-risk groups, who continue to be monitored for AFP cases [27].

Multiple studies have documented improvements in African polio public health interventions after the implementation of eSURV and ISS [19,26,28-32]. A preliminary analysis of the transition from a paper-based system to eSURV showed that the average case-reporting time in Africa decreased by 90% from a mean of 87.6 to 8.7 hours [32]. The time from case identification to entry into a polio case tracking database was reduced by 72% from a mean of 30 to 8 days. Another analysis by the same author group showed that the completeness of data collected via eSURV and ISS was >95% across 5 countries in sub-Saharan Africa, with error rates that ranged from 0.01% to 0.03%. The authors attributed the low error and high completion rates to inherent aspects of the software, including data entry constraints, skip logic, and enforced validation, as well as real-time data sharing. The latter enabled supervision and monitoring of the flow of data that permitted supervisors to detect data fabrication and omissions, which have been documented with paper-based systems [19].

In another study, the implementation of ISS in Zambia improved the updating of immunization records in 7 of 9 provinces studied, with up to a 53% increase in the proportion of health facilities with updated immunization records [26]. In addition, ISS had a statistically significant, positive impact on knowledge of AFP case definition and documentation of AFP case files.
Development of a Consolidated Health Facility Masterlist—Integration of eSURV and ISS With MFLs

Despite improvements in the frequency and accuracy of data collection with eSURV and ISS, several challenges have been identified. First, the total number of health facilities and TM sites in Africa has not been fully documented. The lack of a denominator of known sites limits the ministry of health and WHO personnel’s ability to assess the completeness of surveillance coverage. In addition, the lack of a validation mechanism for the reference list of health facilities has the potential to contribute to database errors. To address these gaps, the consolidated Health Facility Masterlist was developed and is managed by the WHO AFRO GIS Centre in close cooperation with the respective health ministries of each member state.

The Health Facility Masterlist uses existing MFLs provided by each country as references, which are compared with health facility lists compiled from eSURV and ISS data. In a structured process that is still ongoing, WHO AFRO surveillance and GIS personnel cross-reference eSURV and ISS lists from each country against that country’s MFL and consolidate the information for health facilities (including TM sites) with ≥95% similarity. Facilities with <95% similarity are returned to each country, where a designated focal person validates the site information using a health facility validation checklist web app that was developed for Masterlist consolidation and validation. The validated facility uses a standard operating procedure, developed at the operational level by the AFRO GIS Centre and validated by the country surveillance and data management team, to agree on the nomenclatures of a health facility name and health facility type as they differ from one country to another. Validated facilities are then added to the consolidated Health Facility registry form as updates. As data collectors discover and record new facilities (e.g., previously unrecognized TM sites) using the Site Registry form, the in-country focal person validates the facilities using the MFL validation checklist. After verification by the in-country teams (ministry of health and WHO), the new facilities are added to the consolidated health facility registry. The AFRO GIS Centre, in collaboration with the surveillance team at the country level, conducts updates on a priority level twice yearly. The Masterlist includes the following variables: site name, location based on administrative boundary, geocoordinates, priority level, type of site, level of care, operating status, vaccination status, ownership, and auto-generated unique identifier.

Once the consolidated Health Facility Masterlist is complete, the data can be used by the WHO AFRO to more accurately evaluate the completeness of surveillance and immunization efforts, as well as assess the performance of health facilities (Figure 6).

The Health Facility Masterlist is specifically designed for use at the operational level. Any digital surveillance or data collection activities carried out at the health facility level can make use of the health facility registry as a valuable resource. The primary objective is to continually update and enhance the registry, aiming to benefit all countries in the WHO African region. Additionally, the registry will be maintained as a web service, enabling automated integration with various forms such as eSURV, vaccine management, and other ODK forms that are administered at the health facility level, facilitating a comprehensive and streamlined approach. The primary
limitation of the use of the Health Facility Masterlist lies in elucidating the core method for updating the health facility process after consolidation. At present, the AFRO GIS Centre is working on a well-defined system commonly called the “eSURV Companion App” for maintaining, consolidating, and validating health facility data with eSURV. This issue will necessitate further examination in subsequent research related to this paper.

To prevent stagnation in consolidating the Health Facility Masterlist, the primary strategy is to uphold a robust active surveillance system through the use of eSURV, which is currently used by 46 countries in the WHO African region for reporting active surveillance visits. The exclusion of even one country, however, remains a significant hurdle in achieving full consolidation. Furthermore, a key recommendation is to encourage the adoption of the latest eSURV Companion App by 2024, ensuring comprehensive maintenance of the consolidated Health Facility Masterlist over time.

**Figure 6.** Schematic depiction of the development of consolidated Health Facility Masterlist. AGC: World Health Organization Regional Office for Africa Geographic Information Systems Centre; CFP: Country Focal Point; eSURV: electronic surveillance; ISS: integrated supportive supervision; ODK: Open Data Kit; SOP: standard operating procedure.

![Figure 6](image)

**Conclusions**

The eSURV and ISS mobile health solutions have played a vital role in efforts to eradicate polio and control other communicable diseases in Africa. Data compiled from eSURV and ISS over the last 5 years are being used to develop, update, and validate MFLs throughout Africa. Moreover, the ongoing consolidation of health facility information into a single health facility registry that covers the WHO African region should further improve disease surveillance and facilitate epidemiologic research. These efforts will support not only polio eradication in Africa, but also public health efforts directed at other diseases across the continent.

**Acknowledgments**

This paper was supported by the Bill & Melinda Gates Foundation. Writing and editorial support was provided by Amanda Justice and Geoff Marx of BioScience Communications, New York, New York, United States, funded by the Bill & Melinda Gates Foundation.

**Data Availability**

The data that support the findings of this study are openly available from the WHO AFRO GIS Centre at the Rapid Response Team Hub [33]. The authors confirm that the data supporting the findings of this study are available within this paper.

**Authors’ Contributions**

MABN handled the initial drafts, as well as paper and Health Facility Masterlist initiative conceptualization. KT did the planning and supervision. TJM was the active surveillance conceptualization lead and provided guidance. GAU provided Health Facility
Masterlist and eSURV conceptualization. RON provided Health Facility Masterlist standard operating system conceptualization. BIM provided Health Facility Masterlist and eSURV conceptualization. IR, DO, EMOB, and FL handled the Health Facility Masterlist implementation in the WHO African region. JK provided this paper’s coordination and support. HHWG dealt with administration and planning, resource mobilization, program management, and activity monitoring. VS provided resource mobilization, conceptualization, draft reviews and editing, resource mobilization training quality reviews and assurance, and supervision. MN provided administration and management, supervision, resource mobilization and approvals, and draft reviews and editing. JAA provided funding acquisition, editing, resource mobilization, supervision, administration, and management.

**Conflicts of Interest**
None declared.

**References**


Abbreviations

AFP: acute flaccid paralysis
AFRO: Regional Office for Africa
cVDPV: circulating vaccine-derived poliovirus
eSURV: electronic surveillance
GIS: Geographic Information System
GPEI: Global Polio Eradication Initiative
ISS: integrated supportive supervision
MFL: master facility list
ODK: Open Data Kit
TM: traditional medicine
WHO: World Health Organization
WPV: wild-type poliovirus
HIV Prevention and Treatment Interventions for Black Men Who Have Sex With Men in Canada: Scoping Systematic Review

Jemal Demeke1, BSc; Pascal Djiadeu2, PhD; Abban Yusuf1, MPH; Darren Lovell Whitfield3, PhD; David Lightfoot1, PhD; Fiqir Worku4, MSc; Gamji Rabiu Abu-Ba'are5, PhD; Lawrence Mbuagbaw2, PhD; Sulaimon Giwa6, PhD; LaRon E Nelson7, PhD

1MAP Centre for Urban Health Solutions, Li Ka Shing Knowledge Institute, St Michael’s Hospital, Unity Health Toronto, Toronto, ON, Canada
2Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada
3School of Social Work, University of Maryland, Baltimore, MD, United States
4Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada
5Center for Interdisciplinary Research on AIDS, Yale School of Public Health, New Haven, CT, United States
6School of Social Work, St John's College, Memorial University of Newfoundland, St John's, NL, Canada
7School of Nursing, Yale University, New Haven, CT, United States

Corresponding Author:
Jemal Demeke, BSc
MAP Centre for Urban Health Solutions
Li Ka Shing Knowledge Institute
St Michael’s Hospital, Unity Health Toronto
209 Victoria Street
Toronto, ON, M5B 1W8
Canada
Phone: 1 4163604000
Email: jemalwdemeke@gmail.com

Abstract

Background: Black men who have sex with men (MSM) experience disproportionately high HIV incidence globally. A comprehensive, intersectional approach (race, gender, and sexuality or sexual behavior) in understanding the experiences of Black MSM in Canada along the HIV prevention and care continuums has yet to be explored.

Objective: This scoping review aims to examine the available evidence on the access, quality, gaps, facilitators, and barriers of engagement and identify interventions relevant to the HIV prevention and care continuum for Black MSM in Canada.

Methods: We conducted a systematic database search, in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist, of the available studies on HIV health experience and epidemiology concerning Black MSM living with or without HIV in Canada and were published after 1983 in either English or French. Searched databases include MEDLINE, Excerpta, Cumulative Index to Nursing and Allied Health Literature, the Cochrane Library, the NHUS Economic Development Database, Global Health, PsycInfo, PubMed, Scopus, and Web of Science. From the 3095 articles identified, 19 met the inclusion criteria and were analyzed.

Results: Black MSM in Canada consistently report multiple forms of stigma and lack of community support contributing to an increased HIV burden. They experience discrimination based on their intersectional identities while accessing HIV preventative and treatment interventions. Available data demonstrate that Black MSM have higher HIV incidences than Black men who have sex with women (MSW) and White MSM, and low preexposure prophylaxis knowledge and HIV literacy. Black MSM experience significant disparities in HIV prevention and care knowledge, access, and use. Structural barriers, including anti-Black racism, homophobia, and xenophobia, are responsible for gaps in HIV prevention and care continuums, poor quality of care and linkage to HIV services, as well as a higher incidence of HIV.

Conclusions: Considering the lack of targeted interventions, there is a clear need for interventions that reduce HIV diagnoses among Black MSM, increase access and reduce structural barriers that significantly affect the ability of Black MSM to engage with HIV prevention and care, and address provider’s capacity for care and the structural barriers. These findings can inform future interventions, programming, and tools that may alleviate this HIV inequity.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-043055
Introduction

The Canadian publicly funded health care system aims to provide universal access to medically necessary services [1]. This equitable goal alludes to Canadian values of equity and fairness in the administration of health and social services and serves as a source of collective pride for many Canadians [1,2]. However, this narrative does not reflect the state of access to Canadian health care services and the poor health outcomes experienced by marginalized populations [3-10]. These gaps are demonstrated through the HIV inequities seen in Black (descendants of Africa, African diaspora communities in Canada and the Caribbean) communities in Canada. For example, although Black Canadians comprise less than 4% of the national population, according to the 2017 HIV Surveillance report, they made up 25% of new HIV diagnoses [11]. In Ontario alone, between 2012 and 2017, the proportion of new HIV diagnoses increased for Black men while it decreased for White men [12]. Moreover, HIV exposure for gay, bisexual, and other men who have sex with men (MSM) represents the most significant proportion of reported adult HIV diagnoses [11]. Black MSM exist at the intersection of these 2 populations that demonstrate disproportionately high HIV incidence in Canada. Thus, sexual behavior and racialization are significant factors to HIV transmission in adult male HIV diagnoses; this illustrates the relationships between social identities and HIV health inequities in Canada.

Black MSM experience systemic anti-Black racism in addition to the sexual minority stigma and homophobic discrimination [13]. As other high-income countries work to address these inequities, it becomes increasingly important to identify the drivers of HIV inequity through the intersectional experiences of Black MSM in Canada and develop effective evidence-informed intervention models [14-16]. Unfortunately, HIV surveillance by Canadian federal and provincial health agencies does not provide the necessary disaggregated data to elucidate HIV transmission patterns for Black MSM fully. For example, only half of the Public Health Agency of Canada reported cases of integrated exposure category and race data in 2016 and 2017, while the 2018 report did not report this [11,17,18]. Even so, the limited available data reveal that Black MSM are overrepresented in the new HIV diagnoses [11]. Canada aimed to meet the Joint United Nations Programme on HIV/AIDS global targets by 2020 and significantly reduce HIV transmission by 2030; to achieve this goal, Canada must develop a clear strategy to respond to the increased burden of HIV for Black MSM.

Socially constructed identities, including assignment to racial and sexual minority or gendered groups, are associated with inequitable and disproportionate HIV infections [19]. The factors that lead to poor health outcomes are linked to intersectional and systemic barriers, including inaccessible health care services for Black MSM. Additional factors driving poor health outcomes for Black MSM include precarious employment and housing, stress-related mental health challenges, low sexual health and HIV literacy, as well as poor psychosocial health and general health [20-22]. These factors also limit the ability of Black MSM to meaningfully and consistently engage with HIV health care services [23]. Moreover, Black MSM are exposed to anti-Black racism and homophobia within health care institutions, resulting in psychological trauma and medical trauma [13,21,24-27]. Thus, Black MSM are ultimately prevented from accessing HIV-related services by inequitable care, discrimination in health care settings, and systemic marginalization [13,23]. Yet, as these powerful barriers persist, evidence demonstrates that multilevel interventions have successfully mitigated HIV inequity by improving the accessibility to and quality of health care services [24,28,29].

The current HIV prevention and treatment interventions are considered as a continuum of services [30]. This provides a framework through which key points in prevention, care, and existing gaps are engaged [30]. Specifically, the HIV prevention, engagement, and care cascade models the steps to prevent HIV infection or receive treatment if diagnosed with HIV [30-32]. This cascade can be used to identify the gaps in HIV prevention and care and the necessary interventions. However, Canadian HIV surveillance and response effortsquantify the state of HIV transmission and service access by compartmentalizing gender, sexual minority status and sexuality, and race as individual factors. This approach does not accurately characterize the intersectional marginalization of Black MSM. Intersectionality, coined by Crenshaw [33] in 1989, established that singular identities could not explain the complex forms of oppression experienced by marginalized peoples.

Similarly, in the context of Black MSM, their experiences must investigate their gender, sexuality, and race holistically. This may then inform HIV prevention and care interventions and strategies to reduce HIV incidence for Black MSM effectively. This intersectional approach to HIV research, however, has not been widely examined [34]. The results from such investigation could promote access to and improve the quality of prevention and care along the continuums for Black MSM in Canada. Nonetheless, the limited availability of HIV interventions, research, and literature for Black MSM in Canada necessitated this investigation into past and recent HIV research as well as health care provision of HIV-related services for this vulnerable population.

This scoping review examines the state of HIV-related research and health care provision for Black MSM in Canada. Overall, this review aims to comprehensively evaluate and determine the state of HIV prevention and care for Black MSM in Canada as a public health topic while incorporating the lived experience of Black MSM that influence health care engagement. The primary objective is to assess the available literature regarding the influence, access to, and quality of HIV prevention and treatment for Black MSM living in Canada. Secondary
Objectives are to explore the facilitators and disincentives or barriers to HIV-related health care services for Black MSM and the mechanisms through which they influence retention and adherence to the HIV care continuum. Although this review is focused on Black MSM in Canada, its findings will be useful for ethnic minority MSM anywhere in the world.

Methods

Overview
The current investigation used a scoping systematic review methodology. Peterson et al [35] identified this approach as a methodology to advance emerging research topics. Standardized systematic reviews and scoping reviews differ in terms of the specificity of their topic of interest. Scoping reviews tend to investigate broad topics compared to standardized, systematic reviews, which examine a specific, detailed question.

Criteria for Including Studies
This review included evidence syntheses and qualitative, experimental (randomized or nonrandomized), observational (longitudinal and cross-sectional), and mixed methods studies. For inclusion in the review, studies must (1) include data on self-identified Black MSM living with or without HIV, (2) research HIV prevention and care among Black MSM, and (3) have been published in at least 1 of the 2 official languages in Canada—English and French.

For exclusion in the review, studies must not (1) focus on Black MSM outside of Canada unless Canadian data are analyzed separately and (2) be published before the 1983 formation of Canada’s AIDS National Task Force.

Objectives of Interest
The primary objectives of interest of this scoping, systematic review include (1) identifying existing literature that examines the HIV prevention and care continuum in Canada for Black MSM, including epidemiological trends, determinants of engagement, and health care experience and (2) determining health access and availability of specific health resources related to the Canadian HIV prevention and treatment interventions for Black MSM.

The secondary objectives of interest include (1) examining the effects of social determinants of health on the HIV prevention and treatment interventions for Black MSM in Canada, (2) consolidating research on unilevel and multilevel interventions that address the social determinants of health for Black MSM in Canada, and (3) identifying existing health promotion for Black MSM in Canada.

Patients and Public Involvement
There was no direct involvement of the public and patients or participants who are Black MSM in the study.

Ethical Considerations
The approval of the Research Ethics Board was not required. As a scoping systematic review, information was based on secondary published data and not on human subjects. Peer-reviewed manuscripts, conference presentations, and students’ rounds can be strategically used to disseminate study findings. The findings may also influence policy within government health agencies and local HIV or AIDS service organizations.

Search Strategy for Identification of Studies
A health sciences librarian conducted a comprehensive literature search of studies published through the Health Sciences Library in Toronto, Ontario’s St Michael’s Hospital, Unity Health Toronto. This occurred in April 2020. The list of our search terms included “Quality of Health Care,” “Health Status Disparities,” “Social Stigma,” “Human Immunodeficiency Virus,” “Same-Sex Intercourse,” “Blacks OR African OR Caribbean.” The complete search strategy can be found in the published protocol [36].

Electronic Searches
These searches were done on MEDLINE, Excerpta, Cumulative Index to Nursing and Allied Health Literature, the Cochrane Library, the NHUS Economic Development Database, Global Health, PsycInfo, PubMed, Scopus, and Web of Science.

Reference Lists
Related articles were searched for in the reference lists of all pertinent citations.

Gray Literature
Available thesis and conference posters were searched while reports from relevant organizations such as the African and Caribbean Council on HIV or AIDS in Ontario (ACCHO), Black Coalition for AIDS Prevention (Black CAP), Africans in Partnership Against AIDS (APAA), Committee for Accessible AIDS Treatment (CAAT), TAIBU Community Health Centre, Ontario HIV Treatment Network (OHTN), and Canada’s source for HIV and hepatitis C information were explored.

Screening
The studies were deduplicated in advance, and the web-based app Rayyan QCRI (Rayyan) was used to import and screen citations found using this search strategy [37].

A data collection form customized to reflect the inclusion criteria was pilot tested by 2 independent reviewers. These reviewers played a role in generating and using the given form. A total of 50 abstracts were used as a sample to create consistency of use and establish the instrument’s validity. Interrater reliability was measured using the Cohen κ statistic. The screening began upon the achievement of a 60% (n=979) agreement [38].

The study’s selection process began with title and abstract screening to identify potentially relevant articles. This was followed by retrieving the full text for detailed screening using the inclusion and exclusion criteria before the data extraction procedure. All screening and data extraction were completed in duplicate and blinded by JD, PD, FW, AY, and GRAB. Any disagreements were settled via consensus. However, a third author was available to arbitrate (SG) if an agreement could not be reached.
Data Extraction

Bibliometric information such as author names, journal, year of publication, the study location, design, number of participants, outcomes reported, outcome measures overall, and outcome measures in Black MSM participants were extracted. In addition, each outcome is reported through measures of magnitude mean (SD) or percent (95% CIs) where possible, comparing the effect of the intervention in Black MSM versus men in other racialized groups or Black MSW; (odds or risk ratios, mean differences, accompanied with 95% CIs) [39].

Assessment of Methodological Quality of the Included Studies

We did not appraise the methodological quality and risk of bias in the studies as this is not required in a scoping review [40].

Analyses and Reporting

Study findings were reported as per the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [41,42] through the use of narratives and tables. This is outlined in Multimedia Appendix 1. Data were grouped according to outcomes, with the number of studies and their design displayed using tables. In addition, a narrative synthesis of the data was conducted to identify overlapping themes and knowledge gaps.

Results

Results of the Search

Overview

The literature search of the database identified 3085 studies, with 10 studies found through gray literature and other sources (Figure 1). Thus, a total of 1630 studies were screened after the duplicates were removed. A total of 1607 studies were excluded as they did not study Black MSM in Canada or did not include any data on HIV prevention and treatment interventions. Of the remaining 24 studies, 4 were removed during full-text extraction; among them, 2 were abstracts, 1 study reported Canadian and American Black MSM together, and the other 2 did not provide information on HIV. This left 19 studies that met the inclusion criteria and were included in the study.

Figure 1. PRISMA flowchart. MSM: men who have sex with men; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Characteristics of the Study

There were 8 cross-sectional descriptive studies, 3 qualitative studies, 2 mixed methods studies, 1 meta-analysis, and 5 organizational reports. The studies’ publication years ranged from 2004 to 2020. The study characteristics are summarized in Table 1. Further description of each study is provided in Multimedia Appendix 2.
Table 1. Summary of results by themes and number of studies included.

<table>
<thead>
<tr>
<th>Theme or subtheme</th>
<th>Studies included, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Engagement with HIV prevention and care cascades</strong></td>
<td></td>
</tr>
<tr>
<td>HIV epidemiology and its driving factors</td>
<td>4</td>
</tr>
<tr>
<td>Factors that influence HIV testing</td>
<td>7</td>
</tr>
<tr>
<td>Preexposure prophylaxis use among Black MSM</td>
<td>2</td>
</tr>
<tr>
<td><strong>Experiences in health care for Black MSM</strong></td>
<td></td>
</tr>
<tr>
<td>Discrimination in HIV health care</td>
<td>3</td>
</tr>
<tr>
<td>Inadequate HIV health care provision</td>
<td>2</td>
</tr>
<tr>
<td>Missed opportunities for linkage to HIV prevention and care</td>
<td>3</td>
</tr>
<tr>
<td><strong>Social determinants of health and HIV health care access</strong></td>
<td></td>
</tr>
<tr>
<td>Stigma and homophobia</td>
<td>6</td>
</tr>
<tr>
<td>Employment, income, and housing</td>
<td>6</td>
</tr>
<tr>
<td>Immigration</td>
<td>3</td>
</tr>
<tr>
<td>Education and access to knowledge about HIV prevention and intervention methods</td>
<td>3</td>
</tr>
<tr>
<td>Mental health and emotional well-being</td>
<td>6</td>
</tr>
<tr>
<td><strong>Canadian HIV interventions for Black MSM</strong></td>
<td></td>
</tr>
<tr>
<td>Biomedical interventions</td>
<td>5</td>
</tr>
<tr>
<td>Behavioral interventions</td>
<td>5</td>
</tr>
<tr>
<td>Structural interventions</td>
<td>1</td>
</tr>
</tbody>
</table>

aMSM: men who have sex with men.

**Reporting of Results**

The results of the studies included in the scoping systematic review can be grouped into 4 themes identified from the data and are discussed below. They are the engagement with HIV prevention and care cascades, experiences in HIV health care services, the effects of the social determinants of health on HIV prevention and care access, and interventions for Black MSM.

**Engagement With HIV Prevention and Care Cascades**

**Overview**

A total of 10 studies explored the engagement of Black MSM with the various steps in the HIV prevention and care continuum.

**HIV Epidemiology and its Driving Factors**

Multiple studies assessed HIV infections for Black MSM and their predictors. In Ontario, Black MSM reported the highest increase of HIV diagnoses among racialized MSM, compared to the decreasing HIV diagnoses for White MSM over 2009-2010 and 2011-2012 [43]. This disproportionate increase for Black MSM is supported in other studies in this review [44,45]. Moreover, 2 studies explored differences in HIV diagnoses and associated factors between Black MSW and Black MSM in Toronto. Black MSM demonstrated considerably higher HIV and syphilis prevalence compared to Black MSW [44,46]. Significant HIV diagnosis predictors were the number of male partners (6 or higher), older than 24 years of age, and syphilis diagnoses [44,46]. In addition, Black MSM showed higher transmission of sexual and blood-borne infections (SBBIs) as well, including chlamydia, herpes simplex virus (HSV) 1, HSV-2, and active hepatitis B (HBV) infection compared to other MSM [44,45].

**Factors That Influence HIV Testing**

HIV testing behaviors among Black MSM were also investigated. Generally, African, Caribbean, and Black MSM had lower HIV testing rates than other ethnicities [43,47]. Straight-identified Black MSM were more likely to have had an HIV test than gay-identified Black MSM [45]. Canadian-born Black MSM were less likely to get tested for HIV [48]. Several studies assessed factors that influence HIV testing for Black MSM. Black MSM were shown to have a stronger likelihood of being tested for HIV if they were older than 21 years of age, engaged in condomless sex, had a history of a chlamydia infection, or had relatives or friends who live with or died from HIV [44,48]. Caribbean- and African-born MSM were more likely to know someone who currently lives with HIV or know someone who died from complications due to HIV than Canadian-born Black MSM [49,50]. Nontesters commonly cite low perception of risk and practicing safe sex as reasons for not testing for HIV [48]. Overall, Black MSM are most likely to be tested through their family doctor [48].

**Preexposure Prophylaxis Use Among Black MSM**

It is important to highlight that preexposure prophylaxis (PrEP) is a highly effective oral antiretroviral medication used in the prevention of HIV [51]. While the implementation of PrEP through the Canadian guidelines remains controversial due to...
concerns about accessibility for Black populations. PrEP is identified as a promising tool to reduce HIV incidence in Canada [52,53].

This review finds that similar to the trends seen with HIV testing, Black MSM in younger age groups and Canadian-born were less likely to accept PrEP [54]. However, Black MSM were found to accept PrEP at greater rates than Black MSW and similar rates to other MSM populations [54]. As a whole, PrEP acceptance was not significantly associated with self-risk perception [54]. More than half of Black MSM could not accurately describe their actual risk for HIV acquisition [54]. For those who did not accept PrEP, the most popular explanations for not accepting PrEP were concerns about side effects and low perception of risk for HIV transmission [54,55].

Experiences in Health Care for Black MSM

Overview
A total of 4 studies assessed the specific experiences that Black MSM have while accessing HIV health care services in Canada and the quality of care in the context of their intersectional experiences as Black MSM.

Discrimination in HIV Health Care
Black MSM consistently reported discrimination and inequitable access to HIV care and preventative interventions like PrEP [55-57]. In general, Black MSM noticed that health care staff neglected them in favor of White patients [55,56]. Black MSM who self-identify as gay, bisexual, transgender, or queer (GBTQ+) felt that their Blackness remained a barrier to care [55,57]. Furthermore, upon entering Black-centered health care organizations, many Black MSM often did not feel comfortable disclosing their sexual practices [55,57].

Inadequate HIV Health Care Provision
Black MSM describe low-quality and uncompassionate care, which facilitates gaps in HIV care and disrupts retention of care [56]. In addition, health care providers often use scientific or medical terms that are too complex [56]. This contributes to a more prominent theme of depersonalized care, where health care providers appear neither genuine nor have investments in their sexual health [55,56]. Moreover, mental health and emotional well-being were not incorporated into the HIV care provided to Black MSM, especially when health care providers were disclosing HIV diagnoses [56].

Missed Opportunities for Linkage to HIV Prevention and Care
Health care providers fail to inform Black MSM about the HIV resources and the prevention and treatment interventions that they could access, despite health care providers being described as a key source of sexual health information [55,58]. Black MSM needed to advocate for their care in response to the negligence of their health care providers [55,56]. GBTQ+ Black MSM who seek sexual health services from queer-friendly providers often have long wait times for services, which affects HIV care adherence [55]. Furthermore, many Black MSM did not know about PrEP or had inaccurate knowledge about PrEP despite regular HIV testing [55].

Social Determinants of Health and HIV Health Care Access

Overview
A total of 10 studies reported various social factors as facilitators and barriers to HIV care and prevention services for Black MSM.

Stigma and Homophobia
HIV stigma and sexual minority stigma are significant stressors for Black MSM and impair HIV prevention and care access and engagement [56,59]. Also, Black MSM living with HIV fear stigmatization and rejection can reduce their likelihood of disclosing their HIV status to their sexual partners [56]. GBTQ+ Black MSM have an even increased stigma burden due to homophobia [45,57,58]. The fear of ostracization from their communities also prevents Black MSM from talking about their sexual practices and HIV as a whole [57]. HIV stigma and sexual minority stigma also affect their perception of risk for HIV acquisition [59]. Specifically, heterosexual-identified Black MSM internalize messaging that associates HIV risk with being gay, compromising their own perception of risk [56,57]. For instance, “PrEP stigma” is prevalent as many Black MSM associate PrEP with promiscuity [59]. This stigma translates to increased concerns about PrEP and decreased knowledge about PrEP [59].

Regardless of sexual identity, these stigmas make Black MSM apprehensive about disclosing their sexual practices and deter them from accessing critical HIV interventions [45,55,56,58]. This limits the providers’ ability to inform Black MSM about HIV prevention interventions usually promoted to MSM [55]. Black MSM are also concerned about judgment upon entering HIV health care organizations or having their privacy and confidentiality breached [45,56].

Employment, Income, and Housing
Black MSM experience hardship securing full-time and part-time work, and many report unemployment as a priority concern [50,57]. Black MSM in the MaBwana study had lower incomes than the average income of a sample of MSM living in Ontario [50]. This may contribute to difficulties with securing housing and poverty [57]. These disparities affect their use of HIV prevention interventions such as PrEP, which is the most significant biomedical HIV prevention intervention [49,60]. The cost was described as a significant barrier to PrEP use—even if Black MSM were willing to accept PrEP, they report that its unaffordability prevents them from discussing PrEP with their health care providers [55,59]. Nonetheless, while there are various ways to access PrEP through public and private insurance, many Black MSM are unaware of PrEP coverage to mitigate PrEP costs [55,59].

Immigration
Migrants have distinct experiences in accessing HIV prevention and treatment interventions. While many Black MSM favor Canadian health care, they also describe language barriers and stress with navigating complex immigration laws [56]. Without having the necessary immigration documentation, some Black MSM have been denied medication and HIV care [56]. Many
Black MSM born outside of Canada have low HIV prevention literacy and have experienced medical trauma [58]. Also, they may doubt the efficacy of PrEP by emphasizing its adverse effects [58]. Moreover, migrant GBQT+ Black MSM face anti-Black racism and xenophobia in White gay communities, where more people may be using or know about PrEP [50]. This leaves many migrant Black MSM without the HIV and PrEP knowledge other communities may have [50].

**Education and Access to Knowledge About HIV Prevention and Intervention Methods**

The MaBwana study demonstrated lower achievement of formal education for Black MSM than other MSM [50]. In addition, some Black MSM disclose low reading and writing skills, which limit their ability to learn about important sexual health and HIV information and interventions [56]. Even when they can meaningfully engage with educational institutions, Black MSM describe heteronormative sexual education and a significant lack of GBQT+ representation [58]. Thus, they do not learn about biomedical HIV interventions and prevention methods relevant to their sexual practices [58].

**Mental Health and Emotional Well-Being**

The stigmatization of the sexual practices of Black MSM by their cultural communities has a considerable effect on their mental health [57]. This ostracization can lead Black MSM to experience feelings of isolation, alienation, and internalized homophobia. It also contributes to higher levels of depression and psychological distress [49]. These factors can impair the ability of Black MSM to access HIV treatment and care [45,50,57,61,62]. These dynamics act as barriers for Black MSM to meaningfully engage with HIV biomedical preventative interventions and mitigate the transmission of HIV [49]. Furthermore, research on Black MSM in the United States has shown that the intersection of discrimination experiences related to race or ethnicity, sexual orientation, and HIV status can significantly mediate the relation between socioeconomic status and mental health concerns [63]. These findings suggest that having low socioeconomic resources could increase exposure to discrimination and worsen mental health.

**Canadian HIV Interventions for Black MSM**

**Biomedical Interventions**

Condoms remain the most accessible biomedical intervention for Black MSM [49,50,55]. Initiatives that promote condom use in settings where Black MSM engage in unplanned sex or with casual sex partners may increase condom accessibility [49,50]. The Black CAP provided condoms through their bathhouse outreach program and similar programs in other places frequented by Black MSM like clubs, barbershops, Pride, and other community events [64]. The Black CAP developed a partnership with a local sexual health clinic, Hassle-Free Clinic, to implement programming that provided quick and easy access to HIV and syphilis testing for Black MSM [65].

**Behavioral Interventions**

Several behavioral interventions identified through these studies aim to improve risk perception and HIV literacy, promote HIV prevention services, and adapt sexual behaviors to mitigate HIV transmission [49,50,59]. For example, a workshop was provided to Black-Canadian MSM and focused on increasing knowledge of sexually transmitted infections and factors contributing to their transmission and promoting sexual health service use [59]. The ACCHO campaign, “Keep It Alive,” was a multimedia initiative in 2006 that used postcards, posters, and popular gay media to increase HIV literacy and knowledge of HIV in Black communities, promote testing, and reduce HIV stigma. Likewise, the “Be Real” campaign by the Ontario Gay Men’s Strategy in 2006 had similar aims as “Keep It Alive” and used related dissemination methods. Black MSM described both campaigns as significant in providing information about sexual behavior habits and reducing the risk for HIV transmission [50].

Black CAP’s MSM Outreach program focuses on African, Caribbean, and Black MSM and aims to increase HIV awareness, HIV testing rates, and access to HIV resources and services [66]. They facilitate regular group discussions and sexual health education for Black MSM. This expands on their bathhouse outreach program, where they continue to distribute sexual health materials in bathhouses and Pride events as well as in webpages designed for GBQT+ Black MSM. Moreover, Black CAP’s Men’s Prevention Program provides culturally responsive HIV-focused workshops to Black MSM and includes homophobia, community support, and self-esteem. Also, they developed a resource to promote PrEP knowledge, use, and access for Black MSM in Toronto [65].

**Structural Interventions**

Black CAP reported initiatives that aim to reduce social inequity and other driving factors for HIV transmission for Black MSM [65]. For example, they linked Black MSM with support services that increased access to food, housing, income, and immigration aid. They also created an employment program, “Be Your Own Boss Series,” which educates Black trans-MSM on entrepreneurship and supports developing skills to develop and maintain a successful business.

**Discussion**

**Principal Findings**

Overall, the purpose of this study was to evaluate engagement with HIV services and interventions, identify the relevant facilitators and barriers via the social determinants of health, and determine the research and available interventions for Black MSM in the HIV prevention and care cascade through the lens of intersectional, overlapping identities. This scoping review presents a comprehensive evaluation of the state of the science as it pertains to HIV prevention and treatment interventions for Black MSM in Canada through available literature and reports. Through the 19 published studies and reports included, this review has identified critical points in the HIV prevention and care cascades, where Black MSM experience poor outcomes in knowledge, access, engagement, and quality of HIV care and prevention.

PrEP remains the most accessible and well-established biomedical intervention available in Canada. Yet, there were no identified PrEP delivery strategies that target Black MSM in this study. To inclusively reduce HIV incidence in Canada,
it is crucial to identify the factors that promote PrEP access and use. Motivations to use these prevention methods are characterized by the perception of HIV risk, adequate knowledge of the intervention and its benefits, and acceptability within the relevant sociocultural contexts for Black MSM. Limited discussion on HIV risk and prevention methods with health care providers contributes to the low-risk perception and PrEP motivation [55,56,58]. For instance, some health care providers refuse to discuss PrEP with their patients due to being uncomfortable discussing sex, sexuality, and sexual health with same-gender-loving patients [67,68]. Moreover, social stigmas including homophobia, racism, and xenophobia may affect motivation to seek HIV preventative services. These experiences can prevent Black MSM from disclosing their sexual practices, which providers need to assess HIV risk. Ultimately, these gaps expose the need for timely interventions for Black MSM [55,56,58]; priority areas include promoting the relevance of PrEP to health care providers, highlighting culturally responsive strategies, and generating space spaces actively for Black MSM to engage with providers honestly.

Access in the prevention cascade is influenced by numerous factors, including income, employment, proximity to health care facilities, housing, and other social determinants of health [64,69,70]. Many Black MSM describe precarious work and housing statuses, low income, issues related to poverty, and difficulty accessing formal education [50,57]. These disparities are associated with inadequate access to HIV prevention interventions, as high PrEP costs consistently reduce PrEP uptake [49,55,59]. More research is necessary to fully elucidate the mechanisms through which key factors influence health care engagement such as income, housing, and mental health and emotional well-being [71,72]. In the literature, MSM living with HIV experienced depression associated with physical, educational, social, financial, psychological, and short- and long-term health consequences [69]. People living with HIV who experienced depression are exposed to poor health outcomes like poor quality of life and worsening of their disease states. This could have even more consequences on Black MSM who are already impacted by other structural factors [70]. Therefore, if depression is untreated in MSM living with HIV, this can lead to risky sexual behavior, alcohol, and drug misuse and abuse, including suicide [73]. Poor adherence to antiretroviral drugs has also been associated with depression in newly diagnosed people living with HIV, resulting in poor immunological and virological outcomes.

To reduce HIV incidence for Black MSM, interventions that increase access might be beneficial including mass distribution and outreach programs, integrated health services, or legal changes that empower Black MSM to engage with HIV prevention interventions [74]. There is still limited research on intervention strategies to increase access to prevention interventions for Black MSM across Canada. Similar research in the United States investigated specific legislative and structural opportunities for Black MSM to successfully engage in HIV prevention [75,76]. Canada should commit to adopting effective programs in other countries or funding Canadian research that uses implementation science to address the identified challenges in this review for Black MSM.

While HIV prevention and care services exist along the same continuum, HIV care differs from the prevention cascade. People who remain at risk and test negative for HIV need to be linked to prevention services, but an HIV diagnosis requires a specific set of services to maintain viral suppression. The HIV care cascade models HIV testing and diagnosis steps as the starting point to viral suppression through a series of HIV health care services [30]. In comparison to other MSM, Black MSM demonstrate lower testing rates [43,47]. Still, this review shows various HIV testing behaviors among Black MSM dependent on age, sexuality, country of birth, religious affiliation, history of SBSI, and even personal relationships that have been affected by HIV [44,48-50]. Fortunately, this can inform the generation of testing interventions targeting Black MSM subgroups. For instance, many Black MSM were found to receive HIV testing from their family doctor, which can be used to build programming to link Black MSM to HIV tests directly from their primary care physician [48]. Outside of Canada, there have been many HIV testing interventions for Black MSM and other priority populations that have improved HIV testing and knowledge, testing access, and social support [77-81].

Moreover, HIV diagnosis rates are higher among Black MSM than other MSM and Black MSW [43-45]. Once diagnosed with HIV, Black MSM must be linked to HIV primary and specialist care services and antiretroviral therapy to achieve viral suppression [30]. While there were no studies on engagement at these stages by Black MSM, studies address the discriminatory barriers to access for Black MSM [55-57]. Nevertheless, the use of HIV care interventions is shown to be effective with other populations through improvements in retention of care, engagement with providers, and even condom use [82-85]. This review did not identify any HIV care interventions for Black MSM, which exposes the health care gap for Black MSM despite the epidemiological significance. Canada’s goal to dramatically reduce HIV incidence by 2030 is unachievable unless Black MSM are explicitly highlighted by service agencies.

The unilevel interventions mentioned in this study were reported in organizational reports. These are Black community–based initiatives driven by Black-focused organizations such as ACCHO, Black CAP, and APAA. Although they have a strong engagement and trust with Black communities, they are limited by a lack of research to support their ability to scale up including a focus on their effectiveness, feasibility, appropriateness, and vitality. Health care organizations, government, and research funding agencies need to invest in interventions led by community stakeholders to reduce HIV incidence in Black MSM. Black MSM in Canada are entitled to an equitable standard of care and should not be confined to select organizations outside of provincially mandated health service providers.

This review provides information that can be used to develop many interventions, including programs that provide support for mental health, immigration, housing, and employment; train providers to deliver culturally responsive HIV prevention and care services; and provide accessible and affordable HIV testing and PrEP services. Combining systems-level, provider-level, and patient-level interventions may be necessary for reducing
the effect of these factors, as evidence suggests these interventions improve HIV-related outcomes in Black and nonheteronormative communities [86-88]. Targeting critical points in the HIV prevention and care continuum may lead to improvements for Black MSM. Thus, there is a strong need for unilevel and multilevel HIV prevention and treatment interventions as well as more comprehensive clinical research studies that investigate the outcomes of these interventions across Canada for Black MSM.

Limitations
The findings of this scoping review are limited by the fact that all study interventions were conducted in the Greater Toronto Area, with English as the dominant language. This approach may not be practical for capturing the diverse experiences of Black MSM across Ontario and in other Canadian provinces and territories. In addition, as no French studies were identified in this review, the specific barriers that Francophone Black MSM experience may not be included.

Conclusions
MSM and Black communities are priority populations most affected by HIV in Canada. Black MSM experience disparities in HIV incidence and are disproportionately affected by the driving factors of HIV transmission, including stigma and discrimination. Singular-focused MSM or Black-specific interventions may not be effective for Black MSM, as they need intersectional intervention strategies that affirm their racial and sexual identities and sexual practices. This review demonstrates evidence that Black MSM experience structural discrimination, stigma, and poor quality of care and linkage to HIV services, and as a result, higher HIV and other SBBIs impact. There is some evidence of HIV interventions in Canada for Black MSM. Still, none of the interventions identified in this scoping review were assessed by randomized control trials to evaluate their effectiveness in alleviating HIV transmission disparities for Black MSM. Thus, an important next step is to develop evidence-based HIV interventions to reduce HIV inequities for Black MSM in Canada. The benefit of this review is that it provides a comprehensive overview of the barriers for Black MSM in HIV care from the individual to environmental, institutional, and structural. Additionally, this review has highlighted gaps in care that need to be addressed and sheds light on the intersectional interplay of identities in Black MSM care. This review can inform the development of these interventions for Black MSM and outline the factors that need to be addressed. Furthermore, community agencies and researchers can use these findings to create the necessary programming and tools to reduce HIV incidence for Black MSM.

Acknowledgments
This paper was supported by the Ontario HIV Treatment Network Applied HIV Research Chair in HIV Program Science with African, Caribbean, and Black Communities (#AHRC-1066) as well as the Yale University Centre for Interdisciplinary Research on AIDS, National Institutes of Health-funded program (P30 MH062294).

Authors’ Contributions
LEN, PD, and JD contributed to the study conception and design. Study screening was performed by JD, PD, AY, FW, and GRAB. The analysis was performed by JD. Reporting was performed and informed by JD, PD, DLW, LM, and SG. The first draft of the paper was written by JD and all authors commented on previous versions of the paper. All authors read and approved the final paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA-ScR appendix.
[PDF File (Adobe PDF File), 549 KB - publichealth_v10i1e40493_app1.pdf ]

Multimedia Appendix 2
Results of scoping review search: quantitative studies and meta-analysis, qualitative studies, and reports.
[DOCX File, 22 KB - publichealth_v10i1e40493_app2.docx ]

References

https://publichealth.jmir.org/2024/1/e40493


47. Liu J, Remis RS, Myers T, Husbands W. Special report: ethnicity analysis in the lambda survey of men who have sex with men, Ontario 2007. Dalla Lana School of Public Health, University of Toronto and AIDS Committee of Toronto. 2009. URL: https://www.dlsph.utoronto.ca/ [accessed 2023-12-07]


55. Absalom D, Boyce T. One of these things ain't like the other: exploring the HIV prevention needs of young adult Black same gender loving men—a report prepared for the Gay Men's Sexual Health Alliance. Gay Men's Sexual Health Alliance. 2020. URL: https://gmsm.ca/wp-content/uploads/2022/03/One-of-these-things_GMSH_cooper_August-18.pdf [accessed 2023-12-07]


Abbreviations

**ACCHO:** African and Caribbean Council on HIV or AIDS in Ontario  
**APAA:** Africans in Partnership Against AIDS  
**Black CAP:** Black Coalition for AIDS Prevention  
**CAAT:** Committee for Accessible AIDS Treatment  
**GBTQ+:** gay, bisexual, transgender, or queer  
**HBV:** hepatitis B  
**HSV:** herpes simplex virus  
**MSM:** men who have sex with men  
**MSW:** men who have sex with women  
**OHTN:** Ontario HIV Treatment Network  
**PrEP:** Preexposure prophylaxis  
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews  
**SBBI:** sexual and blood-borne infection
Changes to Public Health Surveillance Methods Due to the COVID-19 Pandemic: Scoping Review

Emily C Clark¹, MSc; Sophie Neumann¹, MPH; Stephanie Hopkins¹, MPH; Alyssa Kostopoulos¹, MPH; Leah Hagerman¹, MPH; Maureen Dobbins¹, RN, PhD

¹National Collaborating Centre for Methods and Tools, Hamilton, ON, Canada
²School of Nursing, McMaster University, Hamilton, ON, Canada

Corresponding Author: Maureen Dobbins, RN, PhD
National Collaborating Centre for Methods and Tools
175 Longwood Rd S
Suite 210A
Hamilton, ON, L8P 0A1
Canada
Phone: 1 9055259140 ext 20450
Email: dobbinsm@mcmaster.ca

Abstract

Background: Public health surveillance plays a vital role in informing public health decision-making. The onset of the COVID-19 pandemic in early 2020 caused a widespread shift in public health priorities. Global efforts focused on COVID-19 monitoring and contact tracing. Existing public health programs were interrupted due to physical distancing measures and reallocation of resources. The onset of the COVID-19 pandemic intersected with advancements in technologies that have the potential to support public health surveillance efforts.

Objective: This scoping review aims to explore emergent public health surveillance methods during the early COVID-19 pandemic to characterize the impact of the pandemic on surveillance methods.

Methods: A scoping search was conducted in multiple databases and by scanning key government and public health organization websites from March 2020 to January 2022. Published papers and gray literature that described the application of new or revised approaches to public health surveillance were included. Papers that discussed the implications of novel public health surveillance approaches from ethical, legal, security, and equity perspectives were also included. The surveillance subject, method, location, and setting were extracted from each paper to identify trends in surveillance practices. Two public health epidemiologists were invited to provide their perspectives as peer reviewers.

Results: Of the 14,238 unique papers, a total of 241 papers describing novel surveillance methods and changes to surveillance methods are included. Eighty papers were review papers and 161 were single studies. Overall, the literature heavily featured papers detailing surveillance of COVID-19 transmission (n=187). Surveillance of other infectious diseases was also described, including other pathogens (n=12). Other public health topics included vaccines (n=9), mental health (n=11), substance use (n=4), healthy nutrition (n=1), maternal and child health (n=3), antimicrobial resistance (n=2), and misinformation (n=6). The literature was dominated by applications of digital surveillance, for example, by using big data through mobility tracking and infodemiology (n=163). Wastewater surveillance was also heavily represented (n=48). Other papers described adaptations to programs or methods that existed prior to the COVID-19 pandemic (n=9). The scoping search also found 109 papers that discuss the ethical, legal, security, and equity implications of emerging surveillance methods. The peer reviewer public health epidemiologists noted that additional changes likely exist, beyond what has been reported and available for evidence syntheses.

Conclusions: The COVID-19 pandemic accelerated advancements in surveillance and the adoption of new technologies, especially for digital and wastewater surveillance methods. Given the investments in these systems, further applications for public health surveillance are likely. The literature for surveillance methods was dominated by surveillance of infectious diseases, particularly COVID-19. A substantial amount of literature on the ethical, legal, security, and equity implications of these emerging surveillance methods also points to a need for cautious consideration of potential harm.

(JMIR Public Health Surveill 2024;10:e49185) doi:10.2196/49185
KEYWORDS
public health; surveillance; digital surveillance; COVID-19; screening; infodemiology; big data; mobility tracking; wastewater; ethics; decision making; public health surveillance

Introduction

Decision-making in public health draws on many sources of evidence. The best available research evidence, local health issues and context, current community and political climate, and evidence for existing public health resources all contribute to an evidence-informed decision-making approach [1].

Population-specific surveillance data provides critical evidence to inform decision makers on contextually relevant health issues. Public health surveillance data have been used to directly inform policies like animal importation laws; food recalls due to contaminated food; and communication to the public, such as promoting awareness of emerging environmental threats such as Lyme disease [2,3].

Public health surveillance plays a vital role in decision-making for maintaining and promoting population health. While surveillance programs are most well-known for tracking infectious disease outbreaks, programs also include screening for noncommunicable diseases. In Canada, public health surveillance occurs at national, provincial or territorial, regional, and municipal levels. At the national level alone, nearly 30 surveillance programs, networks, and systems operate to monitor disease and illness across Canada—such as the National Enteric Surveillance Program which monitors the incidence of foodborne illness, and FluWatch, which reports on the spread of influenza [4,5]. There are also regional and local surveillance processes that collect data, including the Ontario integrated Public Health Information System, which allows for reporting and surveillance of several diseases including enteric illness and influenza [6]. The surveillance data collected through these programs inform public health planning, programming, and response.

The COVID-19 pandemic and subsequent public health response led to a pivot in public health priorities. As the urgency to understand and respond to COVID-19 transmission grew, many public health teams prioritized pandemic response including the redeployment of many staff to contact tracing and eventually mass vaccination. For example, over 500 contact tracers were recruited in Ontario alone over a 2-month period in late 2020 [7], and, health care workers in British Columbia were hired or redistributed to form the COVID-19 Immunize BC Operations Centre to coordinate the province’s vaccine response [8].

Non–COVID-19 disease surveillance programs were impacted by emergency public health measures as well as reallocation of resources. Measures including physical distancing and service shutdowns impacted existing programs that were not considered essential, such as cancer screening programs that were paused or deferred during the height of the pandemic [9,10].

Access to “smart” or networked technology has become ubiquitous both in and outside of public health, with nearly 90% of Canadians owning a smartphone and 75% of Canadians using social media regularly [11-13]. The progression of public health surveillance methods has reflected a similar integration of technology, seeing many changes since the prominent SARS-CoV-1 epidemic in 2003. Newer surveillance tools include electronic health records, advancements in wastewater surveillance infrastructure and technology, and digital health surveillance [14]. Electronic health records are the technological response to navigate the complex Canadian health care system by digitally housing individuals’ full health histories [15]. Digitization of records also allows for centrally tracking infectious diseases that may enter the country or region [14]. Prior to the COVID-19 pandemic, wastewater surveillance had gained popularity for tracking infectious disease dynamics and tracing the origins of foodborne illness outbreaks [16,17]. As it is difficult to estimate the true burden of disease of foodborne illnesses, wastewater surveillance has proven to be an accurate and relatively inexpensive way of collecting health data [17,18]. Similarly, digital surveillance can use internet search data to track where and when disease outbreaks occur [19]. More individually, connection to WiFi and GPS tracking can provide access to millions of smartphone users’ locations, enabling accurate and specific tracking of outbreaks and other public health issues [20,21].

Despite advances that occurred prior to the COVID-19 pandemic, the need for improvement of public health surveillance capacity and methods has persisted in Canada and internationally [22-26]. Existing barriers in public health include low-quality data and obstacles to data sharing, as well as political and professional barriers such as incompatible technical lexicons and territorial borders [26]. This adds to the pertinence in investigating how the COVID-19 pandemic impacted public health surveillance and how it may continue to do so [27]. For example, the singular worldwide focus on COVID-19 allowed for an unprecedented dedication of resources (time, money, and intellect) to leverage pre-existing surveillance processes, creating the opportunity to apply these learnings to non–COVID-19 disease surveillance contexts [28]. The severity of the COVID-19 pandemic necessitated a further push in the development of innovative surveillance methods to track and manage SARS-CoV-2 transmission [29,30]. As a result, there have been marked improvements in the ability to predict, track, and respond to COVID-19 outbreaks in Canada and around the world [31-34].

However, there is a need for ethical consideration to be at the forefront of these novel surveillance applications [18,35]. Although the ubiquity of personal technology like smartphones presents convenient and cost-effective disease surveillance opportunities, concerns around privacy and stigma through potential applications have emerged [36-39]. Digital surveillance tools that capture data through personal devices such as smartphones have the ability to capture personally identifiable information including race-based and health history data [35]. As digital surveillance itself is fairly novel there are few regulatory and legal frameworks in place to protect the misuse of this data [35]. Transparency, in terms of what information is
being collected and how it is being used, is an additional ethical consideration that has yet to be fully addressed [38].

The National Collaborating Centre for Methods and Tools’ Rapid Evidence Service [40] was established at the onset of the COVID-19 pandemic to conduct rapid reviews on priority public health topics. This review was requested of the rapid evidence service by the Public Health Agency of Canada to understand how public health surveillance was changed by the COVID-19 pandemic, resulting in a knowledge synthesis to inform future decision-making. This rapid scoping review investigates the question: What is known about changes to public health methods at the population level for governments globally due to the COVID-19 pandemic?

Methods

Study Design

A scoping review capturing literature on changes to public health surveillance in response to the COVID-19 pandemic was conducted and is reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [41]. Review methods were based on previously described rapid review methods [40], with adaptations to a scoping review framework [42]. Inclusion and exclusion criteria were defined based on population, concept, and context criteria rather than population, intervention or exposure, comparison, and outcome criteria. Findings from included studies were not extracted and studies were not assessed for study quality, since the review question focused on identifying a breadth of evidence, rather than individual study findings. Data were analyzed for overall trends and presented visually.

Information Sources and Search Strategy

The search was conducted on January 25, 2022. The following databases were searched: Medline, Embase, Emcare, Global Health Database, PsycInfo, MedRxiv, COVID-19 living overview of evidence, Web of Science, and the World Health Organization’s global literature on coronavirus disease. Each data were searched using combinations of the terms “monitor,” “surveillance,” and “coronavirus.” Additionally, international government websites were hand-searched for relevant gray literature. The full search strategy is included in Multimedia Appendix 1 [14,28-30,32-34,36,43-274].

Eligibility Criteria

Literature on population-level surveillance was included if surveillance was established in response to the COVID-19 pandemic, or directly impacted by the COVID-19 pandemic, through reprioritization of resources or public health restrictions. Studies and reports published between March 1, 2020 and January 25, 2022, were included. While the first cases of COVID-19 were identified in December 2019, the search was timeframe limited to March 2020 or after to allow for literature to capture changes in surveillance and reduce the overall results set for this rapid scoping review. English- or French-language, peer-reviewed sources, and sources published ahead of print before peer review were included. Gray literature was also included. Detailed criteria for inclusion and rationale are outlined inTextbox 1.
Table 1. Inclusion and exclusion criteria for rapid scoping review of changes to public health surveillance methods during the COVID-19 pandemic March 2020-January 2022.

<table>
<thead>
<tr>
<th>Population:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion criteria</strong></td>
</tr>
<tr>
<td>National, provincial, territorial, state, and regional population-level surveillance.</td>
</tr>
<tr>
<td><strong>Exclusion criteria</strong></td>
</tr>
<tr>
<td>Surveillance of an individual.</td>
</tr>
<tr>
<td><strong>Rationale</strong></td>
</tr>
<tr>
<td>To inform the planning of population-level surveillance programs and systems.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion criteria</strong></td>
</tr>
<tr>
<td>Surveillance programs and systems that have been directly affected by the COVID-19 pandemic.</td>
</tr>
<tr>
<td><strong>Exclusion criteria</strong></td>
</tr>
<tr>
<td>Surveillance programs and systems not directly impacted by the COVID-19 pandemic.</td>
</tr>
<tr>
<td><strong>Rationale</strong></td>
</tr>
<tr>
<td>To explore shifts in surveillance approaches in response to the COVID-19 pandemic.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion criteria</strong></td>
</tr>
<tr>
<td><strong>Rationale</strong></td>
</tr>
<tr>
<td>To capture changes in surveillance methods due to the COVID-19 pandemic and reduce overall scope.</td>
</tr>
</tbody>
</table>

Studies were screened using DistillerSR software (DistillerSR). Titles and abstracts of retrieved studies were screened by a single reviewer (SN). Full texts of included studies were screened by a second reviewer (ECC) and reviewed by a third (MD). The screening was not completed in duplicate to maintain a rapid timeline.

Data Extraction

Data extraction was completed by a single reviewer (ECC or SH). Details on the surveillance method (eg, wastewater sampling and digital data collection), as well as surveillance subjects (eg, COVID-19 and population mental health) for each included source were extracted. The type of evidence source was also extracted (eg, syntheses, single study, and opinion).

Studies, critical reviews, editorials, and newspaper articles commenting on the ethics of surveillance methods that were captured by the search were briefly summarized. Details of the method of surveillance and the focus of the ethical implications (ie, ethics, security, legality, and equity) were extracted. Consistent with a scoping review approach, quality assessments of included studies were not completed.

Data Analysis

Sources were grouped by surveillance method and surveillance subject to identify overall trends in the literature. Papers presenting ethical implications of novel approaches to surveillance were grouped by the paper’s focus.

Findings are presented as overall trends in the literature [42]. Results are discussed narratively and summarized in tables. To visually depict overall findings, a Sankey diagram was generated pairing surveillance methods and surveillance subjects, using the software SankeyMATIC (SankeyMATIC). Sankey diagrams can represent the flow of material or ideas through a system. The scale of each section of a Sankey diagram is proportional to the amount of material flowing between stages [43,275-277].

Peer Reviewers

Involving patient and public partners in rapid reviews has emerged as a valuable addition to review methodology. Gaining perspectives from patients and public partners can help ensure that rapid reviews meet knowledge users’ needs and are relevant for decision-making [278,279]. The Cochrane collaboration, a world leader in systematic review methodology, has identified consumer involvement throughout the process of evidence production and dissemination as a vital component [280,281]. Given the more technical focus of this review, rather than seeking partners with lived experience from the general public, epidemiologists who worked closely with surveillance efforts during the COVID-19 pandemic were invited as peer reviewers through organizational connections. Partners were invited to provide feedback on review findings based on lived experiences working in the field. Specifically, they were asked to review the Executive Summary and reflect on their own experiences with shifts in public health surveillance during the COVID-19 pandemic, and to highlight gaps in the literature. Their perspectives are not presented as a source of objective data.
alongside review findings, but rather as contextualizing perspectives of the objective data found by the review. Both partners provided written feedback that informed the review’s conclusions.

**Results**

**Overview**

Database searching retrieved 36,791 records and scanning of key government websites retrieved 8 records. After removing duplicates, 14,238 records were screened by title and abstract, resulting in 1980 reports for full-text review. Of those 1980 records, 241 papers that described surveillance methods and 108 papers that discussed the implications of changes to public health surveillance programs from ethical, security, legal, and equity perspectives were included in the scoping review and underwent data extraction. See Figure 1 for a PRISMA-ScR flowchart illustrating the paper search and selection process.

**Study Characteristics**

Of the 241 papers that described public health surveillance methods, 80 were review papers or evidence syntheses and 161 were single studies. Literature was dominated by reports of COVID-19 surveillance (n=187), but other infectious diseases were also represented, including other respiratory and viral pathogens (n=3), and gastrointestinal (n=3), sexually transmitted (n=2) and zoonotic infections (n=4). There were several papers reporting on surveillance related to immunization, including vaccine efficacy (n=1), vaccine safety monitoring (n=4), and vaccine hesitancy (n=4). Some papers reported on cancer screening (n=5) during the COVID-19 pandemic. Surveillance of mental health (n=11) and substance use (n=4) were also described. One paper described surveillance of healthy nutrition (n=1) and other papers described surveillance of maternal and child health (n=3). Papers described monitoring of antimicrobial resistance (n=2). Several papers described monitoring misinformation on the web (n=6). Table 1 summarizes the number of syntheses and single studies by public health surveillance subject, organized by public health topic area. A detailed list of included studies is available in Table S1 in Multimedia Appendix 1 [14,28-30,32-34,36,43-274].
Table 1. Numbers of syntheses and single studies of public health surveillance method changes among public health surveillance subjects from March 2020 through January 2022.

<table>
<thead>
<tr>
<th>Public health topic area and surveillance subject</th>
<th>Syntheses, n</th>
<th>Single studies, n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antimicrobial stewardship</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antimicrobial resistance</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Chronic diseases and conditions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer screening</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td><strong>Health promotion</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nutrition</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Maternal and child health</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td><strong>Immunization</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaccine efficacy</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Vaccine safety</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Vaccine hesitancy</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td><strong>Infectious diseases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19</td>
<td>67</td>
<td>120</td>
</tr>
<tr>
<td>Other respiratory pathogens (general)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Other viral pathogens (general)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Vaccine-preventable diseases</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Gastrointestinal infections</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Sexually transmitted infections</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Zoonoses</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>Mental health and substance use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental health</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Substance use</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><strong>Health communication</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misinformation</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

Singles studies were conducted in the United States (n=51), Europe (n=27), East Asia (n=17), United Kingdom (n=9), Canada (n=5), South America (n=4), Africa (n=3), Australia (n=3), South Asia (n=2), and Russia (n=1). A total of 12 studies were conducted across several countries, and 27 studies were conducted globally. Global studies were mostly digital surveillance studies using global data sets.

Of the 80 included evidence syntheses, the majority were nonsystematic literature reviews (n=61). Syntheses also included systematic reviews (n=10), scoping reviews (n=7), 1 rapid review, and 1 guideline.

**Surveillance Methods**

The Sankey diagram in Figure 2 illustrates the number of included papers that report on the various surveillance methods used for the target surveillance subjects, listed by public health topic area.
Most of the papers on surveillance of COVID-19 transmission described digital surveillance (eg, 130 papers on the use of big data through mobility tracking and infodemiology and 40 papers on wastewater epidemiology). Other methods for monitoring COVID-19 transmission were designed to supplement limited COVID-19 testing capacity, including analysis of data from various health care settings, for example, hospitals and blood banks (n=11). Strategies to track COVID-19 transmission and detect new SARS-CoV-2 variants through genome sequencing were also described (n=5).

Literature for select public health surveillance functions that operated prior to the COVID-19 pandemic, such as cancer screening, child health assessments, and sexually transmitted infection screening, indicated that these functions adapted to using technology for remote screening in response to mobility restrictions and physical distancing. For example, 2 studies described videoconferencing for remote autism screening in young children, and another study described a pilot program for remote sexually transmitted infection self-testing under the supervision of public health staff through videoconference.

Other public health surveillance programs took approaches similar to COVID-19 surveillance. For example, digital surveillance technologies were applied to monitor population mental health (n=8), misinformation (n=5), and zoonoses (n=2). In addition to COVID-19, wastewater surveillance was used to monitor antibiotic resistance (n=2), mental health (n=2), substance use (n=1), as well as other infectious diseases.

Artificial intelligence (AI) was used to augment digital surveillance of COVID-19 transmission (n=29), immunization (n=2), substance use (n=2), and misinformation (n=3; Table 2). Most of these used AI to comb through population-level cell phone mobility data, social media content, or search engine queries for indications of COVID-19 transmission. AI was also used to analyze social media content and search engine queries for indications of vaccine hesitancy, substance use, and misinformation. One paper described using AI to analyze global wastewater data for COVID-19 transmission pathways.
Several studies (n=4) described the development of standardized, digitized data repositories to enable interjurisdictional surveillance [44-47]. One of these described standardized electronic reporting for cancer screening to support central cancer registries [44]. A literature review described distributed data networks to monitor COVID-19 vaccine safety issues [45]. Two papers described data federation technologies for infectious diseases other than COVID-19. A literature review by Baker et al [46] described the standardization of data for malaria, and a pilot study by Colella et al [47] described the development of a biorepository platform to monitor animal-to-human disease transmission. These papers emphasized the interoperability of data systems to facilitate large-scale surveillance of public health issues.

One study used Indigenous methods to monitor COVID-19 through a community-focused COVID-19 surveillance program [48]. In this study, the Hopi Tribe, a sovereign nation in the United States, collaborated with the US Centers for Disease Control and Prevention to deploy community health representatives to every household in 2 Hopi villages [48]. Community health representatives screened household members for COVID-19 symptoms, exposures, and provided education, all in a culturally safe manner [48]. No other studies using Indigenous methods were found.

### Ethical Implications

In addition to papers describing public health surveillance methods, 108 papers exploring the implications of changes to public health surveillance programs during the COVID-19 pandemic, from ethical, legal, data security, and equity perspectives were retrieved. These papers focused on digital surveillance (n=104) or wastewater surveillance (n=4) and discussed potential consequences of surveillance in terms of ethics (n=80), data security (n=28), legality (n=28), and equity (n=21), noting that some papers covered more than 1 topic area. A detailed list of included studies is available in Table S2 in Multimedia Appendix 2 [37-39,282-386].

Aside from 2 systematic reviews, most of the papers discussing the implications of surveillance were critical literature reviews, conducted nonsystematically and written to express a clear opinion or bias (n=71). There were 36 editorials that did not cite or infrequently cited other literature and were published in academic journals (n=29) or newspapers (n=7).

### Perspectives From Public Health Epidemiologists

Two public health epidemiologists who worked in Ontario during the COVID-19 pandemic provided feedback on the findings of this review. They suggested that nearly every surveillance system or program was impacted by the COVID-19 pandemic; some paused due to resource reallocation. Several key metrics that were previously reliable indicators became difficult or not possible to interpret due to the impact of COVID-19. These changes may not be captured in academic literature featured in this review, as many public health professionals working during the pandemic simply did not have the time, energy, or academic support to publish their work.

Epidemiologists shared that the overall approach to surveillance of COVID-19 also changed over the course of the pandemic. While case counts were an early key indicator, there has been a general shift to measures such as hospital and intensive care unit occupancy, reflecting the change in priority from eliminating COVID-19 altogether to reducing and managing the health care system burden.

It was also suggested by epidemiologists that while data for equity has been limited, some governments have made efforts to improve measures of equity. For example, in 2020 the Ontario government amended the Health Protection and Promotion Act and augmented the provincial case and contact management tool to allow for the collection of social determinants (eg, race and income) for COVID-19 cases. Similar changes were made for COVID-19 vaccine surveillance data.

It was suggested that substantial resources are required to restore public health surveillance systems during pandemic recovery and improvements can be made by leveraging or investing in better data.
Discussion

Principal Findings

This scoping review provides a comprehensive overview of changes to public health surveillance methods for programs, systems, and strategies at the population level implemented since early 2020. The review includes 241 papers, including 80 evidence syntheses and 161 single studies describing surveillance methods, as well as 108 papers discussing the implications of changes to public health surveillance programs from ethical, security, legal, and equity perspectives. A Sankey diagram (Figure 2) illustrates the dominance of COVID-19 surveillance in this literature, particularly through digital surveillance strategies.

The heavy focus on COVID-19 in the surveillance literature reflects public health’s drastic reallocation of resources to the pandemic response [387,388]. As nearly all countries of the world entered some form of public health lockdown, unique research opportunities presented themselves [389,390]. Health researchers redirected efforts to contribute to global focus to understand and mitigate COVID-19 transmission and harm [391]. Additionally, academic journals expedited the publication process for papers on COVID-19 [392,393]. While other public health subjects, such as maternal and child health, chronic diseases, and infectious diseases other than COVID-19 were also represented in the surveillance literature, these subjects were largely overshadowed by COVID-19. As suggested by the public health epidemiologists included in this review, those working in surveillance for other public health topics likely did not have time or resources to publish during the pandemic.

Within the COVID-19 surveillance literature, a substantial proportion of papers discussed digital forms of surveillance. This was largely using sources of big data, such as social media posts, search engine queries, and mobility tracking. Prior to the COVID-19 pandemic, there had been a steady rise in publications on big data, as such the field was sufficiently established to contribute to COVID-19 tracking and analysis at the onset of the pandemic [394,395]. Publications were largely proof-of-concept, demonstrating if and how big data analysis correlated with levels of COVID-19 infections seen in communities. Following the interest and advancements in the analysis of big data, these methods are likely to continue to see novel applications in public health.

Wastewater surveillance emerged as a dominant method in the COVID-19 surveillance literature. Wastewater testing is an attractive option for disease surveillance as it is a cost-effective method to monitor populations without requiring active participation or invasive testing procedures [396-398]. Similar to the literature for big data analysis, much of the literature on wastewater provided proof of concept by correlating wastewater detection of the SARS-CoV-2 virus with reported levels of COVID-19 in the community. The impact of wastewater surveillance on public health decision-making is likely to grow [399]. Recently, the detection of vaccine-derived poliovirus in wastewater in New York City led to the addition of the United States to the World Health Organization’s list of countries with polio transmission [400]. To expand coverage across Canada’s population, the Public Health Agency of Canada invested in a pan-Canadian wastewater surveillance network for COVID-19, in collaboration with other federal departments; provincial, territorial, and municipal governments; and academic institutions [401]. While the focus for wastewater surveillance was largely on COVID-19, advances in wastewater monitoring capacity allowed for progress in wastewater monitoring of other conditions and issues, such as mental health and substance use. Given the investment in wastewater epidemiology and advancements in its applications, it is likely to emerge as a key indicator for public health post the pandemic.

In the context of COVID-19 pandemic lockdowns and physical distancing, it is not surprising that some public health services, such as childhood nutrition, cancer screening, and sexual health, were adapted to remote provision of services. While several papers describing the implementation of remote services were included in this review, it is likely that these papers represent only a small sample of public health efforts to continue service provision during the pandemic. As indicated by the public health epidemiologists consulted for this review, many changes to existing public health surveillance programs are likely not published in the academic literature, due to the demands on public health professionals’ time.

The COVID-19 pandemic also highlighted the importance of interjurisdictional data systems which can enable timely access to data across jurisdictional boundaries to inform public health functions. Several papers included in this review explored standardization of data to enhance the public health response, for monitoring zoonoses [46,47] or rare vaccine-related harms [45] at a population level. Solutions to the current limitations of data interoperability must be implemented in future public health surveillance efforts to improve access to essential data in real-time that can inform decision-making within critical health organizations and governing bodies.

While technology has enabled many advancements in public health surveillance, the findings of this scoping review also demonstrate the alarm caused by these changes. The ethical implications of digital surveillance, in particular, were discussed in 80 included papers. Since smartphones and the internet of things (a system of interrelated computing devices to connect people or things identified by unique identifiers), have only become widely used relatively recently, with the potential consequences of their use in public health surveillance not being fully realized yet. The COVID-19 pandemic accelerated the development and testing of digital surveillance, and the literature reflects that experts in these fields have cause for concern. The relative lack of literature discussing the impact of changes in public health surveillance on equity is also concerning. The pandemic exacerbated existing inequities, and care must be taken so that these emerging methods do not further widen equity gaps [402-406]. Data collection and analysis for equity-deserving populations must be a priority as the identification of inequities allows for mitigation. However, these data must be collected and handled with care and respect to avoid potential harm and stigmatization [407,408].
Limitations

Two public health epidemiologists who contributed their perspectives to this review raised that many changes to surveillance are not captured in the literature, primarily due to the lack of time and academic support for public health in publishing. This is an important limitation of this review, as the full scope of changes to public health surveillance is likely not represented. Academic partnerships and support for public health can help share innovations [409].

The peer perspectives shared in this review are also limited to 2 epidemiologists based in Canada, which may not fully represent the perspectives of those involved in public health surveillance worldwide. However, the epidemiologists involved provided contextual insights to the findings of this review.

This review is also limited to papers from March 2020 to January 2022. Studies published after January 2022 may be more likely to capture changes to public health surveillance that are not directly related to COVID-19. For example, examples of wastewater surveillance of pathogens other than SARS-CoV-2 may not have been captured by this search.

As a scoping review of the literature for changes to public health surveillance methods during the COVID-19 pandemic, there are limitations to the overall conclusions that can be supported by this review’s findings. Given the scoping review approach, findings for the effectiveness of surveillance methods and the quality of included papers were not evaluated. A systematic analysis of included papers, including the extraction of outcome data, rigorous quality assessment of included papers, and judgments of the certainty of evidence, is strongly recommended when considering the potential future implementation of any of the surveillance methods discussed in this report.

Conclusions

A considerable body of literature emerged related to advances in surveillance during the COVID-19 pandemic. Some areas have seen much more literature, such as wastewater epidemiology, while others remain sparse, such as health equity for racialized populations. Possible next steps include conducting systematic reviews in areas for which multiple studies have been identified. In other areas with sparse literature, more research is needed to advance surveillance in the postpandemic era. A large volume of papers expressing concerns about the potential ethical, legal, security, and equity implications of emerging surveillance methods indicates a need for a thorough assessment of the potential effects of large-scale implementation of these surveillance methods.

Acknowledgments

The authors would like to acknowledge epidemiologists Jasmine Song and Mackenzie Slifierz for sharing their perspectives on public health surveillance as peer partners in this review. The National Collaborating Centre for Methods and Tools is hosted by McMaster University and funded by the Public Health Agency of Canada. The views expressed herein do not necessarily represent the views of the Public Health Agency of Canada. The funder had no role in the design of the study, collection, analysis, or interpretation of data or in writing the manuscript.

Data Availability

The data sets used and analyzed in the conduct of this review are available from the corresponding authors on reasonable request.

Authors’ Contributions

ECC, LH, and MD designed the study. SN, SH, AK, and ECC completed screening, and data extraction. ECC, SH, and MD analyzed the study results. AK prepared the Sankey diagram. ECC, SN, and LH wrote the manuscript in consultation with MD. SN, SH, AK, LH, and MD provided feedback on all drafts and the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Included articles that describe public health surveillance methods.
[DOCX File, 251 KB - publichealth_v10i1e49185_app1.docx ]

Multimedia Appendix 2

Included articles that describe implications of emerging public health surveillance methods.
[DOCX File, 72 KB - publichealth_v10i1e49185_app2.docx ]

Multimedia Appendix 3

PRISMA-ScR Checklist.
[PDF File (Adobe PDF File), 243 KB - publichealth_v10i1e49185_app3.pdf ]
References


2. Importing terrestrial animals to Canada. Canadian Food Inspection Agency. 2022. URL: http://tinyurl.com/4yzs8bec [accessed 2023-12-23]


6. iPHIS resources. Ontario PH. 2022. URL: http://tinyurl.com/vv3et4pz [accessed 2023-12-23]


12. The evolution of social media: how did it begin, and where could it go next? Maryville University. 2022. URL: https://online.maryville.edu/blog/evolution-social-media/ [accessed 2023-12-23]


16. Brinkman NE, Fout GS, Keely SP. Retrospective surveillance of wastewater to examine seasonal dynamics of enterovirus infections. mSphere 2017;2(3):e00099-e10017 [FREE Full text] [doi: 10.1128/mSphere.00099-17] [Medline: 28630939]


89. Kilby AE, Denhart C. Location inference on social media data for agile monitoring of public health crises: an application to opioid use and abuse during the Covid-19 pandemic. ArXiv Preprint posted online on 02 Nov 2021 [FREE Full text]


Andrei S. Coronavirus will pass, but Moscow's mass surveillance is here to stay. The Moscow Times. 2020. URL: http://tinyurl.com/4eyyiuh5 [accessed 2023-12-23]


Bhardwaj K. Digital surveillance systems to combat COVID-19 may do more harm than good. Econ Pol Wkly 2020;55(23):1-10 [FREE Full text]


Cofone I. Immunity passports and contact tracing surveillance. Stan Tech L Rev 2021;24:176 [FREE Full text]


Cox D. Alarm bells ring for patient data and privacy in the COVID-19 goldrush. BMJ 2020;369:m1925 [FREE Full text] [doi: 10.1136/bmj.m1925] [Medline: 32423900]


Goldenfein J, Green B, Viljoen S. Privacy versus health is a false trade-off. SSRN J. Preprint posted online on May 26, 2020 [FREE Full text]


306. Greenleaf G, Kemp K. Police access to COVID check-in data is an affront to our privacy. we need stronger and more consistent rules in place. SSRN J. Preprint posted online on September 7, 2021 [FREE Full text] [doi: 10.2139/ssrn.3922035]


325. Oliva JD. Surveillance, privacy, and app tracking. SSRN J. Preprint posted online on July 31, 2020 [FREE Full text]


https://publichealth.jmir.org/2024/1/e49185 JMIR Public Health Surveill 2024 | vol. 10 | e49185 | p.276 (page number not for citation purposes)


368. Rozenstein AZ. Digital disease surveillance. SSRN J. Preprint posted online on October 9, 2020 [FREE Full text] [doi: 10.2139/ssrn.3698431]


374. Mahapatra S. Digital surveillance and the threat to civil liberties in India. GIGA Focus Asia 2021(3):1-12 [FREE Full text] [doi: 10.2139/ssrn.3658228]


Abbreviations

AI: artificial intelligence
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

©Emily C Clark, Sophie Neumann, Stephanie Hopkins, Alyssa Kostopoulos, Leah Hagerman, Maureen Dobbins. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 19.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Counseling Supporting HIV Self-Testing and Linkage to Care Among Men Who Have Sex With Men: Systematic Review and Meta-Analysis

Siyu Chen1, MPH; Yuan Fang2, PhD; Paul Shing-fong Chan1, PhD; Joseph Kawuki1, PhD; Phoenix Mo1, PhD; Zixin Wang2, PhD

1Centre for Health Behaviours Research, Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
2Department of Health and Physical Education, The Education University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Zixin Wang, PhD
Centre for Health Behaviours Research
Jockey Club School of Public Health and Primary Care
The Chinese University of Hong Kong
Room 508, School of Public Health
Prince of Wales Hospital, Shatin, NT
Hong Kong, 999077
China (Hong Kong)
Phone: 852 2252 8740
Email: wangzx@cuhk.edu.hk

Abstract

Background: Counseling supporting HIV self-testing (HIVST) is helpful in facilitating linkage to care and promoting behavior changes among men who have sex with men (MSM). Different levels of counseling support for MSM HIVST users may lead to variance in the linkage to care.

Objective: This study aims to synthesize evidence on counseling supporting MSM HIVST users and to conduct a meta-analysis to quantify the proportion of MSM HIVST users who were linked to care.

Methods: A systematic search was conducted using predefined eligibility criteria and relevant keywords to retrieve studies from the MEDLINE, Global Health, Web of Science, Embase, APA PsycINFO, and Scopus databases. This search encompassed papers and preprints published between July 3, 2012, and June 30, 2022. Studies were eligible if they reported counseling supporting HIVST or quantitative outcomes for linkage to care among MSM and were published in English. The screening process and data extraction followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The quality of the included studies was assessed by the National Institutes of Health quality assessment tool. Data were extracted using random effects models to combine the proportion of HIVST users who were linked to care. Subgroup analyses and metaregression were conducted to assess whether linkage to care varied according to study characteristics. All analyses were performed with R (version 4.2.1; R Foundation for Statistical Computing) using the metafor package.

Results: A total of 55 studies published between 2014 and 2021, including 43 observational studies and 12 randomized controlled trials, were identified. Among these studies, 50 (91%) provided active counseling support and 5 (9%) provided passive counseling support. In studies providing active counseling support, most MSM HIVST users were linked to various forms of care, including reporting test results (97.2%, 95% CI 74.3%-99.8%), laboratory confirmation (92.6%, 95% CI 86.1%-96.2%), antiretroviral therapy initiation (90.8%, 95% CI 86.7%-93.7%), and referral to physicians (96.3%, 95% CI 85%-99.2%). In studies providing passive counseling support, fewer MSM HIVST users were linked to laboratory confirmation (78.7%, 95% CI 17.8%-98.4%), antiretroviral therapy initiation (79.1%, 95% CI 48.8%-93.7%), and referral to physicians (79.1%, 95% CI 0%-100%). Multivariate metaregression indicated that a higher number of essential counseling components, a smaller sample size (<300), and the use of mobile health technology to deliver counseling support were associated with better linkage to care. The quality of the studies varied from fair to good with a low to high risk of bias.

Conclusions: Proactively providing counseling support for all users, involving a higher number of essential components in the counseling support, and using mobile health technology could increase the linkage to care among MSM HIVST users.
Introduction

Background

Globally, the estimated median HIV prevalence among men who have sex with men (MSM) ranges from 5% in Southeast Asia to 12.6% in Eastern and Southern Africa [1]. The risk of acquiring HIV is 26 times higher among MSM compared with the general population worldwide [1]. In 2022, the proportion of new HIV cases attributed to MSM was 44% in Asia and the Pacific [2], 38.7% in Europe, and 70% in the United States [3-5].

HIV testing is one of the key strategies for controlling the spread of HIV [6,7]. Both the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the Centers for Disease Control and Prevention recommend MSM to undergo HIV testing every 3 to 6 months [6,7]. The UNAIDS established its 95-95-95 targets in 2014 [6]. The aim was to diagnose 95% of all individuals testing positive for HIV, provide antiretroviral therapy (ART) to 95% of those diagnosed, and achieve viral suppression for those treated by 2030 [6]. High coverage of HIV testing was the first step in achieving the 95-95-95 targets. However, the overall HIV testing coverage among MSM was 86.2% in Africa and 89% in North America [8]. There is a need for further improvement. HIV self-testing (HIVST) could serve as an alternative strategy for enhancing HIV testing coverage. Systematic reviews have demonstrated that HIVST can overcome barriers faced by MSM when accessing HIV testing services, such as perceived stigma from providers and inconvenience [9-11]. Previous meta-analyses consistently showed that HIVST has doubled the frequency of HIV testing compared with facility-based testing [12,13]. As a result, the World Health Organization (WHO) recommends offering HIVST as an additional approach in addition to the existing HIV testing services [14].

On the basis of the presence of counseling support, HIVST can be categorized into assisted and unassisted HIVST. Several studies investigated the linkage to care in assisted and unassisted HIVST. Individuals who received positive results through unassisted HIVST faced more difficulties in accessing care than those who were identified by facility-based HIV testing and counseling [15]. According to a systematic review, <25% of unassisted HIVST users were able to complete the test without any errors, and many of them had difficulties interpreting the HIVST results [16]. A meta-analysis showed that the absence of assistance would lead to a 17% decrease in the linkage to care rate among HIVST users [12]. Across countries, studies have consistently shown that implementing assisted HIVST could increase linkage to care among different populations [17-19]. A very high linkage to care (99%-100%) was observed among users of assisted HIVST in the United States and Zimbabwe [19]. Therefore, the WHO recommends that counseling support be provided to HIVST users [20].

As recommended by the WHO, 8 essential components should be included in the pretest and posttest counseling of a standard-of-care client-initiated HIV testing and counseling. The pretest counseling should include (1) assessing the risks and window periods, (2) informing clients of the benefit of taking the test and the implications of both negative and positive results, (3) assuring the clients’ right to refuse to take the test, (4) encouraging the clients to anticipate the possibility of beneficial disclosure of serostatus status, and (5) providing preventive information and materials [21]. Essential components of posttest counseling include (1) interpreting testing results; (2) offering psychological support to individuals testing positive for HIV, facilitating beneficial disclosure of their positive serostatus, and referring them for further care and support services; and (3) providing HIV-negative individuals with preventive information and materials [21]. However, the level of counseling support varied across previous HIVST programs. Some programs proactively provided pretest or posttest counseling support to all HIVST users unless they refused [22]. This mode of counseling support was categorized as active counseling [23-25]. Providing active counseling increased ART adherence among people living with HIV who had an unsuppressed viral load [26]. Other programs did not provide active counseling to users. Users could report their results via a web-based platform and obtain optional posttest counseling [27]. This mode of counseling support was categorized as passive counseling [24,25]. In addition, the number of essential components involved in counseling support varied significantly. Some studies only involved a single essential component (eg, providing additional HIV care for HIVST users who received reactive results), whereas others provided more comprehensive support (eg, assessing the risks and window periods, delivering preventive information and materials, and providing additional HIV care for HIVST users who received reactive results) [26,27].

Objectives

Existing systematic reviews and meta-analyses have investigated the digital support [28], effectiveness [29], and acceptability of HIVST [30]. However, few studies have summarized the levels of counseling support for HIVST among MSM. It is also unclear whether different levels of counseling support would result in differences in the linkage to care among MSM HIVST users. To address this knowledge gap, we systematically reviewed global evidence on counseling support for MSM HIVST users. We also summarized the linkage to care under different modes of counseling support, including (1) the proportion of users who reported HIVST results; (2) the proportion of users with reactive results who were linked to laboratory confirmation, ART initiation, and physicians; and (3) the proportion of users with...
negative results who were given information related to sexual risk behavior reduction, related to pre-exposure prophylaxis (PrEP), and linked to PrEP initiation.

Methods

This systematic review and meta-analysis was registered with PROSPERO (CRD42022346247) and conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [31].

Search Strategy

We searched the MEDLINE, Global Health, Web of Science, Embase, APA PsycINFO, and Scopus databases for studies (including both published papers and preprints) between July 3, 2012 (the date when HIVST was approved by the Federal Drug Administration), and June 30, 2022, in any country or setting [32]. Keywords were selected based on the PICOS (participants, intervention, comparison, outcome, and study) criteria to address the research question (where P=MSM, I=HIVST with counseling, C=none, O=linkage to care, and S=randomized controlled trial [RCT] or observational studies). The Boolean operator was used in the search strategy, with “OR” and “AND” used to link search terms, whereas the asterisk “*” was used as a wildcard symbol appended at the end of the terms to search for variations of those terms. Full search strategies are available in Multimedia Appendix 2.

Additional studies were identified through the UNAIDS and WHO websites. We also reviewed databases listing ongoing RCTs, such as ClinicalTrials.gov, the WHO International Clinical Trials Registry Platform, and the Pan African Clinical Trials Registry, as well as reference lists of published reviews, meta-analyses, and articles.

Inclusion and Exclusion Criteria

Inclusion and exclusion criteria are presented in Table 1. The exposure categories and outcomes of the study are presented in Textbox 1.

Table 1. Summary of the inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article or study type</td>
<td>• Population-based original research studies</td>
<td>• Reviews, narratives, commentaries, and editorials</td>
</tr>
<tr>
<td></td>
<td>• Quantitative studies</td>
<td>• Qualitative studies</td>
</tr>
<tr>
<td></td>
<td>• Multicountry studies</td>
<td>• Dissertations, government reports, newspaper articles, textbooks, book chapters, and protocols</td>
</tr>
<tr>
<td></td>
<td>• Gray literature and preprints</td>
<td>• Laboratory studies, model and framework studies, and validation studies</td>
</tr>
<tr>
<td>Language</td>
<td>• English language</td>
<td>• All other non-English languages</td>
</tr>
<tr>
<td>Publication period</td>
<td>• July 3, 2012, to June 30, 2022</td>
<td>• All periods outside July 3, 2012, to June 30, 2022</td>
</tr>
<tr>
<td>Study setting</td>
<td>• All countries</td>
<td>• None</td>
</tr>
</tbody>
</table>

Textbox 1. Exposure categories and outcomes included in the review of linkage to care following HIV self-testing (HIVST), along with counseling.

**Exposure categories**

1. Studies that included HIVST along with active counseling support (eg, studies proactively provided pretest or posttest counseling to all HIVST users unless refused)
2. Studies that included HIVST along with passive counseling support (eg, studies provided certain level of counseling support to HIVST users upon request)

**Outcomes**

1. Reporting test results (defined as the proportion of men who have sex with men [MSM] who reported test results)
2. Laboratory confirmation (defined as the proportion of receiving confirmatory test among MSM with reactive HIVST results)
3. Antiretroviral therapy (ART) initiation (defined as the proportion of ART initiation among MSM HIVST users who are confirmed to be HIV positive)
4. Referred to physicians (defined as the proportion of seeking physicians among MSM HIVST users who are confirmed to be HIV positive)
5. Prevention strategies (defined as the proportion of MSM with negative HIVST results who received information related to sexual risk behaviors reduction, pre-exposure prophylaxis [PrEP])
6. PrEP initiation (defined as the proportion of starting PrEP among MSM with negative HIVST results who had the risk of HIV infection)

Data Extraction

Critical information from this study was extracted using a data extraction form, as outlined in Table 2. The study outcome was the proportion of MSM HIVST users who were linked to care. Two independent reviewers (SC and YF) assessed the eligibility, evaluated the quality, and extracted information from the
included publications. Any disagreements during the data extraction and quality assessment process were resolved by a senior reviewer (ZW).
Table 2. Characteristics, active counseling, and passive counseling support of included studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Study setting or country</th>
<th>Study design</th>
<th>Sample size</th>
<th>Age (y)</th>
<th>Counseling delivering modes</th>
<th>Active counseling support</th>
<th>Passive counseling support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marlin et al [33], 2014</td>
<td>United States</td>
<td>Cross-sectional study</td>
<td>641</td>
<td>90% of the participants were aged between 18 and 35 years</td>
<td>Technology or mobile health</td>
<td>They were provided linkage-to-care consultation via telephone survey</td>
<td>N/A</td>
</tr>
<tr>
<td>Tao et al [34], 2014</td>
<td>China</td>
<td>Cross-sectional study</td>
<td>220</td>
<td>Median 28 (IQR 22-29)</td>
<td>Technology or mobile health</td>
<td>They were provided pretest and posttest counseling via a telephone hotline or QQ Group and instructions by a web-based video posted to the website.</td>
<td>N/A</td>
</tr>
<tr>
<td>Sabharwal et al [35], 2015</td>
<td>United States</td>
<td>Cross-sectional study</td>
<td>53</td>
<td>Mean 32</td>
<td>Peer or community</td>
<td>They were provided partner services and linkage to care to all persons newly diagnosed with HIV by New York City health department.</td>
<td>N/A</td>
</tr>
<tr>
<td>Huang et al [36], 2016</td>
<td>United States</td>
<td>Cross-sectional study</td>
<td>334</td>
<td>65% of the participants were aged between 18 and 30 (range 18 to &gt;41)</td>
<td>Technology or mobile health</td>
<td>They were provided test result instructions, linkage-to-care activities via email or SMS text message, and 69% of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Rosengren et al [37], 2016</td>
<td>United States</td>
<td>Cross-sectional study</td>
<td>125</td>
<td>93% of the participants were aged between 18 and 40 (range 18 to &gt;41)</td>
<td>Technology or mobile health</td>
<td>They were provided reminder emails to complete a web-based survey to report testing results and posttest counseling, and 77% of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Volk et al [38], 2016</td>
<td>Brazil</td>
<td>Cohort study</td>
<td>103</td>
<td>51% of the participants were aged between 18 and 25 (range 18 to &gt;25)</td>
<td>Technology or mobile health</td>
<td>They were provided written instructions that included pictures, pretest and posttest counseling materials, a list of local HIV/AIDS resources, and condoms by mobile phones or email access, and 98% of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Jamil et al [39], 2017</td>
<td>Australia</td>
<td>RCTb</td>
<td>178</td>
<td>Mean 35.8 (SD 11.1)</td>
<td>Technology or mobile health</td>
<td>They were provided expedited confirmatory testing, clinical review, and supportive counseling to any participant with a reactive self-test result at the study clinics, and 90% of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Qin et al [40], 2017</td>
<td>China</td>
<td>Cross-sectional study</td>
<td>341</td>
<td>Mean 24.4 (SD 6.3)</td>
<td>Peer or community</td>
<td>They were provided confirmation of self-test results (at a CDC or hospital), posttest counseling, and potential harms (coercion, feelings of suicidality, and violence), and 47% of participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Zhong et al [41], 2017</td>
<td>China</td>
<td>Cross-sectional study</td>
<td>198</td>
<td>54.2% of the participants were aged between 25 and 34 (range 18 to &gt;35)</td>
<td>Technology or mobile health</td>
<td>They were provided the results interpretation, counseling services, confirmation testing, and linkage to care.</td>
<td>N/A</td>
</tr>
<tr>
<td>Study</td>
<td>Study setting or country</td>
<td>Study design</td>
<td>Sample size</td>
<td>Age (y)</td>
<td>Counseling delivering modes</td>
<td>Active counseling support</td>
<td>Passive counseling support</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>----------------</td>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Choko et al [42], 2018</td>
<td>Uganda</td>
<td>Cohort study</td>
<td>95</td>
<td>Median 41 (IQR 23-62)</td>
<td>Peer or community</td>
<td>They were provided results interpretation and the opportunity to undergo confirmatory HIV testing, and 99% of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Green et al [43], 2018</td>
<td>Vietnam</td>
<td>Cross-sectional study</td>
<td>803</td>
<td>Most of the participants were aged ≤30 years</td>
<td>Peer or community</td>
<td>They were provided confirmatory testing at the closest district health center, and those that were HIV diagnosed were helped with treatment enrollment by peer and staff.</td>
<td>N/A</td>
</tr>
<tr>
<td>Katz et al [44], 2018</td>
<td>United States</td>
<td>RCT</td>
<td>230</td>
<td>Median 35.5 (IQR 27-45.5)</td>
<td>Peer or community</td>
<td>They were provided instructions, pretest and posttest counseling materials, a list of local HIV-related resources, and condoms.</td>
<td>N/A</td>
</tr>
<tr>
<td>Lippman et al [45], 2018</td>
<td>South Africa</td>
<td>Cohort study</td>
<td>127</td>
<td>65% of the participants were aged between 18 and 24 years</td>
<td>Technology or mobile health</td>
<td>They were provided logs to document the use of the tests, a list of local psychosocial and medical resources and referrals should the participant test HIV positive—including a 24-hour study phone number—and safer sex supplies (ie, condoms and lubricant), and 97% of participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Pant Pai et al [46], 2018</td>
<td>Canada</td>
<td>Cross-sectional study</td>
<td>451</td>
<td>Mean 34 (range 18-73)</td>
<td>Technology or mobile health</td>
<td>They were provided instructions on pretest counseling, staging, conducting results, and storing their results on from HIVSmart app, and 99% of participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Tun et al [47], 2018</td>
<td>Nigeria, West Africa</td>
<td>Cohort study</td>
<td>319</td>
<td>Median 25</td>
<td>Peer or community</td>
<td>They were provided information on HIVST kit use, counseling, and referrals for HIV care and treatment and other support services by a certified HIV testing counselor provide.</td>
<td>N/A</td>
</tr>
<tr>
<td>Wray et al [48], 2018</td>
<td>United States</td>
<td>RCT</td>
<td>65</td>
<td>Mean 34.1 (SD 13.9), range 18-72</td>
<td>Technology or mobile health</td>
<td>They were provided pretest counseling by opening their test kits within 24 hours to answer any questions and offer referrals to other sexual health services over the phones and posttest counseling referral within 24 hours after 100% of detected tests.</td>
<td>N/A</td>
</tr>
<tr>
<td>Wang et al [49], 2018</td>
<td>China</td>
<td>RCT</td>
<td>430</td>
<td>63.1% of the participants were aged between 18 and 30 years</td>
<td>Technology or mobile health</td>
<td>They were provided standard-of-care pretest and posttest counseling by the administrators, and 81% of participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Jin et al [50], 2019</td>
<td>China</td>
<td>Cross-sectional study</td>
<td>879</td>
<td>Median 28 (IQR 24-34)</td>
<td>Technology or mobile health</td>
<td>They were provided with a peer navigator to accompany them to receive confirmatory testing, as well as initial visits for treatment and care following formal diagnosis if they were interested in linkage services.</td>
<td>N/A</td>
</tr>
<tr>
<td>Study</td>
<td>Study setting or country</td>
<td>Study design</td>
<td>Sample size</td>
<td>Age (y)</td>
<td>Counseling delivering modes</td>
<td>Active counseling support</td>
<td>Passive counseling support</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>---------</td>
<td>----------------------------</td>
<td>---------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>De Boni et al [51], 2019</td>
<td>Brazil</td>
<td>Cross-sectional study</td>
<td>2526</td>
<td>Median 25 (IQR 22-31)</td>
<td>Technology or mobile health</td>
<td>They were provided free anonymous HIVST and to enhance linkage to HIV care for those with a confirmed HIV positive status by an internet-based HIVST [electronic testing (e-testing)] approach.</td>
<td>N/A</td>
</tr>
<tr>
<td>Gashobote [52], 2019</td>
<td>Burundi, East Africa</td>
<td>Cross-sectional study</td>
<td>231</td>
<td>18-50</td>
<td>Peer or community</td>
<td>They were provided support, and confirmatory testing by peer educators and health care workers.</td>
<td>N/A</td>
</tr>
<tr>
<td>Hidayat et al [27], 2019</td>
<td>Indonesia</td>
<td>Cross-sectional study</td>
<td>317</td>
<td>Mean 29.9</td>
<td>Peer or community</td>
<td>N/A</td>
<td>They were provided interpreting the results and confirmatory test only determining how many lines appear on the result display.</td>
</tr>
<tr>
<td>Nguyen et al [53], 2019</td>
<td>Vietnam</td>
<td>Cross-sectional study</td>
<td>2185</td>
<td>44.4% of the participants were aged between 16 and 25 years</td>
<td>Peer or community</td>
<td>They were provided results interpretation and observed by peer educators.</td>
<td>N/A</td>
</tr>
<tr>
<td>Vera et al [54], 2019</td>
<td>United Kingdom</td>
<td>Cross-sectional study</td>
<td>232</td>
<td>37% of the participants were aged between 45 and 64 (range 25-65) years</td>
<td>Technology or mobile health</td>
<td>They were provided a detailed information about linkage to care (contact for the nearest sexual health clinic) and support in case of a reactive result (helpline details), and 94% of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Wesolowski et al [55], 2019</td>
<td>United States</td>
<td>Cohort study</td>
<td>80</td>
<td>Most participants were aged 30-54 (range 18-80) years</td>
<td>Technology or mobile health</td>
<td>They were provided instructions of how to conduct test, interpret the results, and posttest counseling.</td>
<td>N/A</td>
</tr>
<tr>
<td>Zhu et al [56], 2019</td>
<td>China</td>
<td>RCT</td>
<td>100</td>
<td>68% of the participants were aged 18-29 (range 18 to &gt;30) years</td>
<td>Technology or mobile health</td>
<td>They were provided pretest and posttest counseling after downloading the WeTest mobile app, and confirmatory testing and linkage to care, and 58%-71% of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Balán et al [57], 2020</td>
<td>United States</td>
<td>RCT</td>
<td>272</td>
<td>Mean 36.6</td>
<td>Technology or mobile health</td>
<td>They were provided instructions of how to use the test kits and linkage to care via daily SMS text messages.</td>
<td>N/A</td>
</tr>
<tr>
<td>Carballo-Díéguez et al [58], 2020</td>
<td>United States</td>
<td>RCT</td>
<td>272</td>
<td>Mean 34 (SD 11)</td>
<td>Technology or mobile health</td>
<td>They were provided rapid oral test kits, instructions, and pretest counseling.</td>
<td>N/A</td>
</tr>
<tr>
<td>Edelstein et al [59], 2020</td>
<td>United States</td>
<td>Cohort study</td>
<td>12,182</td>
<td>51% of the participants were aged between 25 and 34 (range 18 to &gt;45) years</td>
<td>Technology or mobile health</td>
<td>N/A</td>
<td>They were only provided informational inserts developed by NYC health department with on HIV testing and pre- and post-exposure prophylaxis, and confirmatory testing and HIV care when there is a need.</td>
</tr>
<tr>
<td>Study</td>
<td>Study setting or country</td>
<td>Study design</td>
<td>Sample size</td>
<td>Age (y)</td>
<td>Counseling delivering modes</td>
<td>Active counseling support</td>
<td>Passive counseling support</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>----------------------------------------------</td>
<td>------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Johnson et al [60], 2020</td>
<td>United States</td>
<td>Cohort study</td>
<td>922</td>
<td>70% of the participants were aged between 18 and 34 (range 18 to &gt;45 years)</td>
<td>Technology or mobile health</td>
<td>They were provided linkage to PrEP®, HIV/AIDS medical care, partner notification, and other prevention and supportive services.</td>
<td>N/A</td>
</tr>
<tr>
<td>MacGowan et al [61], 2020</td>
<td>United States</td>
<td>RCT</td>
<td>2655</td>
<td>57.3% of the participants were aged between 18 and 30 (range 18 to &gt;30 years)</td>
<td>Technology or mobile health</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Okoboi et al [62], 2020</td>
<td>Uganda</td>
<td>Cross-sectional study</td>
<td>297</td>
<td>Median 25 (IQR 22-28)</td>
<td>Peer or community</td>
<td>They were provided counseling by counselors, linkage participants who tested positive for confirmatory testing and to HIV care services.</td>
<td>N/A</td>
</tr>
<tr>
<td>Phanuphak et al [63], 2020</td>
<td>Thailand</td>
<td>Cohort study</td>
<td>465</td>
<td>Median 26.4 (IQR 22.6-31.7)</td>
<td>Technology or mobile health</td>
<td>They were provided pretest counseling and posttest counseling.</td>
<td>N/A</td>
</tr>
<tr>
<td>Yan et al [64], 2020</td>
<td>China</td>
<td>Cohort study</td>
<td>1315</td>
<td>68% of the participants were aged between 15 and 40 (range 15 to &gt;41 years)</td>
<td>Technology or mobile health</td>
<td>They were provided posttest counseling, laboratory confirmation, and further treatment if they received reactive results by Kang Tong clinic in mailing rapid test reagent kit app.</td>
<td>N/A</td>
</tr>
<tr>
<td>Wang et al [65], 2020</td>
<td>China</td>
<td>Cohort study</td>
<td>510</td>
<td>Median 28 (IQR 23-36)</td>
<td>Peer or community</td>
<td>They were provided confirmatory HIV test result, received pretest and posttest counseling, and referred to treatment.</td>
<td>N/A</td>
</tr>
<tr>
<td>Zhang et al [66], 2020</td>
<td>China</td>
<td>RCT</td>
<td>230</td>
<td>Mean 29 (SD 7.7)</td>
<td>Technology or mobile health</td>
<td>They were provided instructions and counseling information, including 24x7 hotlines and an official WeChat study account to reach research assistants to obtain consultation on the HIVST administration and interpretation of testing results.</td>
<td>N/A</td>
</tr>
<tr>
<td>Zhang et al [67], 2020</td>
<td>China</td>
<td>Cross-sectional study</td>
<td>2364</td>
<td>58.5% of the participants were aged &gt;24 years</td>
<td>Technology or mobile health</td>
<td>They were provided pretest and posttest counseling through the WeChat public platform, the individual WeChat app, or by telephone, and social media delivery strategy was faster in recruiting MSM² to attend HIVST, had a higher degree of linkage to care and ART³, and had a lower economic cost than that of its counterpart.</td>
<td>N/A</td>
</tr>
<tr>
<td>Bell et al [68], 2021</td>
<td>Australia</td>
<td>Cross-sectional study</td>
<td>494</td>
<td>48% of the participants were aged between 20 and 29 years</td>
<td>Technology or mobile health</td>
<td>They were provided the pretest information, 3 monthly testing reminders via phone, email, or SMS text messaging, and a link to a posttest survey via an SMS text message, and 24% and 47% agreed with pretest counseling and posttest counseling over the phone, respectively.</td>
<td>N/A</td>
</tr>
<tr>
<td>Study</td>
<td>Study setting or country</td>
<td>Study design</td>
<td>Sample size</td>
<td>Age (y)</td>
<td>Counseling delivering modes</td>
<td>Active counseling support</td>
<td>Passive counseling support</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Chen et al [69], 2021</td>
<td>South Africa</td>
<td>Cohort study</td>
<td>110</td>
<td>67% of the participants were aged between 18 and 24 (range 18 to &gt;25 years)</td>
<td>Technology or mobile health</td>
<td>They were provided laboratory confirmation, posttest counseling if any participant with a positive test and a care call weekly until it was confirmed the participant had linked to care or the study ended.</td>
<td>N/A</td>
</tr>
<tr>
<td>Cheng et al [70], 2021</td>
<td>China</td>
<td>RCT</td>
<td>491</td>
<td>90% of the participants were aged between 18 and 35 (range 18 to &gt;36 years)</td>
<td>Technology or mobile health</td>
<td>They were provided pretest counseling by consisting of a short message and reminded to do the testing at home.</td>
<td>N/A</td>
</tr>
<tr>
<td>Chan et al [71], 2021</td>
<td>China</td>
<td>Cohort study</td>
<td>350</td>
<td>57.1% of the participants were aged between 18 and 30 (range 18-40 years)</td>
<td>Technology or mobile health</td>
<td>They were provided standard-of-care pretest counseling via video chat, web-based and real-time supervision by the administrators, and 72%-98% of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Hecht et al [72], 2021</td>
<td>United States</td>
<td>Cross-sectional study</td>
<td>625</td>
<td>69% of the participants were aged between 18 and 34 (range 18 to &gt;55 years)</td>
<td>Technology or mobile health</td>
<td>They were provided posttest counseling 10 days after their HIV test kit was mailed.</td>
<td>N/A</td>
</tr>
<tr>
<td>Li et al [73], 2021</td>
<td>China</td>
<td>Cross-sectional study</td>
<td>2263</td>
<td>64% of the participants were aged &gt;24 years</td>
<td>Technology or mobile health</td>
<td>They were provided HIV posttest consultation via WeChat or over the phone to help with the interpretation of the test results and referral to services for clinical confirmatory testing and antiviral treatment.</td>
<td>N/A</td>
</tr>
<tr>
<td>da Cruz et al [74], 2021</td>
<td>Brazil</td>
<td>Cross-sectional study</td>
<td>2681</td>
<td>Median 25 (IQR 21-30)</td>
<td>Technology or mobile health</td>
<td>They were provided linkage to HIV treatment to MSM. One-on-one and SMS text messaging were available for up to 3 months following enrollment in linkage services.</td>
<td>N/A</td>
</tr>
<tr>
<td>Wu et al [75], 2021</td>
<td>China</td>
<td>Cohort study</td>
<td>371</td>
<td>Mean 29 (SD 7)</td>
<td>Technology or mobile health</td>
<td>They were provided posttest counseling, including confirmatory testing and treatment services at a local health facility via telephone calls.</td>
<td>N/A</td>
</tr>
<tr>
<td>Zhang et al [76], 2021</td>
<td>China</td>
<td>Cohort study</td>
<td>471</td>
<td>Median 29 (IQR 25-35)</td>
<td>Technology or mobile health</td>
<td>They were provided web-based services on the application of extra testing kits, instructions on self-testing, real-time consultation with the staff, and uploading of test outcomes and posttest counseling.</td>
<td>N/A</td>
</tr>
<tr>
<td>Abubakari et al [77], 2021</td>
<td>Ghana</td>
<td>Cross-sectional study</td>
<td>61</td>
<td>N/A</td>
<td>Technology or mobile health</td>
<td>They were provided posttest counseling by a smartphone with a preinstalled C5 app, and more than three-quarters of the participants were satisfied the service.</td>
<td>N/A</td>
</tr>
<tr>
<td>Maatouk et al [78], 2021</td>
<td>Lebanon</td>
<td>Cross-sectional study</td>
<td>1103</td>
<td>Mean 26, range 18-57</td>
<td>Technology or mobile health</td>
<td>They were provided HIV counseling and guidance on testing by a hotline.</td>
<td>N/A</td>
</tr>
<tr>
<td>Frye et al [79], 2021</td>
<td>United States</td>
<td>RCT</td>
<td>111</td>
<td>Mean 23 (SD 4)</td>
<td>Peer or community</td>
<td>They were provided standard counseling by a counselor.</td>
<td>N/A</td>
</tr>
<tr>
<td>Study</td>
<td>Study setting or country</td>
<td>Study design</td>
<td>Sample size</td>
<td>Age (y)</td>
<td>Counseling delivering modes</td>
<td>Active counseling support</td>
<td>Passive counseling support</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>--------------------------------------</td>
<td>-----------------------------</td>
<td>--------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Girault et al [80], 2021</td>
<td>Thailand</td>
<td>Cross-sectional study</td>
<td>1422</td>
<td>45% aged between 15 and 24, range 15 to &gt;30 years</td>
<td>Peer or community</td>
<td>N/A</td>
<td>They were provided oral instructions before and during the test, and assistance in conducting the test or interpreting the result when requested, and only 57% of the participants were satisfied with the service.</td>
</tr>
<tr>
<td>Phongphiew et al [81], 2021</td>
<td>Thailand</td>
<td>Cross-sectional study</td>
<td>45</td>
<td>Mean 17.6 (SD 1.1)</td>
<td>Peer or community</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Widyantini et al [82], 2021</td>
<td>Indonesia</td>
<td>Cross-sectional study</td>
<td>813</td>
<td>52% of the participants were aged between 16 and 29, range 16 to &gt;39 years</td>
<td>Peer or community</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Wirtz et al [83], 2021</td>
<td>Myanmar</td>
<td>RCT</td>
<td>63</td>
<td>Median 21 (IQR 19-25)</td>
<td>Peer or community</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Dijkstra et al [84], 2021</td>
<td>Kenya, East Africa</td>
<td>Cross-sectional study</td>
<td>452</td>
<td>Median 26 (IQR 22-30)</td>
<td>Peer or community</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>O’Byrne et al [85], 2021</td>
<td>Canada</td>
<td>Cross-sectional study</td>
<td>111</td>
<td>Mean 31 years</td>
<td>Peer or community</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>
Quality Assessment

The National Institutes of Health quality assessment tool was used to assess the quality of RCTs and observational studies [87]. The tool covers 14 domains for RCTs, observational cohorts, and cross-sectional studies, with a total score ranging from 0 to 14. Higher scores indicated better quality, and each study’s summary score was categorized as poor (0-4 out of 14 questions), fair (5-10 out of 14 questions), or good (11-14 out of 14 questions).

Data Analysis

Meta-analyses were conducted using random effects models to combine data and calculate pooled proportions and 95% CIs based on the generalized linear mixed effects method [88]. Heterogeneity was quantified using $I^2$ statistic. $I^2$ values <25%, 25% to 75%, and >75% indicate low, moderate, and high heterogeneity, respectively [89]. We used visual inspection to assess the asymmetry of funnel plots and the Egger test to detect potential publication bias [90]. Sensitivity analysis was conducted by removing one study at a time.

Subgroup analyses and meta-regression were conducted to assess whether the proportion of linkage to care varied and was predicted according to the values of the study characteristics. These included study year (we used 2016 as the cutoff as it was the year when the WHO started recommending HIVST) [91], study sample size (<300 vs ≥300), study countries (high income, upper middle income, lower middle income, and low income based on the new World Bank country classification) [92], HIVST counseling delivery modes (technology and mobile health vs peer and community), presence of pretest and posttest counseling (posttest counseling only vs both pre- and posttest counseling), and quality of counseling. Quality of counseling was measured by the number of essential components involved in the counseling support for MSM HIVST users.

Among all studies, we assessed whether the overall linkage to care varied by study characteristics (type of counseling support, study year, study sample size, study countries, HIVST counseling delivery modes, presence of pretest and posttest counseling, and quality of counseling) using the univariate meta-regression model. Factors with $P<.10$ in univariate meta-regression analyses were entered into the multivariable meta-regression model. Within studies providing active or passive counseling support, univariate and multivariate meta-regression were used to examine whether linkage to care varied by study characteristics. All analyses were performed with R (version 4.2.1; R Foundation for Statistical Computing) using the metafor package.

Results

Study Characteristics

A flowchart of the literature selection process is presented in Figure 1. The initial search yielded 1362 publications through databases and registers, and 55 studies met the eligibility criteria and were included in the systematic review [27,33-86]. All 55 studies were included in the meta-analysis to estimate the pooled proportion of linkage to care among MSM HIVST users, categorized by active and passive counseling support.
The characteristics of the included studies are presented in Table 2, and the results for studies that assessed linkage to care are presented in Table 3. All studies were published from 2014 to 2021, including 53 full-text manuscripts [27,33,34,36-51,53-86] and 2 conference abstracts [35,52]. The 55 included studies reported data on 45,147 MSM aged 15 to 73 years. According to the geographical location, 26 (47%) studies were conducted in the Asia-Pacific region, 16 (29%) in North America, 9 (16%) in Africa, 3 (6%) in South America, and 1 (2%) in Europe. Of the included studies, based on the study design, 43 (78%) were observational studies, including 28 (65%) cross-sectional and 15 (35%) cohort studies, and 12 (22%) were RCTs. Detailed results of the quality assessment of the studies are presented in Multimedia Appendix 3 [27,33-86].
Table 3. Results for studies that assessed linkage to care.

<table>
<thead>
<tr>
<th>Linkage to care</th>
<th>Overall proportion, (95% CI)</th>
<th>$I^2$</th>
<th>Egger test (P value)</th>
<th>Number of estimates (references)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active counseling support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting test results</td>
<td>97.2 (74.3%–99.8%)</td>
<td>100%</td>
<td>.25</td>
<td>9 [41, 51, 56, 64, 66, 67, 73, 75, 76]</td>
</tr>
<tr>
<td>Laboratory confirmation</td>
<td>92.6 (86.1%–96.2%)</td>
<td>57%</td>
<td>.34</td>
<td>35 [34, 37, 39, 47, 51–53, 56, 60, 62, 63, 65, 66, 69, 71, 73–75, 78, 79, 81–83, 86]</td>
</tr>
<tr>
<td>ART* initiation</td>
<td>90.8 (86.7%–93.7%)</td>
<td>0%</td>
<td>.97</td>
<td>21 [37, 43–45, 47, 49, 52, 53, 60, 62, 63, 66, 67, 70, 71, 73, 78, 81, 82, 84, 86]</td>
</tr>
<tr>
<td>Referral to physicians</td>
<td>96.3 (85%–99.2%)</td>
<td>0%</td>
<td>.14</td>
<td>28 [33, 35, 36, 38, 39, 41, 44–47, 49, 54, 55–57, 60, 63, 65, 66, 68–71, 74, 75, 77, 79, 81]</td>
</tr>
<tr>
<td>Prevention strategies</td>
<td>100.0 (0%–100%)</td>
<td>0%</td>
<td>&lt;.001</td>
<td>7 [34, 49, 60, 63, 71, 81, 85]</td>
</tr>
<tr>
<td>PrEP** initiation</td>
<td>27.0 (10.2%–54.6%)</td>
<td>97%</td>
<td>.73</td>
<td>6 [48, 60, 72, 81, 84, 85]</td>
</tr>
<tr>
<td><strong>Passive counseling support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting test results</td>
<td>—</td>
<td>—</td>
<td></td>
<td>1 [50]</td>
</tr>
<tr>
<td>Laboratory confirmation</td>
<td>78.7 (17.8%–98.4%)</td>
<td>82%</td>
<td>.02</td>
<td>5 [27, 50, 59, 61, 80]</td>
</tr>
<tr>
<td>ART initiation</td>
<td>79.1 (48.8%–93.7%)</td>
<td>0%</td>
<td>.06</td>
<td>4 [50, 59, 61, 80]</td>
</tr>
<tr>
<td>Referral to physician</td>
<td>79.1 (0%–100%)</td>
<td>91%</td>
<td>&lt;.001</td>
<td>2 [59, 61]</td>
</tr>
<tr>
<td>Prevention strategies</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrEP initiation</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aART: antiretroviral therapy.

**PrEP: pre-exposure prophylaxis.

*cPooled proportion was not performed because of fewer than 2 studies.

Active Counseling Supporting HIVST

Overall, 91% (50/55) of the studies provided active counseling support for MSM HIVST users [33-49, 51-58, 60, 62-79, 81-86]. These studies were conducted in China (14/50, 28%), the United States (12/50, 24%), Brazil (3/50, 6%), Australia (2/50, 4%), Uganda (2/50, 4%), Vietnam (2/50, 4%), South Africa (2/50, 4%), Canada (2/50, 4%), Burundi (2/50, 4%), Thailand (2/50, 4%), the United Kingdom (1/50, 2%), Nigeria (1/50, 2%), Ghana (1/50, 2%), Lebanon (1/50, 2%), Indonesia (1/50, 2%), Myanmar (1/50, 2%), and Kenya (1/50, 2%). Most of these studies (38/50, 76%) were conducted in high-income and upper middle-income countries.

Moreover, 66% (33/50) of the studies used mobile health technology to deliver active counseling support [33-34, 36-39, 41, 45, 46, 48, 49, 51, 54-58, 60, 63, 64, 66-76]. Mobile phones and the internet (eg, telephone calls, SMS text messages, and emails) were the most commonly used technology (21/33, 64%), followed by social media apps or geospatial dating apps (eg, WeChat, QQ, Blued, and Grindr; 8/33, 24%) and HIVST-specific apps (eg, HIVSmart!, WeTest, and C5 apps; 4/33, 12%). Other studies (n=17) used sex partners (3/17, 18%) and nurses or physicians (14/17, 82%) to deliver [33, 38, 42, 44, 45, 52, 55, 56, 64, 76-83].

Essential Components Involved in the Pretest Counseling of the Active Counseling Supporting HIVST

In the pretest counseling, 32 studies proactively provided at least one essential component to MSM HIVST users (Multimedia Appendix 4 [27, 33-86]). Five studies provided only one essential component, such as informing users of the benefits of taking the tests (1/5, 20%), assuring users’ rights to refuse HIV testing (2/5, 40%), or providing HIV prevention information (2/5, 40%). Five other studies provided 2 essential components. In addition to informing the users of the benefits of taking the tests, these studies provided risk assessment (1/5, 20%), assured users’ right to refuse (2/5, 40%), encouraged beneficial disclosure of HIV serostatus (1/5, 20%), or provided HIV prevention information (1/5, 20%). Four other studies provided 3 essential components. In addition to informing the users about the benefits of taking the test, the combination of other components were (1) providing risk assessment and HIV prevention information (1/4, 25%), (2) encouraging beneficial disclosure of serostatus and providing HIV prevention information (1/4, 25%), (3) assuring users’ right to refuse HIV
testing and providing HIV prevention information (1/4, 25%), and (4) providing risk assessment and assuring users’ right to refuse (1/4, 25%). Another study provided 4 essential components: (1) informing users of the benefits of taking the tests, (2) providing risk assessment, (3) assuring the user’s right to refuse HIV testing, and (4) encouraging beneficial disclosure of serostatus status. The remaining studies (n=13) provided all 5 essential components.

In addition to these essential components, 15 studies provided other supplementary components. These components consisted of (1) the reason for HIVST (1/15, 7%), (2) stories addressing general health concerns of MSM (2/15, 13%), (3) the assessment of potential social support (4/15, 27%), (4) first-person stories about people living with HIV (1/15, 7%), and (5) local data, news, and policies regarding HIV and sexually transmitted infections among MSM (7/15, 47%).

**Essential Components Involved in the Posttest Counseling of the Active Counseling Supporting HIVST**

In posttest counseling, 50 studies proactively provided at least one essential component to MSM HIVST users (Multimedia Appendix 4). Six studies only provided one essential component, offering additional HIV care to MSM who received reactive HIVST results. Moreover, 24 studies provided 2 essential components. These combinations included (1) additional HIV care for users with reactive HIVST results and interpretation of HIVST results (23/24, 96%) and (2) additional HIV care for users with reactive HIVST results and HIV prevention information for users with negative HIVST results (1/24, 4%). The rest of the studies (n=18) provided all 3 essential components (eg, interpretation of testing results, HIV prevention information for users with negative HIVST results, and additional HIV care for users with reactive HIVST results).

**Passive Counseling Supporting HIVST**

Passive counseling support to MSM was provided in 10% (5/50) of the studies [27,50,59,61,80]. These studies were conducted in the United States (2/5, 40%), China (1/5, 20%), Indonesia (1/5, 20%), and Thailand (1/5, 20%). Three studies used mobile health technology to deliver counseling [50,59,61], whereas the other 2 studies used peers and communities to provide counseling [27,80].

**Essential Components Involved in the Pretest Counseling of Passive Counseling Supporting HIVST**

In the pretest counseling, 3 studies offered at least one essential component upon request. One study only informed the benefits of taking the tests. The other 2 studies included 2 essential components: informing the participants of the benefits of taking the tests and providing HIV prevention information. In addition to these essential components, one study provided local data, news, and policies regarding HIV and sexually transmitted infections among MSM.

**Essential Components Involved in the Posttest Counseling of the Passive Counseling Supporting HIVST**

In posttest counseling, all 5 studies provided at least one essential component upon request. Three studies provided one essential component, such as interpretation of the HIVST results (1/5, 20%), or provision of psychological support and referral to HIV care for users with positive results (2/5, 40%). The other 2 studies provided 2 essential components. In addition to the interpretation of HIVST results, these studies provided referral to HIV care for users with positive results (1/5, 20%) or HIV prevention information for users with negative results (Multimedia Appendix 4).

**Meta-Analysis of Linkage to Care Among MSM HIVST Users Along With Active and Passive Counseling**

The main findings of the meta-analysis of the linkage to care among MSM HIVST users are summarized in Figure 2.
Reporting Testing Results
Our meta-analysis comprised 45,147 MSM, using a random effects model (Table 3). Overall, 10 studies measured the proportion of users who reported HIVST results [41,50,51,56,64,66,67,73,75,76]. In studies providing active counseling support, the pooled proportion of reporting HIVST results was 97.2% (n=9; 95% CI 74.3%-99.8%; $I^2=100%$; Figure 3 [41,51,56,64,66,67,73,75,76]). Only one study providing passive counseling support reported a proportion of 77.7% for reporting HIVST results [50]. However, this study was not included in the meta-analysis owing to insufficient data.

Linkage to Laboratory Confirmation
In total, 40 studies assessed linkage to laboratory confirmation among users with positive results [27,34,36,37,39-47,49,51-53,56-63,65,66,68,69,71,73-75,78,81-83,86]. In the studies providing active counseling support, the pooled proportion of linkage to laboratory confirmation was 92.6% (n=35; 95% CI 86.1%-96.2%; $I^2=57%$; Figure 4 [34,36,37,39-47,49,51-53,56-86,65,66,68,69,71,73-75,78,81-83,86]). In studies providing passive counseling support, the pooled proportion of linkage to laboratory confirmation was 78.7% (n=5; 95% CI 17.8%-98.4%; $I^2=82%$; Figure 4).
Linkage to ART Initiation

Overall, 25 studies reported linkage to ART initiation among users with positive results [37,43-45,47,49,50,52,53,59-63,66,67,70,71,73,78,80-82,84,86]. In the studies providing active counseling support, the pooled proportion of ART initiation was 90.8% (n=21; 95% CI 86.7%-93.7%; $I^2=0$%); Figure 5 [37,43-45,47,49,50,52,53,59-63,66,67,70,71,73,78,80-82,84,86]. In studies with passive counseling support, the pooled proportion of ART initiation was 79.1% (n=4; 95% CI 48.8%-93.7%; $I^2=0$%; Figure 5) [50,59,61,80].
Figure 5. Forest plot of the pooled proportion of linkage to antiretroviral therapy among users who were confirmed HIV positive: (A) Studies were provided active counseling along with HIV self-testing and (B) studies were provided passive counseling along with HIV self-testing.

### Linkage to Physicians

In total, 30 studies reported linkage to physicians among users with positive results ([33,35,36,38,39,41,44-47,49,54,55-57,60,63,65,66,68-71,74,75,77,79,81]). In the studies providing active counseling support, the pooled proportion of linkage to physicians was 96.3% (n=28; 95% CI 85% - 99.2%; $I^2=0$%; Figure 6) ([33,35,36,38,39,41,44-47,49,54,55-57,60,63,65,66,68-71,74,75,77,79,81]). In the studies with passive counseling support, the pooled proportion of linkage to physicians was 79.1% (n=2; 95% CI 0%-100%; $I^2=91$%; Figure 6) ([59,61]).
Figure 6. Forest plot of the pooled proportion of linkage to physicians among users with reactive results: (A) studies were provided active counseling along with HIV self-testing and (B) studies were provided passive counseling along with HIV self-testing.

Linkage to Information Related to Sexual Risk Behaviors Reduction and PrEP and PrEP Initiation

Overall, 7 studies reported linkage to information related to sexual risk behaviors reduction and PrEP among users with negative results. In studies with active counseling support, the pooled proportion of linkage to sexual risk behaviors reduction was 100% (n=7; 95% CI 0%-100%; $I^2=0%$; Figure 7 [34,49,60,63,71,81,85]). The pooled proportion of PrEP initiation was 27% (n=6; 95% CI 10.2%-54.6%; $I^2=97%$; Figure 7) [48,60,72,81,84,85] in studies providing active counseling support. No studies with passive counseling support reported a linkage to information related to sexual risk behaviors and PrEP or PrEP initiation among users with negative results.
Publication Bias

Upon examination of the funnel plots (Figure 8; Table 3), there was a publication bias in studies reporting the proportion of linkage to laboratory confirmation ($P=.02$), referral to physicians ($P<.001$), and prevention strategies ($P<.001$). Furthermore, outliers were identified in studies reporting the proportion of reporting test results [51] and linkage to laboratory confirmation [73,82], ART initiation [63], physicians [35], and PrEP initiation [85] (Figure 8).
Figure 8. Funnel plots for assessing the publication bias among the included studies: (A) linkage to reporting test results; (B) linkage to laboratory confirmation with active counseling support; (C) linkage to laboratory confirmation with passive counseling support; (D) linkage to antiretroviral therapy (ART) initiation with active counseling support; (E) linkage to ART initiation with passive counseling support; (F) linkage to physicians with active counseling support; (G) linkage to physicians with passive counseling support; (H) linkage to information related to sexual risk behaviors reduction, and pre-exposure prophylaxis (PrEP); and (I) linkage to PrEP initiation.

Sensitivity Analysis

For studies providing active counseling support, the pooled proportion of reporting test results, and linkage to laboratory confirmation, ART initiation, physicians, information related to sexual risk behaviors reduction and PrEP, and PrEP initiation changed slightly after removing one study at each time.

With regard to the studies providing passive counseling support, the pooled proportion of linkage to ART initiation did not change after removing one study at each time. However, after removing the study conducted by Jin et al [50], the pooled proportion of linkage to laboratory confirmation changed from 78.7% (95% CI 17.8%-98.4%) to 55.5% (95% CI 27%-80.8%). As there were only 2 studies with passive counseling support that reported linkage to physicians, a sensitivity analysis was not conducted. Details of the sensitivity analysis results are presented in Multimedia Appendix 5 [27,33-86].

Subgroup Analysis and Metaregression

Subgroup analysis by study sample size (<300 vs ≥300) revealed different levels of linkage to reporting test results (98.5% vs 82.7%), laboratory confirmation (93.8% vs 85.6%), physicians (92% vs 86%), and PrEP initiation (35.6% vs 23.9%). The subsequent subgroup analysis (posttest counseling only vs both pre- and posttest counseling) also found different proportions of users who reported test results (87.1% vs 93.2%) and initiated PrEP (20.2% vs 53%). Subgroup analysis by other study characteristics did not reveal a large difference in the linkage to care.

Among all studies, univariate metaregression analysis demonstrated that the type of counseling (active vs passive), a smaller sample size (<300 vs ≥300), and a higher number of essential components involved in the counseling support were significantly associated with better linkage to care. Furthermore, the multivariate metaregression analysis confirmed that a larger sample size was linked to a lower linkage to care ($P=.03$). In contrast, mobile health technology counseling ($P=.05$) and a higher number of essential components involved in the counseling support were associated with increased linkage to care ($P=.04$).

In studies providing active counseling support, univariate metaregression analysis indicated a smaller sample size (<300 vs ≥300), provision of both pretest and posttest counseling (vs
posttest counseling only), and a higher number of essential components were significantly associated with better linkage to care. The findings of multivariate metaregression analysis revealed that a smaller sample size ($P = .03$) and using mobile health technology for counseling ($P = .05$) were associated with a higher linkage to care. With regard to different outcomes related to linkage to care, a larger sample size was correlated with a lower linkage to laboratory confirmation ($P = .03$) and prevention strategies ($P < .001$) with active counseling support (Multimedia Appendix 6).

**Discussion**

**Principal Findings**

This systematic review and meta-analysis aimed to summarize the global evidence on counseling support and synthesize the proportion of linkage to care among MSM HIVST users. We categorized counseling support in assisted HIVST into active or passive. More than 90% of the MSM HIVST users with reactive results were linked to laboratory confirmation and ART initiation in studies implementing active counseling support. Such a proportion was higher than that of the studies with passive counseling support (78.7%-79.1%). Therefore, the provision of active counseling support may be helpful in improving the linkage to HIV care and treatment for MSM HIVST users with reactive results.

Relatively few studies (7/55, 13%) provided information related to sexual risk behaviors reduction and PrEP for MSM HIVST users with negative results. One possible explanation was that most resources were used to provide support for users with reactive results, which was considered a priority for some HIVST programs [93]. Hence, there are constraints in resources to provide support for the large number of users with negative results [93]. In addition to identifying individuals testing positive for HIV, facilitating behavior changes is an important purpose of HIV testing and counseling. Future studies should consider providing more comprehensive support for MSM HIVST users with negative results.

The metaregression results identified some significant determinants of the linkage to care among MSM HIVST users. First, more essential counseling components were associated with better linkage to care, which aligns with findings from a previous study [94]. Incorporating a higher number of essential components would enhance counseling quality. Previous studies suggest that delivering high-quality counseling improves linkage to care, reduces risky behaviors, and prevents new infections [95,96]. However, our study found that only 26% to 36% of the studies provided all essential active pretest and posttest counseling support. As a result, future programs should consider offering comprehensive counseling to MSM HIVST users.

Second, a larger sample size was associated with a lower linkage to care among MSM HIVST users. Providing active counseling support for HIVST users was resource demanding. For example, it took 1 hour to prepare and implement one session of real-time pretest and posttest counseling support for each MSM HIVST user [49,71]. Therefore, it is challenging to provide counseling support to a larger number of HIVST users. Furthermore, mobile health technology counseling was associated with a better linkage to care compared with peer and community counseling. Our study found that mobile health technology counseling is the predominant method to support MSM HIVST users, which aligns with a previous study [28]. Owing to high smartphone ownership among MSM (>94%) [97], mobile health technology presents a viable strategy for counseling among MSM HIVST users. In addition, previous studies have demonstrated that using mobile health technology for counseling support reduces the workload of HIV testing administrators [28,98].

Stigma and discrimination against MSM impede access to HIV testing and counseling services [99]. Systematic reviews have shown that perceived stigma remained a significant obstacle to engaging in assisted HIVST [100,101]. Future programs should consider increasing the empathy of health workers who provide counseling to support HIVST. A previous study suggested that the negative effects of perceived stigma or discrimination on HIV testing use could be offset by increasing the empathy of service providers [102]. A recent study applied computerized programs (instead of people) to provide active counseling supporting HIVST [98]. Such an approach could also reduce the concerns of stigma or discrimination when using assisted HIVST among MSM.

**Limitations**

There are several limitations in this study. First, there was high heterogeneity among the studies that reported outcomes on linkage to laboratory confirmation, physicians with the provision of passive counseling support, linkage to PrEP initiation, and reporting test results. Heterogeneity pertains to the diversity observed in the design of studies, the effects of interventions, or the outcomes obtained across different studies. Persistent heterogeneity could not be resolved using sensitivity analysis. Second, publication bias was found in relation to the pooled proportion of linkage to laboratory confirmation, physicians, and prevention strategies, which could impact the validity and generalization of conclusions. Third, because half of the included studies were cross-sectional, a causal relationship could not be established. Furthermore, the use of nonprobabilistic sampling for MSM in all the included studies limited the generalizability of the findings. Finally, the small number of studies in the comparison group (those providing passive counseling) would result in bias when comparing the study outcomes between active and passive counseling support.

**Conclusions**

This study synthesized evidence on active and passive counseling support for MSM HIVST users and quantified the proportion of linkage to care. As compared with passive counseling support, active counseling support had a better linkage to care. Having a higher number of essential counseling components, a smaller sample size, and using mobile health technology to deliver counseling support were also associated with a better linkage to care. As our results showed, proactively providing counseling support for all users, involving more essential components in the counseling support, and using mobile health technology should be considered to increase the linkage to care among MSM HIVST users.
Acknowledgments
This study was supported by the AIDS Trust Fund (MSS349R). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Data Availability
All data generated or analyzed during this study are included in this published paper (and its Multimedia Appendix files).

Authors' Contributions
SC, YF, PM, and ZW conceptualized and designed the study. SC, YF, and ZW were involved in data curation and formal analysis. SC, YF, PSFC, JK, PM, and ZW verified the data. SC and ZW drafted the manuscript. All authors critically revised and drafted the manuscript for important intellectual content, including data interpretation in the broader context. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 28 KB - publichealth_v10i1e45647_app1.docx ]

Multimedia Appendix 2
Search terms.
[DOCX File, 29 KB - publichealth_v10i1e45647_app2.docx ]

Multimedia Appendix 3
Quality assessment of the included studies using the National Institutes of Health quality assessment tool.
[DOCX File, 41 KB - publichealth_v10i1e45647_app3.docx ]

Multimedia Appendix 4
Summary of counseling components in included studies.
[DOCX File, 37 KB - publichealth_v10i1e45647_app4.docx ]

Multimedia Appendix 5
Summary of sensitivity analysis.
[DOCX File, 52 KB - publichealth_v10i1e45647_app5.docx ]

Multimedia Appendix 6
Subgroup analysis and metaregression analyses.
[DOCX File, 52 KB - publichealth_v10i1e45647_app6.docx ]

References


37. Rosengren AL, Huang E, Daniels J, Young SD, Marlin RW, Klausner JD. Feasibility of using Grindr TM to distribute HIV self-test kits to men who have sex with men in Los Angeles, California. Sex Health (Forthcoming) 2016 May 23 [FREE full text] [doi: 10.1071/SH15236] [Medline: 27209076]


**Abbreviations**

ART: antiretroviral therapy  
HIVST: HIV self-testing  
MSM: men who have sex with men  
PrEP: pre-exposure prophylaxis  
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
RCT: randomized controlled trial  
UNAIDS: Joint United Nations Programme on HIV/AIDS  
WHO: World Health Organization
Understanding Gaps in the Hypertension and Diabetes Care Cascade: Systematic Scoping Review

Jie Wang*, BMed; Fangqin Tan*, BMed; Zhenzhong Wang, BM; Yiwen Yu, MPH; Jingsong Yang, BMed; Yueqing Wang, MPH; Ruitai Shao, PhD; Xuejun Yin*, PhD

1School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China
2The George Institute for Global Health, University of New South Wales, Sydney, Australia
*these authors contributed equally

Corresponding Author:
Xuejun Yin, PhD
School of Population Medicine and Public Health
Chinese Academy of Medical Sciences & Peking Union Medical College
31 Beijige San Tiao
Dongcheng District
Beijing, 100005
China
Phone: 86 18600988138
Email: yinxuejun@cams.cn

Abstract

Background: Hypertension and diabetes are global health challenges requiring effective management to mitigate their considerable burden. The successful management of hypertension and diabetes requires the completion of a sequence of stages, which are collectively termed the care cascade.

Objective: This scoping review aimed to describe the characteristics of studies on the hypertension and diabetes care cascade and identify potential interventions as well as factors that impact each stage of the care cascade.

Methods: The method of this scoping review has been guided by the framework by Arksey and O’Malley. We systematically searched MEDLINE, Embase, and Web of Science using terms pertinent to hypertension, diabetes, and specific stages of the care cascade. Articles published after 2011 were considered, and we included all studies that described the completion of at least one stage of the care cascade of hypertension and diabetes. Study selection was independently performed by 2 paired authors. Descriptive statistics were used to elucidate key patterns and trends. Inductive content analysis was performed to generate themes regarding the barriers and facilitators for improving the care cascade in hypertension and diabetes management.

Results: A total of 128 studies were included, with 42.2% (54/128) conducted in high-income countries. Of them, 47 (36.7%) focused on hypertension care, 63 (49.2%) focused on diabetes care, and only 18 (14.1%) reported on the care of both diseases. The majority (96/128, 75.0%) were observational in design. Cascade stages documented in the literature were awareness, screening, diagnosis, linkage to care, treatment, adherence to medication, and control. Most studies focused on the stages of treatment and control, while a relative paucity of studies examined the stages before treatment initiation (76/128, 59.4% vs 52/128, 40.6%). There was a wide spectrum of interventions aimed at enhancing the hypertension and diabetes care cascade. The analysis unveiled a multitude of individual-level and system-level factors influencing the successful completion of cascade sequences in both high-income and low- and middle-income settings.

Conclusions: This review offers a comprehensive understanding of hypertension and diabetes management, emphasizing the pivotal factors that impact each stage of care. Future research should focus on upstream cascade stages and context-specific interventions to optimize patient retention and care outcomes.

(JMIR Public Health Surveill 2024;10:e51802) doi:10.2196/51802

KEYWORDS
care cascade; hypertension; diabetes; scoping review; hypertension and diabetes care; review
Introduction

Noncommunicable diseases (NCDs) constitute a formidable global health challenge, accounting for approximately 80% of NCD-related deaths, and they include cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes [1]. Hypertension is the most pivotal risk factor for cardiovascular diseases [2]. The prevalence of hypertension among adults aged 30-79 years worldwide is estimated to be 1.28 billion, with an alarming 700 million individuals unaware of their condition. Less than half of adults with hypertension are diagnosed and treated. Only approximately 1 in 5 adults with hypertension have their blood pressure controlled [3]. Similarly, the global prevalence of diabetes among adults has surged to 537 million in 2021, with nearly half of these cases (240 million) remaining undiagnosed. Moreover, the treatment rate for diabetes is suboptimal, with only 32.9% of patients receiving appropriate care and a mere 16.5% attaining treatment goals [4]. Evidence suggests that a substantial proportion of patients with hypertension and diabetes reside in low- and middle-income countries (LMICs), wherein the management of these conditions remains persistently low [3,5].

Hypertension and diabetes are often approached differently by distinct clinical subspecialties owing to their clinical complexities. However, it is essential to recognize that the management of these 2 conditions together can be highly beneficial owing to their shared risk factors and bidirectional interaction. The management of hypertension and diabetes also shares the same pathway, which includes early detection, appropriate treatment, and continuous monitoring. The health care systems and implementation strategies designed to ensure the continuity of care exhibit significant overlap and can be harnessed efficiently and effectively to support both hypertension and diabetes patients. The care cascade is a model for evaluating patient retention across sequential stages of care required to achieve a successful treatment outcome [6]. This model has sequential stages, including awareness, screening, diagnosis, appropriate management, and disease control, that patients navigate while accessing health care services. Acknowledging potential lapses at each stage, the care cascade model identifies critical stages where patients may disengage, hindering them from attaining disease control [7]. The care cascade model was originally conceived for HIV care [8]. It has since been extended to monitor and manage infectious diseases, such as hepatitis C [9] and tuberculosis [6], and has been more recently applied to NCDs [10].

The utility of studying the care cascade of hypertension and diabetes goes beyond the mere exploration of their clinical pathways. It encompasses a broader holistic perspective that includes not only clinical aspects but also the impact on health systems, the quality of life of affected individuals, and the efficiency of health care delivery. Cascade analysis for hypertension and diabetes can help understand the common factors that affect the care model in order to identify appropriate strategies to improve health care for these 2 conditions. However, there is limited evidence synthesis regarding the care cascade of hypertension and diabetes. Therefore, we performed a systematic scoping review with the goal of mapping and describing the current state of evidence on a global scale. We sought to understand the process of hypertension and diabetes management, identify the factors that influence each stage of the care cascade, and explore potential interventions that hold promise for improving care continuity. By synthesizing existing evidence, our findings seek to inform future research endeavors, propelling the advancement of management strategies for hypertension and diabetes.

Methods

Study Design

This scoping review was conducted following the stages of a scoping review by Arksey and O’Malley [11] and was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [12].

Identifying the Research Questions

This scoping review focused on mapping the existing evidence on the care cascade of hypertension and diabetes. The specific research questions were as follows: 1. How the care cascade model has been applied in hypertension and diabetes research? 2. Which stage of the hypertension and diabetes care cascade has the current research in high-income countries (HICs) and LMICs primarily focused on? 3. What are the barriers and facilitators of hypertension and diabetes care cascade completion? 4. What strategies have been employed to improve retention in the hypertension and diabetes care cascade? 5. What are the key knowledge gaps that remain in the literature about the hypertension and diabetes care cascade?

Identifying Relevant Studies

To identify relevant studies, a systematic search was conducted in MEDLINE, Embase, and Web of Science, using terms pertinent to hypertension, diabetes, and the specific stages of the care cascade. The framework of population, concept, and context was used to identify core concepts related to the research question and inform the search strategy [13]. A complete overview of the search terms for each database is provided in Multimedia Appendix 1. The population of interest in this review was adults aged 18 years or older who had been screened for or diagnosed with hypertension or type 2 diabetes, as well as patients with hypertension or type 2 diabetes undergoing treatment or management. The key concept of the review was the hypertension and diabetes care cascade, with a focus on studies explicitly applying the cascade care model to one or more stages of screening, diagnosis, treatment, and control. Articles describing interventions targeting specific stages of the cascade or factors influencing interventions or outcomes of at least one stage of the care cascade were included. The review aimed to explore a broad range of influencing factors involving both barriers and facilitators, with all pertinent descriptions included, regardless of statistical associations. The contextual scope of this review was in both HICs and LMICs, where hypertension and diabetes care was provided. The timeframe for database searches spanned from January 2011 to January 2021.
2023 since the concept of the care cascade was introduced in 2011 [8]. There was no restriction on publication language, allowing for an inclusive evaluation of relevant studies worldwide. The eligibility criteria are shown in Textbox 1.

Textbox 1. Eligibility criteria for screening.

1. Population: Adults aged 18 years or older who had been screened for or diagnosed with hypertension or type 2 diabetes and patients with known hypertension or type 2 diabetes currently undergoing treatment or management.
2. Concept: Hypertension and diabetes care cascade, with a focus on studies explicitly applying the cascade care model to one or more stages from awareness to control. Interventions that impact patient outcomes and factors that influence implementation outcomes and service outcomes within at least one stage of the care cascade.
3. Context: No limitation. All clinical and primary care settings.
4. Language: No limitation.
6. Article type: Original articles and protocol papers, including cross-sectional studies, cohort studies, trials, and implementation studies published in peer-reviewed journals.

Study Selection

All identified articles were imported into Covidence, and duplicates were removed. Screening proceeded through 2 distinct stages, where titles and abstracts were assessed independently by 4 researchers (JW, FT, XY, and ZW) in pairs, adhering to predefined inclusion and exclusion criteria to determine potential eligibility. In the event of disagreements, a collaborative discussion within the research team swiftly resolved any discrepancies. Subsequently, full-text screening followed a similar process, again involving 4 researchers (JW, FT, YY, and ZW) in pairs. Articles were excluded if they were (1) observing outcomes unrelated to hypertension and type 2 diabetes health care; (2) case reports, conference abstracts, editorials, commentaries, or reviews; and (3) unavailable in full text. Any unresolved discrepancies in article eligibility were resolved by group discussion until a consensus was reached. Notably, to glean insights into ongoing or planned projects and to identify potential interventions and relevant factors, study protocols were intentionally retained and not excluded in the scoping review.

Charting the Data

A data-charting form was created in Microsoft Excel and pilot tested with 15 articles to establish clarity and consistency in data extraction variables across reviewers. Data extraction was performed by 4 researchers (JW, FT, YY, and ZW). The extracted variables included title, author names, year of publication, study countries, disease of interest (hypertension, diabetes, or both), study method (quantitative, qualitative, or mixed method), study design (cross-sectional study, cohort study, trial, or implementation study), sample size, mean age of participants, stages of the care cascade involved, interventions aimed at improving retention, factors associated with stage completion, and reported outcomes. The care of hypertension and diabetes was divided into multiple stages of the cascade, including awareness, screening, diagnosis, linkage to care, treatment, medication adherence, and ultimately, disease control. The world’s economies were classified based on the World Bank classification as follows: low income, lower-middle income, upper-middle income, and high income [14]. The outcomes were categorized into implementation outcomes, service outcomes, and client outcomes. Implementation outcomes encompassed aspects pertaining to the process of implementing interventions and services for hypertension and diabetes care. This included factors such as acceptability, adoption, appropriateness, cost, feasibility, fidelity, and sustainability of interventions to health care providers or patients. Service outcomes were those related to the quality and effectiveness of the health care services provided to patients with hypertension and diabetes, such as access to health care services, continuity of care, appropriateness of care, equity of service, and health care provider adherence to clinical guidelines. Client outcomes, on the other hand, delved into the impact of health services on patients’ health and clinical conditions, such as blood pressure and blood glucose control, reductions in cardiovascular risk factors, and improvements in overall quality of life. To ensure data accuracy and consistency, a senior researcher (XY) reviewed all extracted data. Any disagreements were resolved by consensus.

Collating, Summarizing, and Reporting the Results

Interventions and influencing factors were analyzed by cascade stages and focused diseases. Studies that reported multiple stages of the care cascade were included in the synthesis of each relevant stage. The resulting information was subjected to rigorous quantitative analysis, employing frequencies and percentages to elucidate key patterns and trends. Inductive content analysis was performed to generate themes regarding the barriers and facilitators for improving the care cascade in hypertension and diabetes management. The initial list of codes was grouped into categories and then themes against the supporting evidence. Throughout this process, subthemes and themes were discussed and refined within the research team.

Ethical Considerations

This review does not involve human subject information, primary data collection, or any form of experimentation on individuals.
Results

Characteristics of the Included Studies

Of the 1321 unique articles identified for the title and abstract screening, 222 were retrieved for full-text review. Of these, 128 were included in the analysis after excluding 94 articles for various reasons (Figure 1).

The 128 studies originated from 40 countries, with 42.2% (54/128) conducted in HICs (Figure 2). Of the 128 studies, 47 (36.7%) focused on hypertension care, 63 (49.2%) focused on diabetes care, and 18 (14.1%) reported on the care of both diseases. Most studies (104/128, 81.3%) employed quantitative methods. The majority were cross-sectional studies (70/128, 54.7%), followed by cohort studies (26/128, 20.3%). There were 24 (18.8%) trials evaluating interventions to promote retention in at least one cascade stage. Only 8 (6.3%) were implementation studies designed to systematically assess service delivery gaps and identify contextually appropriate solutions to address these bottlenecks. A total of 116 (90.6%) studies reported health receivers’ perspectives, and only 8 (6.3%) studies had health system perspectives. Most studies (83/128, 64.8%) reported client outcomes as primary outcomes, and they mainly focused on the measure of the effectiveness of disease control. Moreover, 16 (12.5%) studies reported service outcomes, and they mainly focused on the measure of satisfaction. Furthermore, 29 (22.7%) studies reported implementation outcomes, such as feasibility, cost, and adoption (Table 1). Detailed characteristics of the 128 included studies are summarized in Multimedia Appendix 2 [10,15-131].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.
Figure 2. Map of studies included in the scoping review by country (N=40).

Table 1. Summary of the characteristics of the included studies.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value (N=128), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Income level of countries</strong></td>
<td></td>
</tr>
<tr>
<td>High income</td>
<td>54 (42.2)</td>
</tr>
<tr>
<td>Upper-middle income</td>
<td>25 (19.5)</td>
</tr>
<tr>
<td>Lower-middle income</td>
<td>38 (29.7)</td>
</tr>
<tr>
<td>Low income</td>
<td>11 (8.6)</td>
</tr>
<tr>
<td><strong>Disease</strong></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>47 (36.7)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>63 (49.2)</td>
</tr>
<tr>
<td>Hypertension and diabetes</td>
<td>18 (14.1)</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td></td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>70 (54.7)</td>
</tr>
<tr>
<td>Cohort</td>
<td>26 (20.3)</td>
</tr>
<tr>
<td>Interventional</td>
<td>24 (18.8)</td>
</tr>
<tr>
<td>Implementation</td>
<td>8 (6.2)</td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td></td>
</tr>
<tr>
<td>Health receivers</td>
<td>116 (90.6)</td>
</tr>
<tr>
<td>Health providers</td>
<td>4 (3.1)</td>
</tr>
<tr>
<td>Health receivers and providers</td>
<td>8 (6.3)</td>
</tr>
<tr>
<td><strong>Research methods</strong></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>104 (81.3)</td>
</tr>
<tr>
<td>Qualitative</td>
<td>11 (8.6)</td>
</tr>
<tr>
<td>Mixed</td>
<td>13 (10.1)</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td>29 (22.7)</td>
</tr>
<tr>
<td>Service</td>
<td>16 (12.5)</td>
</tr>
<tr>
<td>Client</td>
<td>83 (64.8)</td>
</tr>
</tbody>
</table>
Completion of the Hypertension and Diabetes Care Cascade

Only 3 studies documented all 7 cascade stages, with 2 of them focusing on hypertension management and 1 addressing both hypertension and diabetes [15,16,132]. They were all population-based cross-sectional surveys aimed to describe disease prevalence and quantify the unmet need for hypertension and diabetes care. The remaining studies included in our analysis examined specific stages of the care cascade. Among the studies focusing on hypertension, 13 highlighted increasing awareness and knowledge related to hypertension, 8 addressed the importance of screening through blood pressure measurements, 14 focused on the diagnosis of hypertension, 13 explored the linkage to care, 34 discussed the initiation of treatment, 16 emphasized medication adherence, and 26 explored blood pressure management and control. For diabetes care, 8 studies addressed the critical aspect of awareness, 16 concentrated on screening, 20 discussed the diagnosis of diabetes, 21 explored the linkage to care, 38 focused on treatment interventions, 23 examined medication adherence, and 37 investigated the factors impacting diabetes control. In addition, 18 studies adopted an integrated approach, encompassing care for both hypertension and diabetes. Among these studies, 6 addressed awareness, 1 addressed screening, 4 addressed diagnosis, 10 addressed linkage to care, 9 addressed treatment, 2 addressed medication adherence, and 10 addressed control (Table 2).

Table 2. Studies across the stages of the care cascade.

<table>
<thead>
<tr>
<th>Stage of the care cascade</th>
<th>Value, n (%)</th>
<th>Hypertension (n=47)</th>
<th>Diabetes (n=63)</th>
<th>Hypertension and diabetes (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness</td>
<td>Awareness</td>
<td>13 (27.7)</td>
<td>8 (12.7)</td>
<td>6 (33.3)</td>
</tr>
<tr>
<td>Screening</td>
<td>Screening</td>
<td>8 (17.0)</td>
<td>16 (25.4)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Diagnosis</td>
<td>14 (29.8)</td>
<td>20 (31.7)</td>
<td>4 (22.2)</td>
</tr>
<tr>
<td>Linkage to care</td>
<td>Linkage to care</td>
<td>13 (27.7)</td>
<td>21 (33.3)</td>
<td>10 (55.6)</td>
</tr>
<tr>
<td>Treatment</td>
<td>Treatment</td>
<td>34 (72.3)</td>
<td>38 (60.3)</td>
<td>9 (50.0)</td>
</tr>
<tr>
<td>Medication adherence</td>
<td>Medication adherence</td>
<td>16 (34.0)</td>
<td>23 (36.5)</td>
<td>2 (11.1)</td>
</tr>
<tr>
<td>Control</td>
<td>Control</td>
<td>26 (55.3)</td>
<td>37 (58.7)</td>
<td>10 (55.6)</td>
</tr>
</tbody>
</table>

Interventions of the Hypertension and Diabetes Care Cascade

**Awareness**

Various interventions were identified to enhance the knowledge of disease prevention. Health education programs for hypertension and diabetes were emphasized as continuous efforts to support ongoing management and care [17]. The provision of comprehensive education was achieved through training classes and consulting at nutrition-based shared medical appointments [17-19]. Automated outreach call services with the integration of electronic health records emerged as effective approaches [20]. Out-of-hospital continuous nursing interventions and community awareness campaigns were used to augment the awareness [21]. Dissemination of awareness campaign information occurred through various channels, including the internet, public awareness events, and targeted home visits [22].

**Screening**

In the pursuit of enhancing the screening process for hypertension and diabetes, a diverse array of interventions has emerged. The Sustainable East Africa Research in Community Health (SEARCH) study implemented a community health campaign that offered universal adult screening, rendering screening services widely accessible [23]. Moreover, innovative approaches like home-based screening interventions empowered individuals by providing convenience and ease of access to early detection services [24]. Early diabetes detection was prioritized through specialized medical check-ups, facilitating timely intervention [22]. The Integrated Tracking, Referral, and Electronic Decision Support and Care Coordination (I-TREC) program incorporated cutting-edge technologies, including electronic case record forms and clinical decision support systems, streamlining patient information and offering guideline-based screening. Enhanced training for health care providers in NCD management and lifestyle interventions further fortified the screening process [25]. Lastly, efforts were made to strengthen health education and outreach services, particularly targeting individuals without symptoms, to foster a proactive approach to screening [26].

**Diagnosis**

Effective interventions have been deployed to enhance the diagnosis of hypertension and diabetes. Continuous and coordinated care among multi-level health institutions was emphasized to enable timely diagnosis and consistent follow-up for hypertension and diabetes. Telephone peer coaching provided personalized support through weekly calls, aiding in timely diagnosis and empowering patients to engage in self-care [27]. Patient-centered integrated care with advanced technologies, such as electronic case record forms and clinical decision support systems, streamlined patient information and referrals to deliver tailored guideline-based care. Enhanced training for primary health care providers further strengthened timely diagnosis among patients [25].

**Linkage to Care**

Follow-up within 6 weeks at NCD clinics for participants with hypertension, coupled with the use of diabetes self-management record sheets and telephone reinforcement, has shown positive
outcomes [133]. Additionally, 8-week training classes encompassing diverse self-care aspects have demonstrated effectiveness [18]. Mobile health applications [28], shared medical appointments [29], telephone peer coaching [27], and regular general practitioner contact [30] have proven to be successful strategies for ensuring a smooth connection to care. Furthermore, providing training in communication skills and self-care education to health providers, along with reduced workload and increased availability of competent diabetes specialist nurses, has contributed to enhancing the linkage to care [31]. Educational group programs, decision support tools, and feedback reports for primary care professionals further reinforced the process [32]. Institution-level continuity of ambulatory care [33], standardized “self-care” programs [22], and patient-held health records [34] have also played pivotal roles in promoting seamless linkage to essential health care services among patients diagnosed with hypertension and diabetes.

**Treatment**

Various interventions have been explored to improve the treatment process for patients with hypertension and diabetes. Lifestyle interventions, including physical activity promotion and heart-healthy diets, have shown promise in improving treatment outcomes [35]. Collaborative care models involving pharmacists and physicians demonstrated positive effects on medication therapy management and overall patient care [36]. Self-monitoring of blood pressure is vital for facilitating appropriate treatment [37]. Additionally, interventions targeting medication affordability [38] and continuity of care [39-41,134] play crucial roles in ensuring optimal treatment adherence. Telehealth and digital interventions, such as continuous remote care and mobile health applications, are being explored for improved treatment accessibility and engagement [28,135]. Integrated care models, employing multidisciplinary teams and decision support tools, have yielded promising outcomes in coordinating comprehensive patient care [20,136]. Targeted education for patients and health care providers can effectively enhance communication and self-care skills [31,32]. Moreover, financial incentive programs, like pay-for-performance schemes, have encouraged optimal health care delivery and reimbursement [42].

**Medication Adherence**

A range of interventions has been investigated to optimize medication adherence in patients with hypertension and diabetes. Community-based interventions with patient education, recall services, and reduced out-of-pocket payments have shown promise in promoting adherence [43]. Self-measured blood pressure monitoring and chronic disease management programs in primary care settings facilitated continuous and comprehensive patient care [38,44]. Telephone peer coaching [27], regular general practitioner contact [30], and continuity of care initiatives [39,40,45,46,134] have also demonstrated positive effects on medication adherence. Collaborative care models, which involve patient-centered coordinated care, referral systems, and diabetes education, have yielded favorable results [47]. Additionally, interventions, such as medication co-payment schemes, enhanced counseling, and training for health care providers in communication skills, have reinforced medication adherence efforts [31,48]. Patient and provider engagement programs, along with pay-for-performance initiatives, have also incentivized optimal medication adherence [32,42]. Integrating pharmacists into multidisciplinary care teams has enhanced medication management and adherence [136].

**Control**

Interventions, including multidisciplinary collaboration, patient education, and technology integration, were adopted to enhance hypertension and diabetes control. The integration of pharmacists into care teams and the transition to specialized diabetes physicians can optimize disease management [49,50,136]. For instance, pharmacist-physician collaborative practice models have shown promise through features like shared medical records, defined interprofessional roles, frequent follow-ups, and collaborative practice agreements [50]. Structured educational programs, both for patients and primary care professionals, offer essential knowledge and support, such as tailored SMS text message communication and telephone peer coaching [27,51]. Patient health records and electronic decision support were used to improve the continuity of care and enable tailored interventions [34]. Additionally, integrated interventions like the EMPOWER-PAR program, grounded in the Chronic Care Model, made contributions to advancing disease control, even in the face of challenges related to health care system constraints [137].

**Barriers and Facilitators of Completing Different Stages of the Care Cascade**

In the completion of the hypertension and diabetes care cascade, several barriers and facilitators were identified, encompassing patient-level and system-level factors (Table 3). Patient-level barriers included factors like low socioeconomic status, unhealthy lifestyle choices, and limited health literacy, hindering effective management. Misconceptions about disease and treatment, high treatment costs, and fear of diagnosis also impeded the care progress. At the system level, inadequate resources, heavy workloads, limited capacity in primary care, and a fragmented health system were identified as significant obstacles to effective care.
<table>
<thead>
<tr>
<th>Stage and country type</th>
<th>Facilitators</th>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Awareness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HICs</td>
<td>• Patient-level: Young age [20,52,132], female sex [52,53], and high socioeconomic status [54,132]</td>
<td>• Patient-level: Male sex [55], low income [55,56], poor health literacy [57], minority group [58], and living in resourcelimited areas [55]</td>
</tr>
<tr>
<td>LMICs</td>
<td>• Patient-level: Aged [53,59], female sex [10,16,59,138], high socioeconomic status [10,16,60], overweight and obesity [59,61], unhealthy lifestyle [59,60], and multimorbidity [61]</td>
<td>• Patient-level: Male sex [60], never married [10,60], Hispanic adults [58], and Asian adults [58]</td>
</tr>
<tr>
<td><strong>Screening</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HICs</td>
<td>• Patient-level: Female sex [62-64], unemployed [61], and single [61]</td>
<td>• Patient-level: Low socioeconomic group [15], multimorbidity [65], and minority group [62]</td>
</tr>
<tr>
<td>LMICs</td>
<td>• Patient-level: Under the age of 25 years [66] and high socioeconomic group [53,60]; Use of antenatal care services [67]</td>
<td>• Patient-level: Low socioeconomic group [16,56], unhealthy lifestyle [66,139], and multimorbidity [65]; Not feeling at risk of hypertension [67]; Not aware of screening services [67]; Low ability to pay for health care [67]; Preference for traditional healers [67]; Perceiving hypertension as a normalized condition [67]</td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HICs</td>
<td>• Patient-level: High socioeconomic group [56]</td>
<td>• Patient-level: Male sex [68,132], living alone [132], multimorbidity [69], unhealthy lifestyle [132], and living in resourcelimited areas [70]</td>
</tr>
<tr>
<td>LMICs</td>
<td>• Patient-level: Aged [60,66,71-73], overweight or obesity [63,66,71,74], and presence of other comorbidities [71]</td>
<td>• Patient-level: Characteristics of individuals: male sex [64,69,73], low socioeconomic group [72,75], and unhealthy lifestyle [63]; Lack of understanding regarding the importance of following a referral after a positive screening result [76]; Unaffordable health care services [67]; Fear of diagnosis, refusal to accept illness, and noncompliance with referrals [67]; Influence of culture and values, including gender norms [69]; Conflicting time with health facility opening hours [69]; System-level: Shortage of physicians [75]; Shortage of health facilities [69,76]; Lack of diagnostic equipment and testing capabilities [26,76]</td>
</tr>
<tr>
<td><strong>Linkage to care</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HICs</td>
<td>• Patient-level: High socioeconomic status [77] and female sex [77]; Regular clinic visit (due to smoke) [65]; Involvement in other health programs [77]</td>
<td>• Patient-level: Male sex [16,68,77], no health insurance [79], and low education level [64,134]; Presence of other diseases that affect physical activity [134]; Language barriers [31]; System-level: Heavy workload affecting patient care [31,77]; Inadequate collaboration among health care team members [31]; Providers’ frustration and aggressive attitudes toward patients [31]</td>
</tr>
<tr>
<td>LMICs</td>
<td>• Patient-level: High awareness [80,81]; Social support or involvement of patients’ relatives [81-83]; Context-specific diabetes education and educational materials [84]</td>
<td>• Patient-level: Without health insurance [80]; Misconceptions about medications [85]; Cultural beliefs [85]; System-level: Absence of guidelines for hypertension and diabetes management [38,44]; Insufficient essential resources and infrastructure [84]</td>
</tr>
</tbody>
</table>

https://publichealth.jmir.org/2024/1/e51802
<table>
<thead>
<tr>
<th>Stage and country type</th>
<th>Facilitators</th>
<th>Barriers</th>
</tr>
</thead>
</table>
| **HICs**               | • Patient-level: Young age [20,26,52,72], female sex [10,52,62], White ethnicity [52], and high health literacy [80]; Medicare beneficiary [20]  
  • System-level: Presence of a chronic disease management program [44,78]; Home delivery of medication [86]; Good doctor-patient relationship [87,140] | • Patient-level: Male sex [68,80], no medical insurance [20,80], low education level [80], unhealthy lifestyle [88], multimorbidity [10,65,80], and language barriers [31]; Psychological fear of treatment [140]  
  • System-level: Health care mistreatment attributed to ethnic discrimination [89]; Heavy workload of health providers [34,90,140]; Lack of collaborative strategies among health care teams [140] |
| **LMICs**              | • Patient-level: High education [55,73], high income [16,55,56,60], and overweight and obesity [59,60,91]; Well-designed education and educational materials [84]  
  • System-level: Physician density [75]; Doctors’ interpersonal behaviors and technical competence [92]; Well-trained health workforce [84] | • Patient-level: Individual characteristics: never married [60], occupation [66], poor socioeconomic status [15,66,83], poor comprehension [83], unhealthy lifestyle [66], and living in resource-limited areas [16,53,55,64,70]; Misconceptions, lack of confidence, and fear about medications [85,141]; High treatment costs and low ability to pay for medication or transport [26,67]; Poor understanding of asymptomatic conditions requiring treatment [67]; Low risk awareness of non-treatment consequences [67]; Wrong understanding of the disease and its therapy [93]; Lack of social support from peers, family, and the community [93]; High time cost of seeking care [93]  
  • System-level: Poor monitoring and lack of a patient follow-up protocol [69,94]; Ineffective medication and physician inertia [69,94]; Ambiguous and inappropriate clinical guidelines in under-resourced areas [93]; Shortage of human resources and equipment for blood pressure monitoring [93]; Limited knowledge and understanding among health care workers [76]; Lack of essential clinical facilities and adequate training of health care workers [76,84]; Absence of organized diabetes services within health care facilities [83]; Rarely receiving feedback on patient management from higher-level facilities [83] |

**Medication adherence**

| **HICs**               | • Patient-level: Female sex [80,95], high income [56,95], and high level of hypertension and diabetes knowledge [95,96]; Less negative general beliefs about medications and few concerns about medications [95] | • Patient-level: Lower socioeconomic group [95]; Not confident about community pharmacists [141]; Fear about medications [85]; Cultural beliefs influencing management [31]; Multimorbidity [80]; Lack of knowledge leading to misconceptions about disease management [141]  
  • System-level: Low primary care visits [80]; Ethnic discrimination in health care settings [89]; Heavy workload [31,90]; Lack of a teamwork approach [31]; Insufficient availability of essential medicines [38]; Ambiguous and inappropriate clinical guidelines [93] |
| **LMICs**              | • Patient-level: Living in urban areas [16,45,53] | • Patient-level: High cost of medication [69]; Personal and cultural beliefs [69]; Wrong understanding of a disease and its therapy among patients [93]; Lack of support from peers, family, providers, and the community [93] |

**Control**
<table>
<thead>
<tr>
<th>Stage and country type</th>
<th>Facilitators</th>
<th>Barriers</th>
</tr>
</thead>
</table>
| HICs                  | • Patient-level: Young age [52], female sex [95], high income [74], and being of White ethnicity [52,58]; High level of hypertension and diabetes knowledge [95]; Partner involvement in care [82]; Better self-perceived health status [95]; Healthy lifestyle practices: regular exercise and weight management [97]  
  • System-level: Trust between physicians and patients [98] | • Patient-level: Characteristics of individuals: male sex [56], ethnic minority [20], and low health literacy [96]; Lack of access to medical care and medications [50]; Using nonoptimized doses of antihypertensive medications [65]; Experiencing adverse events associated with medications [65] |
| LMICs                | • Patient-level: High income [15,54,60,74], older age [19,72], marriage [10,54], fewer complications [65], and health insurance [19,39]; Healthy lifestyle: adopting dietary modifications or engaging in regular exercise [54,99]; Receiving long prescribed medications for hypertension and diabetes [71]; Belief in treatment efficacy and having family support [67]; Timely monitoring of blood pressure and blood glucose control [71]  
  • System-level: Adequate medications [84]; High physician density [75] | • Patient-level: Age ≥75 years [74], male sex [15,55,59,68,100], overweight [59,66,100], low income [55,60,67], specific occupations [66], and low education level [55,68]; Coexisting chronic conditions [68,97]; Living in rural or resource-limited areas [55,60]; Unhealthy lifestyle: smoking and alcohol consumption [66,97,100]; Lack of family/social support [57]; Limited awareness of the lifelong nature of the condition [67]; Complexity of the intervention [68]; Insufficient patient education about the importance of clinical management [67]; Poor communication of treatment monitoring results [67]  
  • System-level: Long waiting times at clinics [67]; Negative staff attitudes toward patients [67]; Weak monitoring schedules [67]; Lack of medical resources [84] |

^aHIC: high-income country.  
^bLMIC: low- and middle-income country.

Conversely, various patient-level facilitators positively impacted the cascade. At the patient level, characteristics like high socioeconomic status, positive health behaviors, and strong belief in treatment efficacy played vital roles. Furthermore, timely monitoring of blood pressure and glucose levels, engagement in health programs, and partner involvement were found to be associated with improved outcomes. System-level facilitators included a well-trained health workforce, existing chronic disease management programs, and improved access to medications.

Notably, certain barriers and facilitators were context-specific, with diverse prominence in HICs and LMICs. For instance, lack of understanding and misconceptions were more prevalent in LMICs, while the influence of cultural beliefs and minority status was more pronounced in HICs. Physician density and adequate resources were often noted as facilitators in HICs, while social support and tailored diabetes education were emphasized in LMICs.

**Discussion**

This scoping review identified a substantial body of literature investigating the hypertension and diabetes care cascade in both HICs and LMICs. While most studies provided descriptive snapshots of each cascade stage, only a limited number of studies applied implementation cascade analysis to explore the barriers and facilitators of patient retention. Furthermore, there was a paucity of studies evaluating the effects of interventions to bridge gaps between cascade stages. In addition to analyzing the characteristics of the included studies, this scoping review comprehensively summarized key interventions, facilitators, and barriers associated with completing cascade stages. These findings provide critical insights into the existing evidence on hypertension and diabetes management, offering valuable directions for enhancing health care delivery for these chronic conditions.

The results of this scoping review have revealed a notable gap in the existing literature concerning the entire continuum of all stages in the hypertension and diabetes care cascade. The majority of studies predominantly focused on treatment and control for both hypertension and diabetes care. There was a relative paucity of studies examining the stages before treatment initiation despite evidence suggesting that over 50% of patients with hypertension and diabetes who could benefit from treatment never start medication [3,72]. These pretreatment losses accounted for a much greater reduction in effective care than nonadherence to medication [101]. Modeling studies showed that treatment losses earlier on can result in a greater overall reduction in the public health benefit of hypertension management [142,143]. Potential gaps exist in identifying problems and developing strategies to improve awareness, screening, and diagnosis of the 2 diseases. Based on microsimulation modeling, it is estimated that scaling up diagnosis, treatment, and control of diabetes to achieve a hypothetical 80% target for each component of the care cascade would be highly cost-effective [143,144]. Regarding interventions to improve retention across cascade stages, the review emphasizes the importance of awareness campaigns and health education programs to improve patient retention in care and medication adherence. Moreover, interventions targeting the health system (ie, multidiscipline collaborative care, training for primary health care providers, and increasing access to medications) showed promise in improving diagnosis and treatment outcomes. Other innovations in hypertension and diabetes service delivery have been developed and could further enhance quality, but they require further study and proof of effectiveness at scale. Examples include electronic case record–based clinical decision support systems and telephone
Acknowledgments

The study was supported by the nonprofit Central Research Institute Fund of the Chinese Academy of Medical Sciences (grant number: 2021-RC330-004). RS (shaoruitai@cams.cn) and XY are cocorresponding authors for this study.

https://publichealth.jmir.org/2024/1/e51802

JMIR Public Health Surveill 2024 | vol. 10 | e51802 | p.319

(page number not for citation purposes)
Data Availability
The data sets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
XY and RS conceived and designed the study. XY developed the search strategy. JW and FT ran the search. JW, FT, ZW, YY, and XY extracted the data. XY verified the data extraction. JW, FT, and XY analyzed and interpreted the data. JW wrote the first draft of the manuscript with XY. All authors contributed to the writing of the manuscript. All authors critically revised the manuscript and approved the final version.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 15 KB - publichealth_v10i1e51802_app1.docx ]

Multimedia Appendix 2
Details of the included studies.
[DOCX File, 54 KB - publichealth_v10i1e51802_app2.docx ]

Multimedia Appendix 3
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File, 71 KB - publichealth_v10i1e51802_app3.docx ]

References


113. Pan CC, Kung PT, Chiu LT, Liao YP, Tsai WC. Patients with diabetes in pay-for-performance programs have better physician continuity of care and survival. Am J Manag Care 2017 Feb 01;23(2):e57-e66 [FREE Full text] [Medline: 28245660]


121. Yan JW, Liu S, Van Aarsen K, Columbus MP, Spaic T. Do adult patients with type 1 or 2 diabetes who present to the emergency department with hyperglycemia have improved outcomes if they have access to specialized diabetes care? Can J Diabetes 2021 Feb;45(1):59-63. [doi: 10.1016/j.cjd.2020.05.011] [Medline: 32847767]


Review

Optimization of Screening Strategies for COVID-19: Scoping Review

Yuanhua Liu1,2*, PhD; Yun Yin1,2*, MS; Michael P Ward3, PhD; Ke Li1,2, MS; Yue Chen4, PhD, Prof Dr; Mengwei Duan2, MS; Paulina P Y Wong5, PhD; Jie Hong1,2, MS; Jiaqi Huang1,2, MS; Jin Shi1,2, MS; Xuan Zhou5, MS; Xi Chen1,2, MS; Jiayao Xu1,2, MS; Rui Yuan1,2, MS; Lingcai Kong5*, PhD; Zhijie Zhang1,2, PhD, Prof Dr

1Department of Epidemiology and Health Statistics, School of Public Health, Fudan University, Shanghai, China
2Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China
3Sydney School of Veterinary Science, The University of Sydney, NSW, Australia
4School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
5Department of Mathematics and Physics, North China Electric Power University, Baoding, China
6Science Unit, Lingnan University, Hong Kong, China
* these authors contributed equally

Corresponding Author:
Zhijie Zhang, PhD, Prof Dr
Department of Epidemiology and Health Statistics
School of Public Health
Fudan University
No.130, Dong’An Road, Xuhui District
Shanghai, 200032
China
Phone: 86 021 54237410
Email: epistat@gmail.com

Abstract

Background: COVID-19 screening is an effective nonpharmaceutical intervention for identifying infected individuals and interrupting viral transmission. However, questions have been raised regarding its effectiveness in controlling the spread of novel variants and its high socioeconomic costs. Therefore, the optimization of COVID-19 screening strategies has attracted great attention.

Objective: This review aims to summarize the evidence and provide a reference basis for the optimization of screening strategies for the prevention and control of COVID-19.

Methods: We applied a methodological framework for scoping reviews and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist. We conducted a scoping review of the present publications on the optimization of COVID-19 screening strategies. We searched the PubMed, Web of Science, and Elsevier ScienceDirect databases for publications up to December 31, 2022. English publications related to screening and testing strategies for COVID-19 were included. A data-charting form, jointly developed by 2 reviewers, was used for data extraction according to the optimization directions of the screening strategies.

Results: A total of 2770 unique publications were retrieved from the database search, and 95 abstracts were retained for full-text review. There were 62 studies included in the final review. We summarized the results in 4 major aspects: the screening population (people at various risk conditions such as different regions and occupations; 12/62, 19%), the timing of screening (when the target population is tested before travel or during an outbreak; 12/62, 19%), the frequency of screening (appropriate frequencies for outbreak prevention, outbreak response, or community transmission control; 6/62, 10%), and the screening and detection procedure (the choice of individual or pooled detection and optimization of the pooling approach; 35/62, 56%).

Conclusions: This review reveals gaps in the optimization of COVID-19 screening strategies and suggests that a number of factors such as prevalence, screening accuracy, effective allocation of resources, and feasibility of strategies should be carefully considered in the development of future screening strategies.

(JMIR Public Health Surveill 2024;10:e44349) doi:10.2196/44349

https://publichealth.jmir.org/2024/1/e44349

JMIR Public Health Surveill 2024 | vol. 10 | e44349 | p.329
(page number not for citation purposes)
KEYWORDS
COVID-19; screening strategy; optimization; polymerase chain reaction test; antigen test

Introduction
SARS-CoV-2 has resulted in >500 million cases of COVID-19 worldwide, causing >6 million deaths, and continues to threaten human health [1,2]. The Omicron variant has become the most dominant variant in the current pandemic, and its insidious transmission makes community spread a big challenge [3,4]. For example, the large-scale Omicron outbreak in Shanghai between March and June 2022 resulted in >600,000 infections, and approximately 90% of them were asymptomatic [5]. The global spread of the pandemic has had a tremendous impact on the health of susceptible populations. In China, where the older adult population exceeds 26.4 million [6], the vulnerability of the older adults is heightened, despite the overall low mortality rate of the COVID-19 Omicron subvariant [5,7].

Symptom surveillance and voluntary nucleic acid testing were ineffective in response to the insidious transmission of the Omicron variant. Proactive screening of COVID-19 is essential to identify asymptomatic infections and break the transmission chain in a timely manner [8]. Under the dynamic zero policy [9], residents in epidemic areas received nucleic acid screening tests regularly, and the scope and frequency of the screening were dynamically adjusted in accordance with the epidemic trend. Antigen testing is used as a supplement to nucleic acid testing to improve the screening efficiency [10]. COVID-19 screening has become a basic prevention and control measure in countries worldwide, although the scope of screening varies [11]. Different strategies have been developed for the general population [10], international travelers [12], and high-risk populations [13].

COVID-19 screening is the rapid identification of potentially infected individuals by testing a massive population to take appropriate measures, such as isolating the patient, providing treatment, and conducting contact tracing. COVID-19 screening primarily involves nucleic acid and antigen tests. Nucleic acid screening relies on polymerase chain reaction (PCR) testing techniques and is the gold standard for the confirmation of infection [14,15], and the test includes 4 steps: sample collection, preservation, transportation, and testing [10]. The entire process is labor and resource intensive, and each step is important for test accuracy. Antigen screening uses different detection techniques, such as colloidal gold immunochromatography, latex methods, and fluorescence immunochromatography, allowing for quick and easy self-testing. However, antigen testing is less accurate and is often used as a supplement to nucleic acid testing [16,17].

To develop a COVID-19 screening strategy, the target population and the timing and frequency of screening should be considered. For mass nucleic acid testing, a pooled sample testing approach is often used to reduce costs and improve detection efficiency. Factors that influence the cost-effectiveness of outbreak control should also be considered when optimizing screening strategies. In this scoping review of published research, we aimed to summarize the evidence and provide a reference basis for the optimization of screening strategies for the prevention and control of COVID-19.

Methods
We followed the methodological framework proposed by Arksey and O’Malley [18] and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [19]. The PRISMA-ScR checklist is available in Multimedia Appendix 1 [19].

Search Strategy
The search strategy adopted in this review was ((test* OR screen* OR detect*) AND (polymerase chain reaction OR PCR OR nucleic acid OR antigen) AND (COVID* OR SARS-CoV* OR Omicron OR Severe Acute Respiratory Syndrome Coronavirus 2)) NOT (diagno* OR clinic* OR Gene* OR cell OR protein OR laboratory OR patholog*).

The words related to “testing or screening” were limited to the title field. The words related to “testing methods,” “COVID-19,” and “clinical diagnosis or laboratory process technology or pathology” were also limited to the title or abstract fields. PubMed, Web of Science, and Elsevier ScienceDirect were searched for studies published as of December 31, 2022. The reference lists of eligible studies were reviewed to identify additional studies.

Selection Criteria
The inclusion criteria were as follows: (1) the literature language was English, (2) the disease studied in the publication was COVID-19, and (3) the research articles were related to screening and testing strategies for COVID-19 infection.

The exclusion criteria were as follows: (1) articles not related to COVID-19, (2) duplicates, (3) articles that did not involve screening detection strategies, (4) clinical diagnosis or pathological research articles, (5) technical articles on laboratory testing or testing reagents, (6) environmental detection research articles, and (7) basic theoretical articles on COVID-19 detection techniques. YL and YY screened the literature by reviewing the titles and abstracts. The full-text review was performed by JS, J Hong, KL, and MD, and then, the filtered document was checked again by 1 of these coauthors. Any discrepancies were discussed by YL and YY (Textbox 1).
Textbox 1. Inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Peer review: Peer-reviewed literature</td>
</tr>
<tr>
<td>• Article type: Original articles</td>
</tr>
<tr>
<td>• Language: English</td>
</tr>
<tr>
<td>• Disease: COVID-19</td>
</tr>
<tr>
<td>• Content: Screening and testing strategies for COVID-19 infection</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Peer review: Literature not peer reviewed</td>
</tr>
<tr>
<td>• Article type: Reviews, meeting articles, comments, and notes</td>
</tr>
<tr>
<td>• Language: Non-English</td>
</tr>
<tr>
<td>• Disease: Diseases other than COVID-19</td>
</tr>
<tr>
<td>• Content: Clinical diagnosis or pathological research; technical articles on laboratory testing or testing reagents; environmental detection research; basic theoretical articles on COVID-19 detection techniques</td>
</tr>
</tbody>
</table>

Data Abstraction
ZZ, YL, and YY determined which variables to extract, and the latter 2 developed and tested the data-charting form using Microsoft Excel. We abstracted data on the last name of the first author, research design, research population, optimization design, testing method, screening strategy, evaluation index, and recommendation. All authors participated in the data abstraction and reconfirmation of the abstraction. YL and YY charted the data, grouped the studies according to the optimization directions of the screening strategies, and summarized the findings.

Results
Overview of Included Studies
A total of 4290 publications were found by the searches conducted, of which 1536 were duplicates, and the titles of the remaining 2770 publications were screened for relevance (Figure 1). Subsequently, 476 abstracts were reviewed, and 95 publications received a full-text review. Finally, 62 publications were included in the synthesis (the data-charting form is available in Multimedia Appendix 2 [20-46]).

The study populations of the publications were the general population (36/62, 58%), travelers or immigrants (10/62, 16%), people in an organization (8/62, 12%, including workers, health care persons, and students), infected people (3/62, 5%), contacts or suspects (3/62, 5%), vaccinated population (1/62, 2%), and people at gathering activities (1/62, 2%). The optimization designs involved screening the population (12/62, 19%); timing (12/62, 19%); frequency (6/62, 10%); and testing procedure (35/62, 56%), including scenarios for adopting a pooling strategy (10/62, 16%), pool size (19/62, 31%), and pooling approach (18/62, 29%), as shown in Figure 2.
Figure 1. Flow diagram of the search and study selection process following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews.

4290 records identified through database screening:
- 1134 PubMed
- 1245 Elsevier ScienceDirect
- 1911 Web of Science

16 additional records identified through other sources

4306 records identified

1536 records left after duplicates were removed

2770 titles screened

2294 records excluded:
- 85 computed tomography or x-ray test studies
- 678 detection technology studies
- 70 pathogen or pathology studies
- 13 storage and transportation studies
- 77 environmental studies
- 142 observational studies
- 649 comparison or evaluation of detection technology studies
- 11 screening or detecting reviews
- 197 clinical or diagnosis studies
- 57 other disease studies
- 80 commentary and editorials
- 20 articles in other language
- 187 irrelevant research topic
- 28 vaccine studies

476 abstracts screened

381 records excluded
- 5 pathogen or pathology studies
- 19 technical methods and processes
- 81 evaluation of detection technology
- 26 studies of economic, cost, or resource
- 21 clinical studies
- 161 screening effectiveness evaluation
- 18 screening or detecting reviews
- 50 commentary and corrigendum

95 full-text articles assessed for eligibility

32 full-text articles excluded
- 12 screening effectiveness evaluation
- 13 detecting effectiveness evaluation
- 6 detecting technology studies
- 2 literature without peer review

62 total studies included in the scoping review
Optimization of Screening Population

**Screening of People With Different Infection Risks**

Previous studies tried to optimize screening strategies according to infection risk levels indicated by infection probability, contact with an infected person, probability of exposure, or presence of COVID-19 symptoms (Table 1). Du et al [20] simulated COVID-19 testing in populations with different severities of clinical symptoms and age groups and compared infection rates and false-positive rates among groups with different strategies. They found that in settings of high infection rate and limited testing capacity, a greater proportion of tests should be allocated to symptomatic individuals. Conversely, in a surveillance scenario characterized by a low infection rate and sufficient testing capacity, the optimal strategy, which involved directing a higher allocation of tests to people aged >50 years, only required 51.5% of available tests [20]. Likewise, the research by Han et al [21] supported that symptomatic testing at health care facilities was more beneficial than asymptomatic testing in the community, until most symptomatic individuals had been tested. It was also suggested that conducting additional tests to screen for asymptomatic infections among household members yielded the greatest benefit after fulfilling the demand for symptomatic testing [21]. Berestizhevsky et al [22] compared total morbidity, testing efficiency, and infection probability under screening strategies such as no testing and no isolation, symptom monitoring combined with random sampling testing, and symptom monitoring combined with “greedy” testing. Testing and quarantine among symptomatic populations using optimized sampling, which is based on community graphs and population risk factors, results in a 30% to 50% reduction in overall prevalence [22]. Kumar et al [23] compared the effectiveness and cost of outbreak control by weekly reverse transcription-PCR (RT-PCR) testing in 100% antigen-negative vaccinated individuals in a high-income country (the United States) and in a low-income country (India). They suggested that in regions with resource-limited vaccination strategies, high-frequency testing is still needed to minimize subsequent outbreaks [23].

**Figure 2.** The number of publications on screening population (N=62).
Table 1. Research on optimization of COVID-19 screening strategy.

<table>
<thead>
<tr>
<th>Research population</th>
<th>Optimization directions</th>
<th>Testing methods</th>
<th>Strategy</th>
<th>Evaluation index</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population</td>
<td>Screening population</td>
<td>PCR(^a) and RAT(^b)</td>
<td>• Optimal testing strategy: allocation of PCR and RAT to different age groups and individuals with varying symptoms while ensuring that all severe patients are tested and total expenditure remains within the budget</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Risk-based strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Symptom-based strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Severe-only strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Universal random testing strategy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCR(^a) and RAT(^b)</td>
<td>• Proportion of infections averted relative to the no-testing baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Number of tests available per 100,000 persons per day</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Number of additional infections averted for every 100 more tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Reduction of days when (R_t&gt;1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Proportion of infections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General population</td>
<td>Screening population</td>
<td>RAT</td>
<td>• Strategy 1: symptomatic testing at health care facilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Strategy 2: asymptomatic testing in the community setting (households, schools, formal workplaces, or religious gatherings) with different distribution: (1) even distribution to as many entities as possible once per week and (2) concentrated distribution to test all individuals in selected entities twice a week who will continue to get tested throughout the epidemic. Without or without quarantine of household members.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• 85% of weekly tests for strategy 2 and the rest for strategy 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• All weekly tests for strategy 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Tests are first used for strategy 1, and any remaining tests are used for strategy 2 next week</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• No testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General population</td>
<td>Screening population</td>
<td>PCR</td>
<td>• No testing and quarantining</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Quarantine people in the state &quot;contagious-symptomatic&quot; only</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Symptom-based plus random or greedy sampling and quarantining of positive people</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Symptom-based plus sampling based on optimization of community graph and population risk factors and quarantining of positive people</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Total and peak morbidity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Personal and global quarantine efficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Number of human-days in different states</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaccinated popula-</td>
<td>Screening population</td>
<td>PCR and RAT</td>
<td>• RAT followed by PCR testing in 100% of the population weekly in the United States and India</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tions</td>
<td></td>
<td></td>
<td>• (N_{infected}^k)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• (N_{cases}^l)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Number of the hospitalized, dead, and recovered</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Cost</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(N_{T-pos}^c\), \(N_{T-neg}^d\), \(N_{F-pos}^e\), \(N_{F-neg}^f\), \(N_{miss}^g\), \(N_{test-pos}^h\), and \(P_{pos}^i\)
<table>
<thead>
<tr>
<th>Research population</th>
<th>Optimization directions</th>
<th>Testing methods</th>
<th>Strategy</th>
<th>Evaluation index</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travelers</td>
<td>Screening population and screening timing</td>
<td>PCR</td>
<td>• PCR testing is not required for travel in all areas</td>
<td>• $N_{\text{infected}}$ and $N_{\text{cases}}$</td>
<td>Zhou et al [24], 2021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• PCR testing is required in all regions within 7, 5, and 3 days before travel</td>
<td>• $N_{\text{test}}^m$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• All those coming from high-risk areas (risk level 3-4) need to be tested within 3 days before travel</td>
<td>• The medical expenditure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• All those from medium- and high-risk areas (risk level 2-4) need to be tested within 3 days before travel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People at gathering activities</td>
<td></td>
<td>PCR</td>
<td>• None of the participants are quarantined before the event unless they are contact traced</td>
<td>• $N_{\text{cases}}$</td>
<td>Wong et al [25], 2022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• All participants traveling from overseas are quarantined for 14 days before the event</td>
<td>• $N_{\text{infected}}$, and percentage change of new and cumulative infections</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• All participants are quarantined before the event</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• All mainland participants are tested before the event</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• All participants are tested before the event</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General population</td>
<td></td>
<td>PCR and RAT</td>
<td>• Using PCR to test symptomatic patients in outpatient settings</td>
<td>• Reduction in cumulative symptomatic incidence</td>
<td>Baik et al [26], 2022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Community-based screening by RAT</td>
<td>• Number of unnecessary isolations</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Symptom-driven outpatient diagnostic testing by RAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workers</td>
<td></td>
<td>PCR</td>
<td>• No RT-PCR$^n$ testing of all workers</td>
<td>• $N_{\text{test}}$</td>
<td>Sandmann et al [27], 2020</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Testing the workers with COVID-19-like symptoms in isolation</td>
<td>• Change of days in quarantine per test</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Testing the workers without COVID-19-like symptoms but in household quarantine</td>
<td>• Change of workers spreading per test</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Testing all staff</td>
<td>• Testing accuracy</td>
<td></td>
</tr>
<tr>
<td>College students</td>
<td></td>
<td>PCR</td>
<td>• Testing the students with COVID-19-like symptoms RT-PCR testing for symptomatic students</td>
<td>• $N_{T_{\text{pos}}}$ and $N_{T_{\text{neg}}}$</td>
<td>Van Pelt et al [29], 2021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Testing for all students</td>
<td>• $N_{\text{test}}$ and $N_{\text{test}}$ per person</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Testing for all students + retesting symptomatic students with a negative first test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Testing for all students + retesting all students with a negative first test.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travelers</td>
<td></td>
<td>PCR</td>
<td>• $N_{\text{cases}}$</td>
<td></td>
<td>Chowell et al [32], 2021</td>
</tr>
<tr>
<td>Research population</td>
<td>Optimization directions</td>
<td>Testing methods</td>
<td>Strategy</td>
<td>Evaluation index</td>
<td>Study</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>----------</td>
<td>------------------</td>
<td>-------</td>
</tr>
</tbody>
</table>
| School students      | Screening population    | PCR and RAT     | • No measures  
• PCR testing of passengers before embarkation and social isolation  
• PCR testing of passengers before embarkation, daily testing on board, and social isolation  |  
• $R_t$  
• The proportion of cases reduction  
• $N_{cases}$  
• Student days lost | Colosi et al [31], 2022 |
| School students      | Screening population and screening timing | RAT | • Isolation of year group bubbles for 10 days  
• Twice weekly mass testing and isolation of year group bubbles for 10 days  
• Tested daily by RATs for 7 days from the day after identification of every case  
• Twice weekly mass testing and tested daily by RATs for 7 days from the day after identification of every case  
• Twice weekly mass testing  
• No testing or isolation |  
• School days miss per person  
• $N_{infected}$  
• Asymptomatic cases  
• $N_{test}$ per person  
• Prevalence  
• Absent persons | Leng et al [30], 2022 |
| Health care workers in the nursing home | Screening population | RAT | • Testing the health care person within the facility when there are ≥1 positive cases  
• Testing all asymptomatic health care persons in the absence of a known outbreak at predetermined intervals from 1 day to 7 days  |  
• Maximum preventable transmission | Zipfel et al [28], 2022 |
| Travelers            | Screening timing        | PCR and RAT     | • Testing and quarantine strategies for fully vaccinated travelers and unvaccinated travelers  
• A negative preboarding  
• A negative preboarding test and a negative arrival test  
• Negative preboarding, arrival, and quarantine exit tests  
• 14 days quarantine |  
• $R_t$  
• Adjusted breakthrough IR$^0$  
• Expected number of subsequent infections | Lee et al [37], 2022 |
| Travelers            | Screening timing        | PCR and RAT     | • The proportional reduction in transmission risk |  
• Johansson et al [33], 2021 |
<table>
<thead>
<tr>
<th>Research population</th>
<th>Optimization directions</th>
<th>Testing methods</th>
<th>Strategy</th>
<th>Evaluation index</th>
<th>Study</th>
</tr>
</thead>
</table>
| Travelers           |clinic; policy           |PCR, RAT       |● Isolate individuals before or during travel when symptoms appear  
● Test 3 days before travel  
● Test on the day of travel  
● Test 1 day before arrival  
● Test 3 days before arrival  
● The best time to conduct a second test after travel in the absence of postentry quarantine  
● Monitor and isolate symptoms before, during, and after travel  
● 14-day, 10-day, and 7-day isolation |● $R_t/R_0^p$  
● The proportion of infected traveler causes, the number of infected travelers that reaches 50 cases from 1 traveler |Steyn et al [34], 2022 |
| Travelers           | clinic; policy          |PCR and RAT    |● Predeparture testing  
● No test  
● PCR test 3 days before departure (on day 3)  
● RAT test 1 day before departure (on day 1)  
● Postarrival restrictions  
● Unlimited  
● PCR on days 0 and 4  
● Daily RAT for 5 days  
● Self-isolation for 5 days with PCR test on days 0 and 4  
● Self-isolation for 5 days and daily RAT test Government-managed isolation for 7 days and quarantine with PCR test on day 5  
● Government-managed isolation for 14 days and quarantine with PCR test on days 3 and 12 |● Cumulative infectious days $N_{infected}$  
● The ratio of $N_{F-positives}$ to $N_{F-positives}$ |Kiang et al [35], 2021 |
| Travelers           | clinic; policy          |PCR            |● Anterior nose PCR testing within 3 days before departure  
● PCR test within 3 days of departure, on the fifth day after arrival, and isolation for 5 days after arrival  
● RAT within 3 days of departure and on the fifth day after arrival  
● RAT on the day of departure, PCR test on day 5 after arrival, and isolation for 5 days after arrival  
● PCR test on arrival for 5 days |● IR and proportions of asymptomatic or presymptomatic cases $N_{miss}$  
● Cumulative probability and hazard rate of developing symptoms |Jen et al [36], 2022 |
| Travelers           | clinic; policy          |PCR and RAT    |● RT-PCR tests on arrival and quarantine for 5 days and a second PCR test at the end of quarantine  
● RT-PCR tests on arrival and quarantine for 5 days  
● Quarantine for 14 days without test |● $N_{miss}$ |Dickens et al [38], 2021 |
<table>
<thead>
<tr>
<th>Research population</th>
<th>Optimization directions</th>
<th>Testing methods</th>
<th>Strategy</th>
<th>Evaluation index</th>
<th>Study</th>
</tr>
</thead>
</table>
| Screening timing and screening frequency | - Isolation only  
- Pretest and inbound testing and isolation  
- Pretest, inbound testing, and outbound isolation and testing  
- Pretest, inbound testing and isolation, and daily testing until the exit  
- Pretest, inbound testing and isolation, and testing every 2 days  
- Pretest, inbound testing and isolation, and testing every 3 days  
- Pretesting, inbound testing, RAT every 3 days, and outbound PCR  
- Pretesting, inbound testing, isolation, and alternative testing at exit (a PCR test or a RAT) | | | | |
| Contacts | Screening frequency | RAT | - Isolation-based strategies: isolation duration of 0, 3, 5, 7, 10, and 14 days after exposure to the case; no testing during isolation or testing on the last day of the isolation period  
- Daily testing strategy: daily RAT of exposed individuals for 1, 3, 5, 7, 10, or 14 days, with no isolation required unless symptomatic or positive testing occurs | | Onward transmission potential from secondary cases | Quilty et al [39], 2021 |
| Travelers | Screening timing | PCR | - Isolation and no testing  
- Test at the beginning of isolation  
- Test at the end of isolation  
- Test at the beginning and end of isolation  
- Test during the isolating period. | | PQTR | Wells et al [40], 2020 |
| Contacts | Screening timing | PCR and RAT | - RAT at 2 best times (day 1 and day 3)  
- RAT at 3 best times (day 1 and day 3) and an additional test (PCR or RAT) | | The expected number of infection days | Foncea et al [41], 2022 |
| Infected people | Screening timing | PCR and RAT | - A RT-PCR test administered 1 or 2 days before the end of quarantine  
- Two RT-PCR tests administered on days 6 or 7 and then on day 8  
- A 6-day quarantine with tests on days 4, 5, and 6 using a highly sensitive RT-PCR test in cases where the shortest quarantine is needed  
- A RAT with test administered on day 9 or 10  
- A 9-day quarantine with tests on days 7 and 8 | | PQTR | Peng et al [42], 2021 |
<table>
<thead>
<tr>
<th>Research population</th>
<th>Optimization directions</th>
<th>Testing methods</th>
<th>Strategy</th>
<th>Evaluation index</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population</td>
<td>Screening frequency</td>
<td>PCR</td>
<td>● Citizens, family members, and recent contacts who test positive in the first round of PCR and those who do not participate must be quarantined for 10 days ● All regions with a positivity rate of ≥0.7% in the first round of testing should undergo a second round of mass testing</td>
<td>● The 7-day rolling average of new infections and Rt</td>
<td>Kahanec et al [43], 2021</td>
</tr>
<tr>
<td>General population</td>
<td>Screening frequency</td>
<td>PCR</td>
<td>● Community transmission: 2 tests per 1000 people (low incidence) ● Outbreak response: 4 tests per 1000 people (higher incidence)</td>
<td>● N&lt;sub&gt;test&lt;/sub&gt; ● The percentage of positive tests and the percentage of transmission reduction</td>
<td>Baker et al [45], 2021</td>
</tr>
<tr>
<td>Migrant workers</td>
<td>Screening frequency</td>
<td>PCR, RAT</td>
<td>● A PCR test every 2 weeks ● Weekly RAT</td>
<td>● Rt, N&lt;sub&gt;infected&lt;/sub&gt;</td>
<td>Koo et al [46], 2022</td>
</tr>
<tr>
<td>General population</td>
<td>Screening frequency</td>
<td>RAT</td>
<td>● Mass testing with a frequency of fortnightly, weekly, or tridaily testing begins on the 30th day ● Mass testing with a frequency of fortnightly, weekly, or tridaily testing begins on the peak of the outbreak</td>
<td>● N&lt;sub&gt;infected&lt;/sub&gt;, N&lt;sub&gt;cases&lt;/sub&gt;, and cases of intensive care unit R&lt;sub&gt;t&lt;/sub&gt;</td>
<td>Koo et al [44], 2022</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
</tr>
<tr>
<td>b</td>
</tr>
<tr>
<td>c</td>
</tr>
<tr>
<td>d</td>
</tr>
<tr>
<td>e</td>
</tr>
<tr>
<td>f</td>
</tr>
<tr>
<td>g</td>
</tr>
<tr>
<td>h</td>
</tr>
<tr>
<td>i</td>
</tr>
<tr>
<td>j</td>
</tr>
<tr>
<td>k</td>
</tr>
<tr>
<td>l</td>
</tr>
<tr>
<td>m</td>
</tr>
<tr>
<td>n</td>
</tr>
<tr>
<td>o</td>
</tr>
<tr>
<td>p</td>
</tr>
<tr>
<td>q</td>
</tr>
</tbody>
</table>

### Screening of People in Different Regions

Table 1 also shows the results of optimizing screening strategies in different regions such as medium-risk or high-risk regions and domestic or foreign regions. Zhou et al [24] compared the total number of infections and daily nucleic acid test loads among the screening strategies and found that the optimal strategy was to test people from medium- and high-risk areas using nucleic acid tests before they traveled. Wong et al [25] assessed infections based on attendance at an event by applying different strategies: no testing and quarantine in all areas, quarantine of attendees from foreign areas, testing of attendees from mainland China, and testing of all attendees at the event. They found that the strategies of quarantining the attendees from foreign areas and testing all (foreign or local) attendees were effective in controlling the number of infections, and they estimated that the total number of new infections was only 1% higher than the current local prevalence [25]. Baik et al [26] simulated the effectiveness of outbreak control in regions with limited resources, such as low- and middle-income countries.
They evaluated 3 screening strategies: using PCR to test symptomatic patients in outpatient settings, community-based screening with rapid antigen tests (RATs), and symptom-driven outpatient diagnostic testing using RATs. This showed that RATs would reduce transmission most efficiently when used to test symptomatic individuals in outpatient settings, and to avoid large numbers of unnecessary isolations, mass testing with lateral flow tests (LFTs) should be considered as a screening tool [26].

**Screening of Occupational Populations and Students**

COVID-19 screening strategies were assessed and optimized for some specific student and occupational populations (Table 1). Sandmann et al [27] compared the number of infections, number of tests, and duration of isolation between individuals with and without COVID-19 symptoms living in worker dormitories. Testing all the workers was associated with a reduced transmission of approximately 67 individuals per 1000 tests. However, screening workers with COVID-19-like symptoms in isolation only was associated with a higher risk of transmission in the workplace compared with the strategy of screening all workers [27]. Zipfel et al [28] simulated transmission in health care workers using 2 strategies: testing all when positive cases were detected and testing all periodically at predetermined intervals from 1 day to 7 days. The study showed that 38% of hospital-based transmission could be prevented if all staff were tested within 1 day when a positive case occurred, whereas 30% to 78% of transmission could be prevented if daily testing was performed [28].

Van Pelt et al [29] analyzed the number of RT-PCR tests required to identify each true-positive case and the true-positive rate in a student population using strategies such as symptom screening only, nucleic acid test screening of symptomatic students, and nucleic acid test screening of all students. Conducting RT-PCR testing for all students and retesting those with initially negative results can effectively identify cases with a correct rate of 86.9% [29]. Leng et al [30] further classified the student population by age and simulated the implementation of daily mass screening or screening after the occurrence of positive cases. The study found that mass antigen screening among students significantly reduced the likelihood of not attending class but often required a large number of tests [30]. Similarly, Colosi et al [31] analyzed strategies such as routine testing, symptom-based testing, screening, and quarantine when a positive case was detected and found that weekly testing of 75% of unvaccinated students would reduce the number of cases by 34% in primary schools and 36% in secondary schools.

**Optimization of Timing of Screening**

**Timing of Screening for Travelers**

To address the risk of outbreak caused by population movement, researchers compared the timing of screening of traveling people (eg, before, during, and after travel; Table 1). Chowell et al [32] simulated the impact of testing at different times before and after boarding a cruise ship on the cumulative number of infected cases. The study found that testing before boarding, daily testing on board, and maintaining social distancing significantly reduced the possibility of onboard transmission [32]. Johansson et al [33] found that PCR testing on the day of departure and isolation at the destination reduced the risk of transmission. Testing on the day of departure reduced the risk of transmission when traveling by 44% to 72% [33]. Steyn et al [34] assessed the transmission potential of SARS-CoV-2 and the number of infections by simulating PCR or LFT screening at different times, such as the day before departure, the day after arrival at the destination, and the fourth and fifth days after arrival. It was found that the combination of testing and home isolation could reduce the risk of community outbreaks to approximately 0.01, and using daily LFTs or a combination of LFTs and PCR testing could reduce the risk to levels comparable with or lower than those using PCR testing alone [34]. Kiang et al [35] evaluated the cumulative number of days of infection and the number of infections when travelers used PCR or antigen testing at timings such as 3 days before departure, the day of departure, and 5 days after arrival at the destination. The results indicated that nucleic acid testing 3 days before departure reduced the risk of infection during the travel, and the cumulative number of infection days was reduced from 8357 to 5401 days [35]. Jen et al [36] compared the morbidity, missed tests, and proportion of asymptomatic and presymptomatic individuals in travelers with different strategies such as PCR testing on arrival and quarantine for 5 days and quarantine for 14 days without testing. It was found that >82% of the cases would progress from the presymptomatic phase to the symptomatic phase during the 5-day quarantine period, and the quarantine time with 2 PCR tests depended on the risk, testing and quarantine strategy, and vaccination status of the country of departure [36]. On the basis of previous screening strategies for travelers, Lee et al [37] analyzed transmission potential, infection rates, and subsequent infections for strategies such as testing before traveling, after arrival, or at the end of quarantine and 14 days of quarantine without testing, with a consideration of vaccination factors. It was found that at an incidence rate of 0.4 and a time-dependent reproduction number of 16, testing with a sensitivity of ≥98% and specificity of ≥97% both before traveling and on arrival ensured lower expected transmission in vaccinated than unvaccinated individuals with a quarantine of 14 days [37]. For entry-exit pandemic prevention, Dickens et al [38] analyzed the number of unidentified infected persons using the strategies of isolation only; predetection combining entry testing and isolation; and predetection combining entry testing, isolation, and daily testing. The results showed that the risk of transmission was greatly reduced by adopting predetection, which combined entry testing and isolation. During the isolation period, if an RAT was performed every 3 days, only 3% of the infected individuals were unidentified at 7 days and 0.7% at 14 days [38].

**Screening Timing in Response to an Outbreak**

When an outbreak occurs, the appropriate timing of screening facilitates the identification of infected individuals and helps control the spread of the disease. Several studies have explored the impact of different screening timings on the spread of the pandemic (Table 1). Quilty et al [39] compared the impact of daily antigen testing of close contacts over 1, 3, 5, 7, 10, or 14 days on the spread of the epidemic. It showed that quarantining for 7 days with an antigen testing on the last day or daily antigen
testing for 5 consecutive days without quarantine was effective in reducing the potential for secondary cases [39]. Wells et al [40] evaluated the impact of screening timing, such as at the start or end of isolation and during the period of isolation, on the risk of continued transmission after isolation. It was suggested that PCR testing at the start and end of isolation could reduce the risk of continuous transmission and shorten the isolation period from 14 days to 7 days. However, testing only at the start of isolation had no notable effect on reducing the risk of transmission and shortening the isolation period [40]. Fonseca et al [41] simulated the screening of close contacts at different timing, such as days 1, 2, and 3 after exposure to an infector, and compared the expected days of infection during outbreaks. This suggested that antigen testing should be performed on days 1 and 3 for epidemic prevention and control. Two tests were sufficient to effectively prevent infection, and the effectiveness was equivalent to a 14-day isolation period when personnel compliance was 80% to 90%. If an additional test (PCR or antigen) was performed, it was equivalent to the 14-day isolation period effectiveness when personnel compliance was 90% to 100% [41]. Peng et al [42] conducted a similar study to assess the risk of spread after the end of the quarantine period for PCR testing at different timings, such as the days 1, 2, 4, 5 and 6 during the quarantine period. The results showed that PCR or antigen testing at different timings reduced the quarantine period to different degrees without increasing the risk of transmission. Combining testing with shorter quarantine periods is more cost-effective in terms of both time and expenses compared with a 14-day quarantine. For instance, using 3 highly sensitive RT-PCR tests along with a 6-day quarantine yielded a similar risk of transmission as the traditional 14-day quarantine [42].

Optimization of Screening Frequency

Screening frequency should maximize the effectiveness of screening testing and minimize the related costs. Studies on the impact of different screening frequencies on COVID-19 control in communities or high-density populations are also presented in Table 1. Kahanec et al [43] found that 14 days after 2 rounds of mass nucleic acid testing, the infection rate decreased by approximately 30% and the basic reproductive number decreased by approximately 0.3. In a simulation study, Koo et al [44] found that the influence of the test frequency was greater than the maximum test sensitivity (range 0.6-0.8) on the number of infections. The average reduction in infections per day between the 2 testing days was 2.2%, whereas each 1% increase in test frequency reduced infections by an average of 0.43% [44]. Baker et al [45] found that if the number of screening tests per day was slightly higher than the daily testing capacity, it would not cause a burden on testing, but more cases could be found and transmission could be reduced more effectively. Koo et al [46] assessed the impact of biweekly PCR tests or weekly RATs on the number of new infections and infectivity in areas with a high population density (such as workers’ dormitories) and found that biweekly PCR testing (39 new cases per month) was as effective as weekly RATs (33 new cases per month) and could prevent local outbreaks.

Optimization of Screening Procedure

The Importance of the Implementation of the Pooling Strategy

Given the huge demand for sampling and testing, it is important to optimize the screening procedure, and the pooling strategy (ie, collecting multiple samples in a pool for testing) has been frequently used. For example, the National Health Commission of China and the Centers for Disease Control and Prevention of America have issued guidelines for pooled sample tests [47,48]. Three issues related to the pooling strategy have been studied: scenarios for adopting the pooling strategy, pool size, and test procedure.

Scenarios for Adopting a Pooling Strategy

In general, the decision to adopt a pooling strategy is determined by comparing the average number of pool tests per person with a baseline number of 1 (Multimedia Appendix 3 [49-83]). A study demonstrated that when the prevalence exceeded 0.1, the average number of tests per person was >1 for pooled testing with a pool size of 32; such pooled testing was no better than the individual test [49]. The average number of tests per person varied with the pool size. Choosing an optimal pool size can minimize the average number of tests per person. In this case, when the prevalence was <0.07, the pooling strategy could save the need to perform more than half of the tests. When the prevalence was close to or >0.3, the number of pooled tests was close to or exceeded that of the individual test [50]. A prevalence of 0.3 may be considered as the threshold for performing pooled testing.

Pool Size

The optimal pool size can be calculated based on expected positive rate and detection accuracy with the objective of minimizing the number of tests, and the number of tests decreases with decreasing prevalence and increasing pool size. If the accuracy was 100% and the prevalence was 0.001, 0.005, or 0.01, the optimal pool sizes were estimated to be 32, 15, and 10, respectively [51]. There is a certain upper threshold value for the pool size that is limited by testing accuracy (Multimedia Appendix 3).

To estimate the optimal pool size, the prevalence (positivity rate or infection rate) must be assumed. It is also assumed that all individuals are independent of each other, and that the probability of infection is uniform. However, in practice, the prevalence remains unknown until the test results are available (Multimedia Appendix 3). Pikovski and Bentele [52] considered the prevalence to be a random variable uniformly distributed between the expected maximum and minimum values substituted in the calculation of the optimal pool size. An optimal pool size of 4, 3, or 5 was acceptable when the prevalence was uniformly distributed between 0 and 0.3 [52]. In addition, there is heterogeneity and correlation in the probability of infection among people. Fewer tests are needed when individuals in the same pool for testing are homogeneous in terms of age, sex, and other risk characteristics [53-55]. Libin et al [56] considered that combining pools with several families for testing was more conducive to home isolation. The propagation dynamics simulation found that a family-based pool size of 32 and testing
volume of 50,000 per day could achieve the weekly testing of the entire population in Belgium [56]. A larger optimal pool size is required considering the correlation of individuals in the pool [57], or even the social graph [58], in which an edge represents frequent social contacts between 2 persons. Furthermore, Augenblick et al [59] showed that if the pool size could be adjusted to be optimal with the infection rate at any time, screening with a high testing frequency could quickly reduce the infection rate. The final number of tests may decrease despite the high testing frequency owing to the increasing optimal pool size, that is, “frequency gain” [59].

Test accuracy, including screening sensitivity and specificity, also affects the optimal pool size and the upper limit of the pool size. Bish et al [60] found that the optimal pool size would modestly increase when the sensitivity of the pooled sample test decreased. In the PCR test, the sensitivity would decrease and the specificity would increase in a pooled sample test owing to the dilution effect [61], which needs to be considered in the calculation of the optimal pool size [62]. The maximum pool size recommended in previous studies varies from 8 to 30 [63-65]. The sensitivity of the individual test and the influence of the dilution effect of pooling on sensitivity are related to the specific techniques of sampling and testing (such as sampling tools, sample processing reagents, detection instruments, and standardization of operation); therefore, test accuracy is an important determinant for the selection of pool size.

**Pooling Approach**

For the original Dorfman pooling approach, each individual in a positive pool is tested separately. If a pooled sample is negative, then all individuals in the pool are regarded as negative. Several suggestions were made to optimize the pooling approach, which are summarized in Table 2.

First, sequential pooling may be used. A positive pool is divided into several subpools, and the samples in the positive subpools are tested individually [55,66]. Binary pooling divides people to be screened into 2 pools, and the positive pool is divided and pooled again until all positive individuals are found [67]. In the nested pooling strategy, the samples in a positive pool are divided into smaller pools with an optimal number of stages, and the optimal pool size of each stage is calculated according to various measures such as the predicted number of positives and time limit [68,69]. Ng et al [70] conducted simulations of a household-based sequential pooling approach to optimize a universal testing scheme in Hong Kong. They showed that the household-based sequential pooling approach could rapidly screen people in high-risk groups for COVID-19 infections and quarantine those who tested positive [70]. Although these approaches reduce the number of tests required, the operability of such strategies for time-critical epidemic control should be carefully assessed.

Second, repeated testing of the same pool of samples may be conducted to reduce false negatives of the pooled test [71]. Litvak et al [72] conducted a second pooled test after reordering and recombining the samples in the negative pools. For the sequential pooled test, some researchers allocated a part of the samples to 2 subpools to improve accuracy [73].

Third, a copy-link optimization strategy may be used to accurately link the results of the pooled test to the individuals in the pool. The primary “copy-link” strategy is matrix pooling [84]. Samples are arranged in the form of a matrix, with each row and column forming a pool, and each sample is tested once in the row pool and once in the column pool. Research on matrix pooling for COVID-19 has only mathematically simulated the number of tests and the accuracy of the pooled sample test. The Dorfman pool test may be more economical when the prevalence is extremely low, whereas matrix pooling may be more economical when the prevalence is relatively high [74]. Zilinskas et al [75] broadened the concept of matrices by dividing each sample into 2 pools to create as many links as possible between pools. Zhou and Zhou [76] applied the copy-link strategy in designing the Pentagram minipool test. Mutesa et al [77] expanded the 2D matrix to a 3D or multidimensional hypercube, where the number of copies of each sample was split into different planar slices of the hypercube. The subsamples on 1 planar slice of the hypercube were tested in a pool together. Investigators have demonstrated the feasibility of this “hypercube testing strategy” in the laboratory, and field trials are underway in Rwanda and South Africa [77]. Wu et al [78] improved the current hypercube testing strategy by calculating the prevalence, edge, and dimension because every edge had a best performance range, and hypercube pooling with edge=3 may not be the optimal strategy in different outbreaks. Daon et al [79] used a Bayesian model to determine the best combination of pool size, detection steps, repeat detection, and split sample detection to maximize the mutual information between the infection status and testing results. However, this is limited to a simulation analysis.
Table 2. Research on optimization of pooling approach in COVID-19 screening strategy.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Evaluation index</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dorfman pool test</td>
<td>Daily detection capacity</td>
<td>Cleary et al [80], 2021</td>
</tr>
<tr>
<td>Pool test for each sample tested in multiple pools</td>
<td>Number of sample results represented by each test</td>
<td></td>
</tr>
<tr>
<td>Individual test</td>
<td>The number of persons per test</td>
<td>Hanel and Thurner [71], 2020</td>
</tr>
<tr>
<td>Pool test repeated multiple times</td>
<td>The upper bound for the fraction of $N_{miss}^b$</td>
<td></td>
</tr>
<tr>
<td>Individual test</td>
<td>$N_{F\text{-pos}}^d$</td>
<td>Litvak et al [72], 2020</td>
</tr>
<tr>
<td>Dorfman pool test</td>
<td>$N_{test}^e$</td>
<td></td>
</tr>
<tr>
<td>Splitting pool test: samples in the negative pools are recombined to new pool tests, and samples with 2 negative results are identified as negative</td>
<td>$N_{F\text{-neg}}^f$</td>
<td></td>
</tr>
<tr>
<td>Dorfman pool test</td>
<td>$N_{test}^g$</td>
<td>Cheng et al [66], 2021</td>
</tr>
<tr>
<td>Sequential pool test: the positive pool is divided into several subpools of pool size of 3, and the samples in the positive subpool are tested individually</td>
<td>Ratio of number of tests</td>
<td></td>
</tr>
<tr>
<td>Random sequential pooling test the positive pool is divided into several subpools, and the samples in the positive subpool are tested individually</td>
<td>$N_{test\text{ Per person}}$</td>
<td>Millioni and Mortarino [55], 2020</td>
</tr>
<tr>
<td>Informed sequential pool test: divide subjects with similar risk of infection (eg, by age and sex into the same pool)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual test</td>
<td>Ratio of number of tests</td>
<td>Perivolaropoulos and Vlacha [67], 2021</td>
</tr>
<tr>
<td>Binary pool test of the best number and depth of branches considering the prevalence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nested pool test: the positive pool is then divided into several small pools</td>
<td>$N_{test\text{ Per stage}}$</td>
<td>Armendáriz et al [68], 2021</td>
</tr>
<tr>
<td>Pool test strategy based on the optimization algorithm: the positive pool is then divided into several small pools and tested in the next stage</td>
<td>$N_{cases}$</td>
<td>Rai et al [69], 2020</td>
</tr>
<tr>
<td>Pooling test of pooling size of 20 in the family with different prevalence and then retest with minipool for batches with positive results using pooling size of 20, 10, 5, 4, and 2</td>
<td>$N_{test}$</td>
<td>Ng et al [70], 2022</td>
</tr>
<tr>
<td>Multistage pool test: ≥3 stages using the overlap strategy (some samples are detected in both pools)</td>
<td>$IR^h$</td>
<td>Gu et al [73], 2021</td>
</tr>
<tr>
<td>Individual test</td>
<td>Costs</td>
<td>Kim et al [74], 2022</td>
</tr>
<tr>
<td>Dorfman pool test</td>
<td>Cost per test</td>
<td></td>
</tr>
<tr>
<td>Matrix pool test</td>
<td>Positive rate</td>
<td></td>
</tr>
<tr>
<td>Individual test</td>
<td>$N_{\text{test-saving}}$</td>
<td>Žilinskas et al [75], 2021</td>
</tr>
<tr>
<td>Matrix pool test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Study Evaluation index Strategy Zhou and Zhou [76], 2022
- Individual test Ntest
- Dorfman pool test
- Pentagram minipool test: for the positive Dorfman pooling with size of 10, double samples are tested by 5 “three-in-one” pools and 1 “five-in-one” pool
- Subsample pool test in the hypercube algorithm Loss of sensitivity compared with individual test Ntest Per person and Ninfected
Mutesa et al [77], 2021
- Pooling test under different prevalence, edge, and dimension using the hypercubic method Ntest
Wu et al [78], 2022
- Dorfman pool test FNR and FPR
- Recursive pool test Ntest
- Matrix pool test
- D-Optimal Pool Experimental design (a novel Bayesian pooling strategy)
Daon et al [79], 2021
- Individual test Confirmed cases per test
- 2-stage Dorfman pool test Time to test the whole population
- Binary splitting pool test N_{T-pos}, N_{F-pos}
- Optimized recursive binary splitting pool test Ncases
- Matrix pool test Number of quarantined individuals
de Wolff et al [81], 2020

Discussion

Principal Findings

Despite ongoing vaccinations worldwide, COVID-19 is still present and causing outbreaks, and screening remains important. First, SARS-CoV-2 has the potential to mutate, and the transmissibility, pathogenicity, and incubation period of the novel variant remain unknown. Screening facilitates the surveillance and tracking of SARS-CoV-2 novel variants, providing essential information for an appropriate response. In addition, screening helps protect vulnerable populations and reduce the pressure on the health care system. Therefore, the continuous optimization of screening strategies to improve cost-effectiveness and reduce resource consumption is still worthy of our attention in the current global situation of relaxed prevention and control. Furthermore, the development of strategies for screening COVID-19 provides a basis for the prevention and control of novel or re-emerging infectious diseases in the future, particularly respiratory infectious diseases.

In previous studies, researchers have been optimizing the screening strategy for COVID-19 based on the target population, timing, frequency of screening, and testing procedure and providing a scientific basis for COVID-19 screening. Specific strategies are designed and developed for different populations according to risk levels, regions, or occupations; different timing and frequencies (eg, before, during, and after traveling or entry and exit); and different testing procedures (eg, individual or pooled test, pool size, and polling approaches). The conceptual model for developing screening strategies is available in Multimedia Appendix 4. As there is a continuous emergence of new variants of SARS-CoV-2, further research is necessary to improve the current screening strategy by addressing the issues on the scale of screening, proper timing and frequency of testing, testing accuracy, and cost-effectiveness.

Dynamic Adjustment of Screening Strategies Based on Variations in Prevalence

The design and optimization of any screening strategy are based on disease prevalence during an epidemic as a hypothetical
condition. This determines the strictness of the screening strategy, including the screening frequency, time interval, target population, and testing procedures. For example, the selection of pool size in pooled testing depends on the prior prevalence. To apply the optimization plan in practice, the information on real prevalence is important for determining the screening plan at a specific time in a specific area, which may be obtained from the screening results at the same time. Nicholson et al. [85], Hamadeh et al. [86], and Chiu and Ndeffo-Mbah [87] explicitly estimated the real prevalence by statistically correcting the reported data, which can be used in the research and practice of optimizing and adjusting screening strategies over time. Additionally, pandemics are constantly evolving. Screening results can help policy makers and researchers understand current risk levels and trends as well as reflect past screening effects. A feedback mechanism should be created to dynamically adjust the screening strategy according to the screening results. When the number of cases increases rapidly, the stringency of screening should be increased to prevent a larger pandemic outbreak. For example, Yu et al. [82] set several alternative pool sizes and updated the pool size within an alternative range every week according to the changes in prevalence in a simulation study. A dynamic screening plan should be relatively stable, easy to implement, and adaptable to the changing trend of an epidemic or outbreak.

**Optimizing Screening Strategies by Combining Multiple Testing Methods**

Various testing methods have been developed. Some methods are more appropriate than others for different screening processes. Self-detection, such as the use of antigen tests, has been widely promoted. As a supplement to laboratory detection methods, such as PCR, it can effectively save the resources of medical workers for sampling and testing. It has been applied for the screening of both small and large populations. However, the accuracy of the current antigen test is lower than that of the PCR test, with an overall sensitivity of approximately 70% and a specificity of approximately 98%, and its accuracy for asymptomatic patients is lower than that for symptomatic patients [88]. Moreover, there are no guaranteed standardization of sampling for self-detection and compliance with self-isolation after diagnosis. Self-detection is often an alternative when professional testing capabilities are insufficient, and self-detection and laboratory testing complement each other. A few studies have compared laboratory tests with self-tests or combined antigen tests with PCR tests [35,38,39,42,46,86]. Yu et al. [82] found that the PCR pooled test tended to be more cost-effective at low prevalence because it allows more people to be tested with existing equipment and quarantines more patients with presymptomatic and asymptomatic infections to prevent future infections. However, a pooled PCR test would cause delays in results and would not facilitate timely quarantine of infected persons and interrupt transmission at a high prevalence. In contrast, high-frequency antigen screening may reverse the epidemic to obtain results quickly and quarantine infected people in a timely manner, despite the large number of false positives in the screening process [82]. Although the sensitivity of the PCR pooled test may be higher than that of the RAT, there is a problem with a large number of tests leading to delays in reporting. RATs are particularly useful in settings such as schools, workplaces, and mass gatherings with a high frequency owing to their characteristics of lower cost, rapid time to result, and increased accessibility. A combination of laboratory testing and self-detection may maximize the benefit under limited resources, which needs to be studied in the future.

**Accuracy of Screening Test**

Whether an infected person can be identified depends on their infection status after exposure because viral nucleic acid, antigen, antibody, and other biomarkers change with infection time. The accuracy of testing results may also change with time. These characteristics can help in the selection of the appropriate timing for screening and quarantine of close contacts and entry-exit persons with a definite exposure time [39].

Screening accuracy is affected by sampling quality and detection accuracy. In China, sampling quality is monitored based on whether human somatic cells are collected from the swab as a laboratory indicator [89]. As the dilution effect caused by pooling is likely to reduce test sensitivity, it is crucial to consider PCR testing with optimal sensitivity and the maximum pool size. The sensitivity may be affected by the sample handling method, selection of the detection kit, and standardization of the detection operation [90]. The accuracy of screening should be evaluated when designing pooling strategies, which is conducive to the selection of pool size and the credibility of the screening results.

**Optimization of PCR Testing Procedures to Improve Detection Capacity**

Optimizing the nucleic acid detection procedure for the pooled sample test using various approaches reduces the number of tests performed and increases the detection speed and capacity, which can improve the speed of obtaining results for quarantine-infected individuals. The primary Dorfman procedure has been implemented during COVID-19 screening in some areas owing to its convenience and operability. The pool size when screening the entire population in China was 5, 10, or 20 [48,91], and the US Food and Drug Administration authorized 5 pool tests with pool sizes ranging from 3 to 10 [47]. The optimal selection and dynamic adjustment scheme of the pool size need to be in accordance with the actual prevalence and detection accuracy. The pooling procedure has been improved using a multistage strategy and “copy-link” strategies for higher detection efficiency [81]. The multistage strategy reduces the number of tests while simultaneously increasing the number of test rounds. The turnaround time from sampling to result reporting may increase, but its operability has not yet been investigated. The “copy-link” strategy not only reduces the amount of testing but also requires only 1 testing stage to locate positive individuals. However, for this strategy, the laboratory requires more complex testing procedures and equipment. It is not feasible to apply “copy-link” strategy to an existing laboratory platform.

**Accurately Define the Scope of Screening Individuals**

A screening strategy for the whole population can effectively control the outbreak, but it also has a significant impact on health care resources and delays the time of transferring cases.
and contacts to quarantine [45]. Therefore, the occurrence of cases does not imply the need for screening the entire population. In China’s dynamic zero policy, screening of the entire population of an area is often initiated in the early stage of an outbreak and interrupts the spread of COVID-19 [92]. However, there is a lack of research on when and under what circumstances a full screening should be initiated, and whether specific thresholds for the number of cases or incidence are necessary in the decision-making process. High-risk populations are usually targeted for screening such as people from high-risk areas or countries or people with special occupations. Screening helps to isolate infectious sources to prevent COVID-19 outbreaks [93].

Consider the Implementability of Screening Strategies

Previous studies on the optimization of screening strategies have mainly focused on how to control an epidemic, reduce the risk of transmission, and shorten the isolation period. However, the design and implementation of screening strategies depend on the resources available, which vary greatly among countries or regions. A large-scale screening program is costly and may not only be sustainable owing to limited resources.

Larremore et al [94] raised the meaningful problem of whether the frequency of testing, time to obtaining results, or sensitivity is more important in new crown screening. Their study revealed that effective testing depends largely on the frequency of testing and the speed of obtaining results, whereas the sensitivity of the test is relatively secondary [94]. This means that the implementation of relatively low-sensitivity PCR pooled tests or antigenic tests can improve the speed of obtaining results, facilitate quarantine of infected individuals, and interrupt transmission. This may be appropriate in areas with high prevalence.

Methods of Comparing Different Screening Strategy

When searching for an appropriate screening strategy, the most commonly used method involved using model simulations to evaluate the effectiveness of different screening strategies. For the selection of models to compare different screening strategies, the appropriate model should be based on different situations, populations, and purposes. In addition, real-world disease data are always available during the pandemic, but very few studies have evaluated the effectiveness of different screening strategies from the perspective of real-world data. The conclusions of evaluating the effectiveness of screening strategies using real-world data are more realistic. Model simulations enable a convenient comparison of the effectiveness of different screening strategies for different situations, but they are based on a hypothetical theoretical setting. Furthermore, if the results from the model simulations can be validated using real-world data, the findings of these studies will be more reliable.

Regarding the comparison of different screening strategies, which is the best strategy depends on the actual situation. The selection of a screening strategy in realistic scenarios requires a balanced consideration of the economic costs and effectiveness of controlling the outbreak. Only a few studies reviewed in this paper analyzed cost-effectiveness in a cursory manner; for example, the simple indexes such as cost per test [74] or cost per infection reduced [82] were used. Furthermore, attention should also be paid to resource consumption, such as medical resources (eg, hospital beds) and quarantine resources (eg, hotel rooms), which determine the feasibility and sustainability of the strategy [95].

Limitations

First, we restricted studies to those publications in English and did not search gray literature and might have missed relevant studies published in other languages and in non-peer-reviewed journals and conference proceedings. Second, we restricted the search terms to the title or abstract field and might have excluded some studies that included the search terms as Medical Subject Headings terms or free text. Third, we did not perform duplicate screening of publications owing to time limitations. However, data extraction was performed by 9 investigators from our team, and 2 investigators in cooperation conducted quality assessments (YL and YY) to ensure consistency. Finally, most publications in this review were simulation studies using mathematical models, which lack accepted criteria for quality evaluation; therefore, literature quality assessments were not performed.

Conclusions

A well-designed and developed COVID-19 screening strategy is conducive to the rapid identification of infected individuals and the control of an epidemic. As new variants continue to emerge, screening strategies should be dynamically adjusted and optimized to achieve expected results. To identify and isolate infected individuals in a timely manner, a screening strategy must produce fast and accurate results. A program is sustainable only when costs can be controlled at the level of available resources. Some key elements for COVID-19 screening strategies are reviewed and discussed, including the screening population, timing and frequency of screening, detection methods, and procedures.

Acknowledgments

This review was supported by the National Key Research and Development Program of China (2022YFC2602900), National Natural Science Foundation of China (81973102), Public Health Talents Training Program of Shanghai Municipality (GWV-10.2-XD21), and Shanghai New Three-year Action Plan for Public Health (GWV-10.1-XK16).

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.
**Authors' Contributions**
ZZ and LK jointly conceived and supervised this scoping review. YL, YY, MPW, YC, PPYW, ZZ, and LK designed the methodology. YL and YY led the retrievals and screening of the publications. YL, YY, KL, MD, J Hong, J Huang, JS, ZX, XC, JX, and RY verified the eligibility of the included publications and extracted the relevant information. YL and YY drafted the manuscript. All authors have revised the manuscript and approved the final version.

**Conflicts of Interest**
None declared.

**Multimedia Appendix 1**
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 38 KB - publichealth_v10i1e44349_app1.docx ]

**Multimedia Appendix 2**
[DOCX File , 41 KB - publichealth_v10i1e44349_app2.docx ]

**Multimedia Appendix 3**
Research on the optimization of SARS-CoV-2 nucleic acid detection strategy.
[DOCX File , 39 KB - publichealth_v10i1e44349_app3.docx ]

**Multimedia Appendix 4**
Conceptual model for screening strategy development.
[DOCX File , 80 KB - publichealth_v10i1e44349_app4.docx ]

**References**


61. Lakdawalla D, Keeler E, Goldman D, Trish E. Getting Americans back to work (and school) with pooled testing. Schaeffer Center for Health Policy & Economics. URL: https://healthpolicy.usc.edu/research/getting-americans-back-to-work-and-school-with-pooled-testing/ [accessed 2022-07-08]


Abbreviations

LFT: lateral flow test  
PCR: polymerase chain reaction  
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews  
RAT: rapid antigen test  
RT-PCR: reverse transcription-polymerase chain reaction
Review

Short- and Long-Term Predicted and Witnessed Consequences of Digital Surveillance During the COVID-19 Pandemic: Scoping Review

Leigha Comer1*, PhD; Lorie Donelle1,2*, PhD; Bradley Hiebert1*, PhD; Maxwell J Smith3*, PhD; Anita Kothari3*, PhD; Saverio Stranges4,5,6,7*, MD, PhD; Jason Gilliland8*, PhD; Jed Long9*, PhD; Jacquelyn Burkell9*, PhD; Jacob J Shelley10*, JD, LLM, LLB, MTS; Jodi Hall11*, PhD; James Shelley11*, BA; Tommy Cooke12*, PhD; Marionette Ngole Dione1*, MSN; Danica Facca9*, MA

1Arthur Labatt Family School of Nursing, Western University, London, ON, Canada
2School of Nursing, University of South Carolina, Columbia, SC, United States
3School of Health Studies, Western University, London, ON, Canada
4Department of Epidemiology and Biostatistics, Schulpich School of Medicine & Dentistry, Western University, London, ON, Canada
5Departments of Family Medicine and Medicine, Schulpich School of Medicine & Dentistry, Western University, London, ON, Canada
6The Africa Institute, Western University, London, ON, Canada
7Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy
8Department of Geography and Environment, Western University, London, ON, Canada
9Faculty of Information and Media Studies, Western University, London, ON, Canada
10Western Law, Western University, London, ON, Canada
11Faculty of Health Sciences, Western University, London, ON, Canada
12Surveillance Studies Centre, Queen’s University, Kingston, ON, Canada
*all authors contributed equally

Corresponding Author:
Lorie Donelle, PhD
School of Nursing
University of South Carolina
1601 Greene Street, Room 617
Columbia, SC, 29208-4001
United States
Phone: 1 803 777 6528
Email: LDONELLE@mailbox.sc.edu

Abstract

Background: The COVID-19 pandemic has prompted the deployment of digital technologies for public health surveillance globally. The rapid development and use of these technologies have curtailed opportunities to fully consider their potential impacts (eg, for human rights, civil liberties, privacy, and marginalization of vulnerable groups).

Objective: We conducted a scoping review of peer-reviewed and gray literature to identify the types and applications of digital technologies used for surveillance during the COVID-19 pandemic and the predicted and witnessed consequences of digital surveillance.

Methods: Our methodology was informed by the 5-stage methodological framework to guide scoping reviews: identifying the research question; identifying relevant studies; study selection; charting the data; and collating, summarizing, and reporting the findings. We conducted a search of peer-reviewed and gray literature published between December 1, 2019, and December 31, 2020. We focused on the first year of the pandemic to provide a snapshot of the questions, concerns, findings, and discussions emerging from peer-reviewed and gray literature during this pivotal first year of the pandemic. Our review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) reporting guidelines.
Results: We reviewed a total of 147 peer-reviewed and 79 gray literature publications. Based on our analysis of these publications, we identified a total of 90 countries and regions where digital technologies were used for public health surveillance during the COVID-19 pandemic. Some of the most frequently used technologies included mobile phone apps, location-tracking technologies, drones, temperature-scanning technologies, and wearable devices. We also found that the literature raised concerns regarding the implications of digital surveillance in relation to data security and privacy, function creep and mission creep, private sector involvement in surveillance, human rights, civil liberties, and impacts on marginalized groups. Finally, we identified recommendations for ethical digital technology design and use, including proportionality, transparency, purpose limitation, protecting privacy and security, and accountability.

Conclusions: A wide range of digital technologies was used worldwide to support public health surveillance during the COVID-19 pandemic. The findings of our analysis highlight the importance of considering short- and long-term consequences of digital surveillance not only during the COVID-19 pandemic but also for future public health crises. These findings also demonstrate the ways in which digital surveillance has rendered visible the shifting and blurred boundaries between public health surveillance and other forms of surveillance, particularly given the ubiquitous nature of digital surveillance.

International Registered Report Identifier (IRRID): RR2-https://doi.org/10.1136/bmjopen-2021-053962

(JMIR Public Health Surveill 2024;10:e47154) doi:10.2196/47154

KEYWORDS
digital surveillance; COVID-19; public health; scoping review; pandemic; digital technologies

Introduction

Public health surveillance has been described as one of the most critical functions and mechanisms of public health [1]. With a history dating back to Snow’s pioneering investigations into cholera epidemics in England in the 19th century, surveillance has long been leveraged as a tool for meeting the objectives of public health, including population health assessment, health surveillance, health promotion, disease and injury prevention, and health protection [1]. The Centers for Disease Control and Prevention describe public health surveillance as “the ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice” [2]. Other definitions include the World Health Organization’s conceptualization of surveillance as continued watchfulness and the monitoring of events linked to action [3]. More broadly, Lyon [4] has defined surveillance as “the focused, systematic, and routine attention to personal details for purposes of influence, management, protection or direction.”

One of the outcomes of the COVID-19 pandemic and the corresponding global response has been the rapid development and deployment of digital technologies used for public health surveillance. New and existing digital technologies, including mobile phone location tracking, mobile phone apps, drones, and closed-circuit cameras, have been used for surveillance purposes in supporting the detection and mitigation of disease spread and to enhance compliance with public health measures [5]. The accelerated use of these technologies during the COVID-19 pandemic has led to some scholars describing the COVID-19 pandemic as the first public health crisis of the “digital age” [6] and as a catalyst for a “digital revolution” in surveillance [7].

Despite the extensive use of a plethora of digital solutions for public health surveillance, their actual efficacy in predicting disease spread or supporting public health responses has been unclear [8,9], and there remain ongoing questions about the ethical dimensions of digital surveillance. For instance, the use of digital tools by the state, its agents, the private sector, and other actors has raised concerns around the potential consequences of surveillance, particularly for those who are most marginalized [10-12]. The rapid pace of technological development has also limited opportunities to consider the consequences of technology use or misuse and whether digital technologies are actually effective in mitigating the health-related, economic, and social effects of the pandemic [13].

The growing attention paid to digital surveillance during the COVID-19 pandemic has also offered an important opportunity to reflect on the function of surveillance within public health. For instance, the ubiquitous nature of contemporary modes of data collection enabled by digital surveillance technologies has prompted a reconsideration of pervasive health monitoring [14]. The collection of large amounts of data from sources that have not traditionally fallen within the purview of public health surveillance (eg, transaction monitoring, closed-circuit cameras, drones, and location tracking through mobile devices) presents an urgent need to contend with the consequences of surveillance and the potentially invasive and coercive aspects of public health. Likewise, the use of digital surveillance technologies by a range of actors and for a variety of different objectives throughout the COVID-19 pandemic signals an important opportunity to consider the blurred boundaries between public health surveillance and other forms of surveillance (eg, criminal–legal surveillance, commercial use of data) and how these modes of surveillance and their aims overlap.

This scoping review examined peer-reviewed and gray literature on the global use of digital technologies for public health surveillance during the COVID-19 pandemic to (1) describe the nature of digital technologies used for surveillance during the COVID-19 pandemic, (2) describe the witnessed and potential short- and long-term implications of using digital technologies for public health surveillance, (3) synthesize the peer-reviewed and gray literature regarding the use of such technologies for surveillance during the COVID-19 pandemic.

https://publichealth.jmir.org/2024/1/e47154

Jmie Public Health Surveill 2024 | vol. 10 | e47154 | p.354

(page number not for citation purposes)
response, and (4) identify gaps within this knowledge base. To meet these objectives, the review was guided by two research questions: (1) What is known about digital technologies used for surveillance during the COVID-19 pandemic? and (2) What are the short- and long-term predicted and witnessed impacts of digital surveillance during the COVID-19 pandemic?

Methods

Study Design
The protocol used to conduct this research followed the 5-stage methodology described by Arksey and O’Malley [15] and Levac et al [16], which included (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the findings. A summary of the study’s methodology is included below; for a complete description, refer to a previously published scoping review protocol [17]. The study’s reporting structure also conformed to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines and checklist (Multimedia Appendix 1 [18]) [19].

Data Collection

Peer-Reviewed Literature Search
In January and February 2021, the MEDLINE (Ovid), PubMed, Scopus, CINAHL, ACM Digital Library, Google Scholar, and IEEE Xplore databases were searched for peer-reviewed literature. Journals and reference lists of relevant papers identified by our interdisciplinary team of researchers were also reviewed for additional publications. Textbox 1 shows the search terms and strategy created with guidance from a specialist research librarian.

Textbox 1. Search terms developed with the assistance of a health specialist research librarian.

1. Population Surveillance or Public Health Surveillance, or surveillance.tw.
2. digital surveillance.tw.
3. biosurveillance.tw or Biosurveillance
4. epidemiological monitoring.tw. or Epidemiological monitoring
5. 1 or 2 or 3 or 4
6. pandemic.t.w. or Pandemics
7. disease outbreak.tw. or Disease Outbreaks
10. H1N1.t.w.
11. SARS.t.w. or SARS Virus
12. 6 or 7 or 8 or 9 or 10 or 11
13. Public Health or public health application.mp.
14. 5 and 12
15. 13 and 14

An initial broad search was conducted to capture all English-language publications on the use of digital technologies for public health surveillance during previous and current pandemics, epidemics, and outbreaks. Documents were retained if they were published between January 1 and December 31, 2020. This preliminary search yielded 18,449 publications for review. Following the removal of 8819 duplicates, 9630 unique documents were subjected to title and abstract screening by 2 researchers (LC and MND) to assess their fit with the study’s aims. In this initial search, the following inclusion and exclusion criteria were used during the title and abstract screening: (1) document title or abstract mentioned the use of a digital technology for public health surveillance, (2) digital surveillance focused on containing or limiting the spread of an infectious disease, (3) public health surveillance monitored humans rather than nonhuman animals, and (4) digital technology used explicitly for surveillance (eg, data collection).

Consequently, 2076 publications were retained for full review following title and abstract screening. Following a full-text review, 888 publications were retained for analysis.

To conduct a feasible review of the literature, the research team narrowed the scope of the review to focus on COVID-19-pandemic–specific apps of digital technologies for public health surveillance. During this secondary screening, documents were retained for analysis if they were published between December 1, 2019 and December 31, 2020, and if they explicitly included any of the terms such as “coronavirus,” “COVID 19,” “SARS-CoV-2,” or “severe acute respiratory syndrome coronavirus 2” in the title or the abstract. This period was chosen not only to obtain a manageable number of publications for review but also to focus our attention on the first year of the COVID-19 pandemic as a means of better understanding the questions, concerns, findings, and discussions that emerged from the academic and nonacademic literature during this pivotal time. This process led to a final study sample...
size of 147 unique documents. Throughout this process, in cases in which the 2 screening researchers disagreed or were uncertain whether a publication met the inclusion criteria, a third researcher read the text and discussed with the 2 researchers to decide whether to include the document for analysis.

**Gray Literature Search**

Our search of the gray literature was guided by our multidisciplinary team of researchers as well as a specialist research librarian, who assisted in identifying relevant organizational websites that explore the use of digital technologies for surveillance. These websites included the Ada Lovelace Institute, Human Rights Watch, and the Munk School. The websites of these organizations were searched to retrieve potentially relevant current and archived publications using the same protocol and search terms used to search peer-reviewed literature. Publications were retained for screening if their content related to the use of digital technologies for surveillance during the COVID-19 pandemic. The initial search of the gray literature yielded 141 documents.

Gray literature publications were screened using the same inclusion criteria used to screen peer-reviewed literature. After 2 rounds of screening, 74 documents were retained for analysis. Five conference proceedings identified while searching the peer-reviewed literature were also included as gray literature for a total of 79 publications.

Due to the breadth and depth of the literature relevant to this review, attention to the social and ethical implications of digital health surveillance for public health purposes during the COVID-19 pandemic response is the focus of this scoping review. A second paper describes the types of digital technologies used during the COVID-19 pandemic response and the factors impacting their effectiveness in public health surveillance [8]. Figure 1 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) chart describing the study’s inclusion process.

**Figure 1. Study selection process.**
Data Analysis
Documents were analyzed through an inductive coding process to identify dominant themes and patterns [20]. Following an initial analysis of 10 documents and discussion among 5 researchers, an extraction table was developed and further refined upon analysis of 5 additional publications. Five researchers piloted the data extraction table to ensure consistency of analysis and collaboratively reviewed all retained peer-reviewed and gray literature documents for data extraction and analysis. The completed table was used by 5 reviewers to extract data from each publication related to (1) authors; (2) author locations based on affiliated institutions as identified in publications; (3) study aim or objective; (4) methodology and research design; (5) countries or regions identified; (6) types of digital technologies identified; (7) intended or stated outcomes of digital technology use; (8) target populations; (9) unintended (or not explicitly stated) outcomes and implications of technology use; and (10) theoretical analyses, arguments, discussions, and debates. The findings were reviewed and discussed among our multidisciplinary team of researchers to ensure a range of expert perspectives from nursing, medicine, public health, epidemiology, surveillance, geography, health information science, law, bioethics, policy, knowledge translation sciences, critical theory, and sociology of health.

Results
Overview
The peer-reviewed literature included commentaries (n=39), reviews (n=39), analyses (n=27), reports (n=8), empirical studies (n=24), legal reviews (n=3), and a policy analysis paper (n=1). The gray literature included commentaries and opinion pieces (n=33); video and audio recordings and transcripts of conferences, panels, and meetings (n=8); news reports (n=33); and conference proceedings (n=5). Identifying the institutional affiliations of authors of peer-reviewed publications allowed us to grasp the international breadth of the literature (authors of more than 1 publication were counted only once). Authors were affiliated with institutions from countries including the United States (n=156), the United Kingdom (n=91), India (n=40), China (n=28), Australia (n=28), Singapore (n=27), Spain (n=25), Italy (n=24), Germany (n=21), Brazil (n=20), Japan (n=19), the Democratic Republic of the Congo (n=16), Switzerland (n=16), Taiwan (n=15), Canada (n=14), Ireland (n=12), South Africa (n=11), and Iran (n=9).

Across the publications reviewed, the use of digital technologies for public health surveillance during the COVID-19 pandemic was identified in 90 countries and regions (Multimedia Appendix 2). Mobile phone apps (particularly apps to support contact tracing), mobile phone location tracking through Bluetooth and GPS technologies, drones, temperature-scanning technologies, and wearable devices were among those most commonly discussed. Digital technologies were most frequently used to support contact tracing (ie, to identify individuals with the disease or symptoms and those with whom they have been in contact) [10,11,21-173]; to inform decision making by public health agencies, states, and individuals [21,25,28,31,44,47,49,51-53,56-60,62,63,66,67,77,81-83,88-90, 94,97,98,102,104,121,129,157-159,167,170,171,174-204]; to monitor compliance with public health measures [21,23,25,28,30-34,38,41,42,44,45,54-58,60,61,63,71,83,85,86,89,91, 94,95,100-102,105,108,110,119,125-127,129,130,135,140,155,156,165,170, 174,182,185,188,190,197,198,205-215]; and to inform the provision of health care services [28,41,44,46,54,57,62,64,66,69,77,80-82,87,94,95,102,108,126,128,159,170,174, 177,182,183,187,191-193,197,198,201,202,216-224].

Thematic Analysis
Key considerations related to the use of digital technologies for COVID-19 public health surveillance were grouped into five themes: (1) data security and privacy, (2) function and mission creep related to digital surveillance, (3) consequences of private sector involvement in surveillance, (4) human rights and civil liberties and impacts on marginalized groups, and (5) ethical principles for digital technology design and use.

Data Security and Privacy
Many authors identified risks to data security and data privacy associated with the use of digital technologies for public health surveillance (n=138) [11,22-26,28,30,31,33-45,47,48,51, 53,55,57-68,70-77,79-81,83,84,86-92,94-98,101-110,114,116, 118,121-123,125,128-130,132,134,137,147-152,154-156,165,168,169,172-174,179, 185,187,191,192,195,198,201,203,204,207,208,210,212,216,221,225-232]. Several authors argued that digital surveillance may exacerbate risks to information privacy given that data gathered through digital technologies are increasingly granular and that health data privacy protections often do not adequately address the collection and use of data through digital technologies, such as mobile apps [22,25,37,42,98,110,132,156,162,164,179,228,229]. Others noted that most technologies used for surveillance are products of private companies, and they advised caution related to potential function creep and exploitation (eg, data monetization) of data intended for public health purposes. The idea of a capitalist data strategy is a relatively new aspect of disease surveillance for many public health systems [34,211], and several authors acknowledged that in many countries, including Canada and the United States, health data regulations generally apply to data collection by the state and its agents but do not necessarily extend to private companies [24,33,132,190,195,211]. Issues with data security were also found to present potential risks to data privacy, particularly with regard to technologies that were rapidly deployed without full consideration of potential security vulnerabilities [10,28,33,97,102,105,110,147,150,233]. The UK government, for instance, posted the code for its mobile contact tracing app for public review and received feedback that data logs were not adequately encrypted and risked revealing individuals’ personal information [150]. Likewise, in a review of 50 mobile apps, the authors of another study found that while 30 of the apps required access to features including contacts, photographs, media files, location data, the camera, and call information, only 16 of the reviewed apps indicated that users’ data would be made anonymous, encrypted, or secured [28]. A number of digital technologies used globally during the COVID-19 pandemic were identified as examples of privacy-
and security-compromising surveillance. A frequently cited example was the use of cell phone location tracking to support contact tracing and quarantine enforcement in countries including Israel, India, the Czech Republic, China, South Korea, Argentina, Bahrain, Kuwait, and Norway [30,36,42,54,61,89,115,155,190,204,225]. The use of thermal cameras to detect fever among potentially symptomatic individuals in countries including the United States, China, South Korea, Thailand, Singapore, Italy, and the United Kingdom was also identified as a form of surveillance that may reveal sensitive information about underlying health conditions (eg, pregnancy) or stigmatized behaviors (eg, substance use) [44,95,221]. While van Natta et al [221] found that thermal cameras are typically construed as minimally invasive, they cautioned that data regarding individuals’ health could be used by data brokers, advertisers, employers, and law enforcement agencies.

Several authors conceptualized individuals’ right to privacy as a trade-off between data privacy and security and public health. The decision by different states to use centralized or decentralized contact tracing databases, for instance, was frequently mentioned as an example of the tensions between mitigating disease spread, protecting public health, and ending public health measures, such as lockdowns, while also protecting individuals’ data privacy and security [31,34,35,46,112-114,120,121,123,137,139,152,163,211]. While data held in centralized databases can be more readily accessed and analyzed (eg, by public health agencies), they are also more vulnerable to data hacks, data leaks, and other security compromises. In contrast, Christou et al [35] provided an alternative perspective in suggesting that the trade-off between the right to health and the protection of fundamental rights, such as privacy, is a false dilemma. They argued instead that these objectives are intertwined and that human rights standards will strengthen global efforts to mitigate the COVID-19 pandemic because greater human rights protections will heighten users’ trust in surveillance tools. The authors challenged the dichotomy between public health and private information and argued that decisions to infringe upon individuals’ data privacy should be informed by wider discussions of social and political values [96].

**Function and Mission Creep Related to Digital Surveillance**

Several authors (n=30) noted that use of digital surveillance may lead to forms of "creep," such as mission creep and function creep [21,22,26,28,31,34,36,44,54,61,75,83,86,92,93,95,96,101,102,123,160,185,190,195,196,210,211,221,226,230]. Broadly, the concept of “function creep” has been used to refer to instances in which there is a slow, nearly imperceptible transformation in a data processing system’s proper or intended activity. Meanwhile, the term “mission creep” has been taken up to identify the expansion of a project or intervention beyond its original aim, particularly if an intervention continues to be used beyond the end of the crisis during which it was implemented. In the literature examined in this paper, discussions of “creep” tended to be largely speculative in nature, as authors pointed to historical instances of emergency surveillance that outlasted the initial crisis or led to the repurposing of surveillance measures (eg, surveillance measures deployed by the United States and other states following the 9/11 crisis) and considered the potential for similar forms of creep that may occur following the COVID-19 pandemic [26,44,55,74,190,226].

While some authors focused on the potential for some states to take advantage of the COVID-19 pandemic to entrench forms of surveillance that might otherwise be seen as extraordinary [28,31,36,115,135,196,210,226], others noted that digital surveillance justified by the need to mitigate the COVID-19 pandemic may normalize surveillance. In this way, civil liberties (eg, the right to privacy and the right to freedom of movement) ceded during the crisis may be difficult to reclaim. Likewise, interventions not normally accepted may be normalized over time [10,11,23,31,36,44,48,76,86,91,93,103,115,135,190,203,205,209,211]. For instance, Leclercq-Vandelannoitte and Aroles [86] drew on Deleuze’s concept of control societies to analyze the normalization of digital surveillance during the COVID-19 pandemic. The authors considered the ways in which digital technologies (eg, mobile apps to support contact tracing used in China, Australia, Denmark, and Singapore) enable “control societies” as they facilitate insidious tracking of individuals’ everyday activities, which may normalize ubiquitous surveillance.

In some cases, authors identified specific technologies and modes of surveillance that may outlast the COVID-19 pandemic and that may repurpose COVID-19 health-related data. For instance, Kitchin [44,190] referenced the Indian state-sponsored mobile contact tracing app, which may be repurposed for political purposes to monitor and discriminate against certain (unspecified) populations. French et al [45] identified historical examples of racial biases in algorithms and questioned whether COVID-19 health data gathered through mobile contact tracing apps might be used, in the future, by law enforcement against racialized and other vulnerable communities. McDonald [135] described vaccination efforts in Bangladesh that may lead to the establishment of a state-supported digital identity program, including a data directory of vaccinated individuals. Pakes [230] discussed “panic buying” of digital monitoring systems for at-home workers (eg, keyboard-tracking technologies, video-monitoring tools) by unspecified employers and argued that the use of these technologies may normalize surveillance that extends beyond the COVID-19 public health mandate and persists after the pandemic has ended.

Tracking disease and its spread is a core mandate of a well-functioning public health system and critical to mitigating the COVID-19 pandemic. The rapid and unprecedented number of technologies employed to contribute to disease tracking sets the COVID-19 pandemic response apart from all other infectious disease and pandemic responses [211]. As such, many authors (n=46) expressed concern regarding the lack of guidelines related to the collection and use of individuals’ information within contemporary digital public health surveillance practices. They also emphasized the tensions created by the omnipresence of digital public health surveillance technologies and the urgency of their employment, which prohibited full discussion of the impacts and implications of digital surveillance strategies on public understanding and acceptance of surveillance practices.
Concerns were also raised around the existence of national and international laws that justify forms of surveillance and contraventions of human rights during times of emergency that would not be acceptable otherwise. The European Union’s general data protection regulation was cited as an example of a data privacy and security law that allows for nonconsensual surveillance and restrictions on privacy to meet public health objectives during crises [35,55,75,98,196,234]. Several authors warned that states may take advantage of public health emergencies to introduce or expand nonconsensual forms of surveillance, as witnessed in states including India, Israel, South Korea, and Singapore, which framed COVID-19 as a threat that required extraordinary measures, including location tracking through mobile devices [30,31,35,42,55,75,98,210,234].

**Consequences of Private Sector Involvement in Surveillance**

Several authors (n=27) highlighted private sector involvement in digital surveillance for public health and how private interests may lead to invasive, inequitable, or harmful surveillance if data are collected and used for commercial purposes [21,23,24,29,31,35,44,45,58,60,75,86,93,114,115,122-125,162,173,185,190,195,204,211,213,230]. Many of the digital technologies identified within the literature were developed and deployed through private companies or private–public partnerships. These include mobile contact tracing apps developed by the private sector in partnership with governments and public institutions, such as the Apple–Google App Programming Interface designed to support mobile contact tracing [23,24,33-36,44,45,53,57-59,61,64,66,67,70,72,74,75,78,83,86,90,94-96,105-107,112,114,144,152,160,190,195,204,211,213,230], and the use of aggregated and anonymized data collected by private companies, including telecommunications providers [37,44,55,76,121,125,128,129,154,179,190,210].

Private sector involvement in data collection and use has led to the development of proprietary technologies that are not interoperable and that may contribute to a fragmented public health system [35,44,58,60,66,190]. Other potential harms are associated more broadly with surveillance capitalism, the widespread collection and commodification of personal data by corporations [235], and disaster capitalism, the exploitation of situational crises to establish practices and policies at a time when individuals are distracted (emotionally and physically) or otherwise inattentive and less likely to consider the implications of or offer resistance to newly proposed changes in practice or policy [236]. There is a risk, for instance, that private companies might wrongly profit from the ability to predict when outbreaks will occur (eg, artificially increasing the price of medication) [31,44,45,55,67,75,134,174,190]. Furthermore, the potential and witnessed consequences of digital surveillance identified in the literature were frequently associated with the rapid innovation of digital technologies and rushed implementation of these technologies. Newlands et al [75] argued that rushed innovation is often a feature of disaster capitalism, as private interests seek to profit from crises; in the case of the COVID-19 pandemic, they noted that rushed innovation may heighten the risk of emergency legislation passed to make exceptions for use of these tools without full consideration of the ethical or legal implications [31,33,35,44,75,79,86,113,116,164,207].

**Human Rights and Civil Liberties and Impacts on Marginalized Groups**

The impacts of digital surveillance on human rights and civil liberties were noted by many authors (n=58) [10,23,24,28-30,34-36,38,43,44,53-55,61,67,71,78,89,91,98,104-106,109,111,115,125,131,135,137,154,157,173,188,190,204,209,210,212,215,220,223,226,230,232,233,237]. The erosion of freedom of movement [29,35,41,98,134], freedom of expression [75,130], freedom of association [41,130], and freedom of the press [33] was identified as a potential implication of digital surveillance [36,43,203,204,220]. A number of digital technologies used globally for public health surveillance were identified as imposing on rights and liberties, including digital proximity tracking technologies used in South Korea, Singapore, India, China, Hong Kong, and elsewhere [24,29,61,129,130]; use of wearable technologies (eg, ankle bracelets and trackable bracelets) in the United States, Hong Kong, Bahrain, and India [155,198]; and new advances in digital technology, including biosensors, artificial intelligence, and the Internet of Things [220]. Through the expansion of existing surveillance capacities, Russia and China, for instance, were described as using surveillance technologies, including mobile contact tracing apps, location data from telecommunications companies, and mobile apps for quarantine enforcement, in ways that did not contribute to proposed COVID-19 public health goals and impeded on human rights [125].

In addition to the impacts on human rights and civil liberties discussed above, the implications of digital surveillance for marginalized groups were discussed in many (n=46) publications [10,11,23,31,35,43,45,57,58,62,74,83,86,89,90,93,95,100,111,116,119,124-126,130,133,137,139,145,146,154,173,183,188,197,205,207,208,210,213,223,226,231,237,238]. Hendel et al [90] summarized this discussion through the recognition that the COVID-19 pandemic has not impacted all people in the same way, and that in addition to the need for epidemiological data on inequities generated by the pandemic response, there is a need to consider the ways in which digital surveillance could exacerbate existing vulnerabilities. A frequently cited example pertained to the use of mobile-based emergency alerts in South Korea sent out by public health authorities. These messages included intimate details regarding people infected with COVID-19 and allowed for ease of surveillance could exacerbate existing vulnerabilities. A frequently cited example pertained to the use of mobile-based emergency alerts in South Korea sent out by public health authorities. These messages included intimate details regarding people infected with COVID-19 and allowed for easy surveillance and may fear an exacerbation of marginalization through COVID-19 public health surveillance.

https://publichealth.jmir.org/2024/1/e47154
Ethical Principles for Digital Technology Design and Use

Authors of peer-reviewed and gray literature publications proposed several recommendations for designing and using digital technologies for public health surveillance in ways that minimize harmful consequences.

Fit for Purpose and Efficacy

Technologies should produce expected outcomes that will lead to actual benefits for public health and fulfill the needs of public health authorities. There should be evidence of this efficacy, and the potential advantages of digital technologies should be compared with other, less invasive options or nontechnological solutions [10,34-42,57,116-119,122,130,132,133,137,148,190,223,231,238,239].

Proportionality

The potential harms associated with digital surveillance should be proportional to the predicted or witnessed threats. Proportionality should include consideration of the disproportionate impacts of surveillance on marginalized groups [21,23,24,29,35,39,90,92,95,125,126,129,149,154,162,173,196,223,232,239].

Transparency

All aspects of digital surveillance should be performed such that they are easily observable. Information about how data are collected, used, and shared should be not only readily accessible but also easy to understand [29,31,35,39,43,44,53,63,73,95,97,102,104,113,117,118,120,125,132,149,154,161,173,178,190,192,196,203,231,232,240].

Temporary

There should be deadlines for data collection, use, and retention, and these should be communicated and upheld. In the case of emergency measures, there should be clearly defined indicators of when the emergency will be considered resolved and when emergency measures will end [21,29,35,36,41,43,53,70,73,75,97,105,122,126,129,149,154,161,173,179,196,226,229,232,237].

Rethinking Consent

There is a need to rethink what “consent” means in the contemporary digital surveillance context. In particular, there is an urgent need for discussions of how to avoid illusions of consent—for example, mandating surveillance to access certain spaces or use of data collected in covert ways for health-related surveillance [22,53,55,75,83,89,97,110,118,123,129,149,162,188,196,210,226,232,234].

Purpose Limitation

Collection and use of data should be limited to public health purposes for mitigating infectious disease spread. Data should not be repurposed [21,22,28,29,31,34,35,41,44,104,110,115,123,125,126,128,149,159-162,173,190,195,196,210,226,230,237].

Privacy and Security

Privacy and data security should be protected through privacy by design and technical tools as well as legislation regulating data collection and use [29,31,34-36,42,53,70,75,83,90,92,96,98,104-107,110,174,190,192,195,196,204,228,229,234,241,242].

Accountability

It should be clear who is responsible for data collection and use. Those engaging in surveillance should be answerable to the public, and individuals should be able to seek redress for harms [29,31,35,44,53,68,83,124,149,160,190,203,230,232,237,240].

Civil Society and Public Engagement

Digital surveillance for public health should require ongoing participation by the public and input from those who will be under surveillance. All aspects of surveillance should be discussed and debated through a democratic process [29,31,53,111,116,118,132,205,220,230,239]. Several authors (n=30) raised concerns that the use of digital technologies for public health surveillance may have long-term consequences for governance, the state, and civil society [11,23,25,30,33,41,43-45,52,55,68,71,79,86,91,95,100,114,123,133,190,203-205,209-211,220,230].

Discussion

Principal Findings

In our review of the academic and gray literature published during the first year of the COVID-19 pandemic, we aimed to explore the use of digital technologies for surveillance; the predicted and witnessed short- and long-term consequences of digital surveillance; and the questions, concerns, findings, and discussions that emerged in this first pivotal year of the pandemic. We found that digital technology use was identified in more than 90 countries and regions for pandemic-related applications, including contact tracing, symptom monitoring, disease tracking, enforcing compliance with public health measures, and supporting the direct provision of health care. We also found a number of consequences related to digital surveillance identified in the literature, including implications for data security and privacy, the potential for function and mission creep, consequences of private sector involvement in surveillance, implications for human rights and civil liberties, and impacts on marginalized groups. Finally, we also discovered ethical principles for digital technology design and use proposed throughout the literature.

Concerns regarding the potential consequences of digital surveillance often intersected. For instance, risks to privacy associated with digital surveillance were discussed by framing privacy as a human right, which, if compromised, may also lead to the undermining of other rights, such as freedom of movement, expression, and association. Likewise, the potential for mission and function creep and the expansion of surveillance
were linked to concerns around governance during the COVID-19 pandemic and whether this state of exception will extend beyond the crisis. Notably, discussions regarding mission and function creep drew parallels between other historical instances in which surveillance both extended beyond the initial crisis and captured data for uses other than those originally intended, such as the 9/11 crisis. These discussions highlight the fact that there is significant overlap between public health surveillance and other forms of surveillance (eg, criminal–legal and commercial) in both their objectives and their mechanisms. The blurred line between public health surveillance and other forms of surveillance is rendered particularly visible by digital surveillance, as the use of digital technologies for data collection and analysis highlights the ubiquitous, ongoing, and increasingly normalized nature of surveillance, which is likely to extend far beyond the COVID-19 pandemic.

The associations between public health and other institutional relations and modes of surveillance are captured in the authors’ descriptions of the impacts of surveillance on individuals’ human rights and civil liberties. In South Korea, the use of mobile-based emergency alerts sent out by public health authorities led to stigmatization and discrimination against lesbian, gay, bisexual, transgender, intersex, queer/questioning, and asexual communities; while these alerts were deployed in support of the public health response to the COVID-19 pandemic, the objectives of reducing disease spread and protecting public health do not detract from the invasive and coercive nature of these alert messages, nor their impacts on marginalized groups. Likewise, many authors raised concerns about the capitalist exploitation of public health crises and the repurposing of data to serve private interests. These findings mark the importance of attending to the intersections between public health, public health surveillance, and other forms of social control. These discussions are situated within long-standing debates around freedom of choice and other human rights, the coercive character of public health, and ethical quandaries brought to light by digital surveillance, as explored in this paper.

Recent publications have underscored the urgency of attending to questions of human rights related to digital surveillance during the COVID-19 pandemic, particularly in the case of marginalized groups. Abdelrahman [243], for instance, describes the impacts of digital surveillance during the COVID-19 pandemic on vulnerable groups, including African minority populations in China who had surveillance cameras installed outside their homes by the state. Marshall explores the potential implications of increased digital surveillance during the pandemic for queer sex workers, as enhanced surveillance through closed-circuit television cameras and facial recognition reduced their ability to work anonymously and increased their risk of being criminalized [244]. Our scoping review contributes to these ongoing discussions through a thematic, qualitative analysis that identifies the “slippage” between public health surveillance and other forms of surveillance, including their impacts on marginalized groups, regardless of the stated objectives of surveillance.

One of the most interesting findings to emerge from our review is the emphasis on the differences between digital surveillance and other forms of public health surveillance. While surveillance (including global health surveillance) has long been a critical function of public health [245], the emergence of the COVID-19 pandemic as the first global pandemic of the “digital age” has meant that the uncertainty of navigating this public health crisis was exacerbated by the unprecedented development and use of digital technologies for surveillance [6]. Given the unique context in which the COVID-19 pandemic spread and in which digital technologies were used for public health surveillance, some authors drew on theoretical concepts (eg, Deleuze’s theory of the control society and Foucault’s concept of the panopticon) to analyze the normalization of digital surveillance during the pandemic as these technologies became increasingly ubiquitous [79]. The consequences of emerging differences between digital surveillance and traditional forms of public health surveillance (eg, the huge amount of data collected, the ubiquity and pervasiveness of surveillance, and evolving techniques of data analysis, including machine learning) remain to be fully explored [6,246]. Further research is needed to elucidate the impacts of the normalization and omnipresence of digital surveillance and the implications raised in this review, including data privacy and security, creep, private sector involvement, and human rights.

In light of the ramifications associated with digital public health surveillance, the value of these forms of surveillance must be carefully weighed. Recent publications have suggested that the wide range of digital technologies implemented during the COVID-19 pandemic for surveillance purposes faced a number of barriers preventing their successful implementation, and while many of these digital innovations have yet to be formally evaluated or assessed, there is significant uncertainty around their value [8,9]. It is difficult to determine the impact of digital surveillance on disease transmission, particularly given the challenges around technology uptake, implementation, and consistency. Likewise, it is equally difficult to draw comparisons between countries given the various technologies used, the public health and health care infrastructures in which they were deployed, and the varying objectives toward which these technologies were used, including commercial, criminal–legal, and state objectives. In the face of this uncertainty, it will be important to continue to carefully assess the value of digital surveillance and its intended and unintended consequences. Also critical will be a holistic approach to surveillance that accounts for the blurred boundaries between public health, public health surveillance, and other forms of surveillance and social control.

Our scoping review provides a snapshot of peer-reviewed and gray literature publications from the first year of the COVID-19 pandemic. The scope of the publications included in this review is global, and we included authors affiliated with institutions across diverse regions, signaling preoccupation with these topics worldwide. The large number of publications retained also points to the urgency of these concerns and interest in digital public health surveillance in both academic and nonacademic spaces. Ongoing discussions continue to explore and generate appropriate practices, policies, and legislation to address the challenges associated with digital surveillance [6,246]. The findings of this scoping review contribute to our knowledge of
digital technologies used for public health surveillance during the COVID-19 pandemic, the potential and witnessed implications of digital surveillance, and ethical principles for technology design and use. This information is critical for leveraging digital public health surveillance in ways that are ethical and that use data to improve health while minimizing potential marginalization associated with surveillance.

Limitations
While our review provides important insights into the use of digital technologies for public health surveillance, this study has some limitations. Most significant is the restriction of our focus to the first year of the COVID-19 pandemic, which excludes the use of digital technologies for surveillance after this first year of the pandemic and during other public health crises, including other infectious disease outbreaks. Limiting the review to English-language documents may have also limited our analysis of global digital surveillance. As the COVID-19 pandemic continues and other outbreaks, such as mpox, emerge, this publication represents an early appraisal of existing knowledge. However, the occurrence of new infectious disease outbreaks and the use of new modes of digital surveillance underscore the urgency of reviewing this literature.

Conclusions
This scoping review explored the potential and witnessed short- and long-term consequences associated with the use of digital technologies for public health surveillance during the COVID-19 pandemic identified in the peer-reviewed and gray literature published during the first year of the pandemic. The review found evidence of concerns raised around risks to human rights and civil liberties, the potential normalization of surveillance, the expansion of the state of emergency of the COVID-19 pandemic, and data security and privacy. The literature also included recommendations for more ethical digital surveillance, including transparency, limited retention and use of data, proportionality, and ongoing efforts to introduce adequate legislation to regulate surveillance.

Although this scoping review focuses on the COVID-19 pandemic, the findings have implications for digital surveillance during other infectious disease outbreaks. The rapid pace of digital technology use and development requires consideration of the impacts of rushed innovation and how legislative, public health, technological, and other mechanisms can be used in striving toward more ethical data collection and use. The consequences and potential harms associated with digital surveillance identified in this review raise important questions about the use of digital technologies for public health surveillance and how these technologies can support public health without impinging on rights, liberties, or democratic processes.

Acknowledgments
We would like to thank Meagan Stanley for her assistance in designing the search protocol for this review. This work was supported by the University of Western Ontario FHS Research Grant 2020. The funder had no role in the study design; the collection, management, analysis, or interpretation of the data; the writing of the report; or the decision to submit the manuscript for publication.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews) checklist. [PDF File (Adobe PDF File), 104 KB - publichealth_v10i1e47154_app1.pdf]

Multimedia Appendix 2
Table indicating, as identified by publications included in this review, digital technologies used for surveillance by country or region. [PDF File (Adobe PDF File), 47 KB - publichealth_v10i1e47154_app2.pdf]

References


72. Hsu J. The Dilemma of contact-tracing apps: can this crucial technology be both effective and private? IEEE Spectr 2020;57(10):56-59 [FREE Full text]


120. Contact tracing: to centralise or not to centralise, that is the question. Ada Lovelace Institute. 2020. URL: https://www.adalovelaceinstitute.org/contact-tracing-to-centralise-or-not-to-centralise-that-is-the-question/ [accessed 2024-04-30]


156. Tsang S. Here are the contact tracing apps being deployed around the world. International Association of Privacy Professionals. 2020. URL: https://iapp.org/news/a/heres-the-contact-tracing-apps-being-employed-around-the-world/ [accessed 2022-04-22]


Abbreviations

**PRISMA**: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

**PRISMA-P**: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols

**PRISMA-ScR**: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

©Leigha Comer, Lorie Donelle, Bradley Hiebert, Maxwell J Smith, Anita Kothari, Saverio Stranges, Jason Gilliland, Jed Long, Jacquelyn Burkell, Jacob J Shelley, Jodi Hall, James Shelley, Tommy Cooke, Marionette Ngole Dione, Danica Facca. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 24.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Review

Drug-Resistant Tuberculosis Case-Finding Strategies: Scoping Review

Susanna S van Wyk¹, MD, MSc; Marriott Nliwasa², PhD; Fang-Wen Lu³, MSc; Chih-Chan Lan³, MSc; James A Seddon¹,², MD, PhD; Graeme Hoddinott⁴,⁵, PhD; Lario Viljoen⁴, PhD; Gunar Günther⁷,⁸, MD, PhD; Nunurai Ruswa⁹, MD, MSc; N Sarita Shah¹⁰, MD, MPH; Mareli Claassens⁴,⁵,⁸, MD, PhD

¹Centre for Evidence Based Health Care, Division of Epidemiology and Biostatistics, Department of Global Health Stellenbosch University, Cape Town, South Africa
²Helse Nord Tuberculosis Initiative, Kamuzu University of Health Sciences, Blantyre, Malawi
³Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan
⁴Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
⁵Department of Infectious Disease, Imperial College London, London, United Kingdom
⁶School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
⁷Department of Pulmonary Medicine and Allergology, Inselspital, Bern University Hospital, Bern, Switzerland
⁸Department of Human, Biological & Translational Medical Science, School of Medicine, University of Namibia, Windhoek, Namibia
⁹National TB and Leprosy Programme, Ministry of Health and Social Services, Windhoek, Namibia
¹⁰Departments of Epidemiology and Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Mareli Claassens, MD, PhD
Department of Human, Biological & Translational Medical Science
School of Medicine
University of Namibia
Bach Street
Windhoek, 10005
Namibia
Phone: 264 612063111
Email: mela@sun.ac.za

Abstract

Background: Finding individuals with drug-resistant tuberculosis (DR-TB) is important to control the pandemic and improve patient clinical outcomes. To our knowledge, systematic reviews assessing the effectiveness, cost-effectiveness, acceptability, and feasibility of different DR-TB case-finding strategies to inform research, policy, and practice, have not been conducted and the scope of primary research is unknown.

Objective: We therefore assessed the available literature on DR-TB case-finding strategies.

Methods: We looked at systematic reviews, trials, qualitative studies, diagnostic test accuracy studies, and other primary research that sought to improve DR-TB case detection specifically. We excluded studies that included patients seeking care for tuberculosis (TB) symptoms, patients already diagnosed with TB, or were laboratory-based. We searched the academic databases of MEDLINE, Embase, The Cochrane Library, Africa-Wide Information, CINAHL (Cumulated Index to Nursing and Allied Health Literature), Epistemonikos, and PROSPERO (The International Prospective Register of Systematic Reviews) using no language or date restrictions. We screened titles, abstracts, and full-text articles in duplicate. Data extraction and analyses were carried out in Excel (Microsoft Corp).

Results: We screened 3646 titles and abstracts and 236 full-text articles. We identified 6 systematic reviews and 61 primary studies. Five reviews described the yield of contact investigation and focused on household contacts, airline contacts, comparison between drug-susceptible tuberculosis and DR-TB contacts, and concordance of DR-TB profiles between index cases and contacts. One review compared universal versus selective drug resistance testing. Primary studies described (1) 34 contact investigations, (2) 17 outbreak investigations, (3) 3 airline contact investigations, (4) 5 epidemiological analyses, (5) 1 public-private partnership program, and (6) an e-registry program. Primary studies were all descriptive and included cross-sectional and retrospective reviews.
of program data. No trials were identified. Data extraction from contact investigations was difficult due to incomplete reporting of relevant information.

**Conclusions:** Existing descriptive reviews can be updated, but there is a dearth of knowledge on the effectiveness, cost-effectiveness, acceptability, and feasibility of DR-TB case-finding strategies to inform policy and practice. There is also a need for standardization of terminology, design, and reporting of DR-TB case-finding studies.

(JMIR Public Health Surveill 2024;10:e46137) doi:10.2196/46137

**KEYWORDS**
tuberculosis; drug-resistant tuberculosis; drug-resistant tuberculosis case finding; drug-resistant tuberculosis case detection; drug-resistant tuberculosis screening; drug-resistant tuberculosis contact investigation; scoping review; TB symptom; anti-tuberculosis drug; strategies; multidrug-resistant; systematic review; drug resistant; drug resistance; medication; tuberculosis; diagnosis; screening

**Introduction**

With the emergence of *Mycobacterium tuberculosis* strains resistant to first-line antituberculosis drugs, strategies to control tuberculosis (TB) have become even more challenging [1]. It is estimated that almost half a million people developed rifampicin-resistant TB, of which 78% had multidrug-resistant tuberculosis (MDR-TB) in 2019 [2]. Although drug-resistant tuberculosis (DR-TB) is not as prevalent as drug-susceptible tuberculosis (DS-TB), it is more difficult to diagnose, treatment is longer and more toxic, outcomes are worse, and costs are higher.

Finding individuals with DR-TB and initiating treatment as early as possible is important to improve patient clinical outcomes and to break the chain of transmission to help control the pandemic. Despite new diagnostic technologies, only a third of the estimated number of people who developed DR-TB initiated treatment in 2020 [3].

TB can be detected after the patient presents passively to health services or follows one of several different screening pathways depending on the case-finding strategy of a TB program [4]. Pathways can also be enhanced through several activities such as health promotion in the community, improved access to TB diagnostic services, or training of health workers to identify presumptive TB at general health services. Multiple activities often result in complex interventions and heterogeneous trials that are difficult to meta-analyze in systematic reviews [5,6].

To our knowledge, systematic reviews assessing the effectiveness, cost-effectiveness, acceptability, and feasibility of different DR-TB case-finding strategies to inform research, policy, and practice, have not been conducted and it is unknown whether enough research exists to conduct such reviews. It is also unknown whether case-finding strategies are similar for DR-TB and DS-TB and whether we can draw on findings from DS-TB reviews to inform decisions on DR-TB case-finding strategies.

Scoping reviews are useful for scoping the literature and to clarify concepts [7,8]. We therefore conducted a scoping review to assess whether enough research exists for a systematic review, to identify priority questions for such a review, and to clarify which case-finding strategies exist for DR-TB specifically.

**Methods**

**Reporting Guidelines and Protocol**
The Arksey and O’Malley framework [9], Levac et al [10], and the Joanna Briggs Institute scoping review methodology [8] guided methods for this scoping review. The review is reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [11]. See Multimedia Appendix 1 for the completed PRISMA-ScR checklist. The protocol for this review was published in *JMIR Research Protocols* [12].

**Defining the Research Question**
The question for our review was, what literature is available on DR-TB case finding and which case-finding strategies are described? We looked at studies that had sought to improve DR-TB case detection.

**Eligibility Criteria**

Textbox 1 lists the inclusion and exclusion criteria for participants, concept, outcome, context, and study design.
Textbox 1. Eligibility criteria.

Inclusion criteria

• Participants
  • Participate regardless of symptoms, for example, contacts, people living with HIV attending HIV care, whole communities
• Concept/Intervention
  • Strategies aiming to improve or enhance participants’ pathways to drug-resistant tuberculosis (DR-TB) case detection specifically
• Outcome
  • Patients diagnosed with tuberculosis (TB)
• Context
  • Community
  • Primary, secondary, or tertiary care
  • Study design
  • Primary studies
• Systematic reviews
• Qualitative studies, where the experiences of individuals who receive the intervention or those who provide the intervention are investigated
• Studies of diagnostic test accuracy if the study describes a DR-TB screening strategy
• Trials comparing different screening or diagnostic tools within a DR-TB case-finding intervention

Exclusion criteria

• Participants
  • Patients with TB symptoms seeking care
  • Patients diagnosed with TB
  • Laboratory samples/isolates
• Concept/Intervention
  • Intervention strategies aiming to improve TB case finding in general, even if they do report the yield of people with DR-TB
• Outcome
  • No report of patients diagnosed with TB
• Context
  • Laboratory based
  • Study design
  • Meta-reviews (review of reviews)
  • Narrative reviews
• Editorial
  • Opinion articles
  • Meeting summaries
• Guidelines
• Prevalence surveys, except if the survey includes an intervention strategy to find individuals with DR-TB specifically
• Conference abstracts

Identifying Relevant Studies

With assistance from an information specialist, we searched the academic databases of MEDLINE (PubMed), Embase (Ovid), The Cochrane Library, Africa-Wide Information (EBSCOhost), CINAHL (EBSCOhost), Epistemonikos, and PROSPERO (The International Prospective Register of Systematic Reviews) using no language or date restrictions. These searches were conducted on August 31, 2021, and after an initial peer review of this article, an updated search was conducted on January 11, 2024. The search string included combinations of the following 3 domains, that are (1) terms related to “TB”; (2) terms related to “drug resistance”; and (3) terms related to “case finding,”
“case detection,” “screening,” “contact investigation,” and “contact tracing.”

Search strategies from each electronic database are detailed per search date in Multimedia Appendix 2.

Study Selection
We used Rayyan systematic review software [13] to screen titles, abstracts, and full-text articles. Decisions were blinded, except when reviewing conflicts. Reviewers screened abstracts in duplicate for inclusion. Conflicts were resolved through discussion. Full-text articles were also screened in duplicate. Disagreements were resolved through discussion to determine final inclusion.

Charting the Data
We developed a data extraction form in Excel (Microsoft Corp). The data extraction form was applied to all primary research reports to collect standard information on each study. Textbox 2 lists the information that was collected.

One reviewer extracted data from included papers and a second reviewer (SSvW) checked the extracted data. Reviewers met regularly to determine whether their approach was consistent and in line with the research question.

Textbox 2. Information collected on each study.

<table>
<thead>
<tr>
<th>Information collected on each study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors, journal, year of publication</td>
</tr>
<tr>
<td>Aim or purpose of the research</td>
</tr>
<tr>
<td>Study design</td>
</tr>
<tr>
<td>Country</td>
</tr>
<tr>
<td>Income</td>
</tr>
<tr>
<td>Tuberculosis prevalence</td>
</tr>
<tr>
<td>HIV prevalence</td>
</tr>
<tr>
<td>Urban or rural setting</td>
</tr>
<tr>
<td>Participants</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>HIV status</td>
</tr>
<tr>
<td>Other reported risk factors</td>
</tr>
<tr>
<td>Target group and how the group was identified if applicable</td>
</tr>
<tr>
<td>Interventions</td>
</tr>
<tr>
<td>All components (activities) of the intervention</td>
</tr>
<tr>
<td>Types of providers</td>
</tr>
<tr>
<td>Screening and diagnostic tools used</td>
</tr>
<tr>
<td>Treatment support, including preventive therapy</td>
</tr>
<tr>
<td>Outcomes assessed</td>
</tr>
</tbody>
</table>

Collating, Summarizing, and Reporting the Results
We provide a narrative report with supporting tables to summarize the data. Table 1 contains definitions we used in charting, collating, summarizing, and reporting our results. A systems-based logic model developed from a synthesis of DS-TB case-finding strategies (Figure 1) was used as a framework to describe different strategies and resulting pathways (care-seeking pathways or screening pathways).

Quality appraisal was not conducted, because this is a scoping review and our interest is in the existing evidence base, regardless of study design and quality.
Table 1. Definitions.

<table>
<thead>
<tr>
<th>Terms</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR-TB&lt;sup&gt;a&lt;/sup&gt;</td>
<td>All types of DR-TB that include DR-TB with resistance to one first-line drug, MDR-TB&lt;sup&gt;b&lt;/sup&gt;, XDR-TB&lt;sup&gt;c&lt;/sup&gt;, and any other DR-TB reported by the authors.</td>
</tr>
<tr>
<td>Systematic screening for TB&lt;sup&gt;d&lt;/sup&gt;</td>
<td>“The systematic identification of people with suspected (presumptive) TB disease, in a predetermined target group, using tests, examinations, or other procedures that can be applied rapidly. Among those screened positive, the diagnosis needs to be established by one or several diagnostic tests and additional clinical assessments, which together have high accuracy.” [14]</td>
</tr>
<tr>
<td>A screening tool</td>
<td>Tests, examinations, or other procedures used for systematic screening for TB disease. Examples of TB screening tools include a structured symptom-based questionnaire, CXR&lt;sup&gt;e&lt;/sup&gt;, or an algorithm [4]. Algorithms may include sequential or parallel tests. With sequential tests, only those who screen positive with the initial test receive a second test. With parallel tests, those who screen positive on any of the tests are regarded as screen positives.</td>
</tr>
<tr>
<td>A diagnostic tool</td>
<td>Tests, examinations, or other procedures used to establish a diagnosis of TB disease in people identified with presumptive TB. Examples of TB diagnostic tools include a clinical algorithm, sputum smear microscopy, Xpert MTB/RIF (Cepheid Inc), or culture [4].</td>
</tr>
<tr>
<td>TB symptom</td>
<td>Any TB symptom, for example, cough, fever, night sweats, weight loss, or combination of TB symptoms as defined by the study authors.</td>
</tr>
<tr>
<td>Care seeking</td>
<td>People seeking care for a perceived health problem.</td>
</tr>
<tr>
<td>TB care seeking</td>
<td>People seeking care for TB symptoms specifically.</td>
</tr>
<tr>
<td>A risk group</td>
<td>Any group of people in whom the prevalence or incidence of TB is significantly higher than in the general population. Examples of risk groups include a whole population within a geographical area or TB contacts [15].</td>
</tr>
<tr>
<td>A clinical risk group</td>
<td>Individuals diagnosed with a specific disease or condition that increases their risk for TB, for example, people living with HIV (PLHIV).</td>
</tr>
<tr>
<td>Presumptive TB</td>
<td>Presumptive TB is identified when a provider identifies a patient with suspected TB disease. In the context of screening, a person who screens positive is a patient with presumptive TB.</td>
</tr>
<tr>
<td>Passive case finding</td>
<td>Care-seeking pathway without TB screening, that is, the green and black dashed pathways in Figure 1 [16].</td>
</tr>
<tr>
<td>Passive case finding with an element of systematic screening or triage</td>
<td>TB screening at general health services, that is, the green pathway in Figure 1.</td>
</tr>
<tr>
<td>Enhanced case finding</td>
<td>TB health promotion with or without TB screening.</td>
</tr>
<tr>
<td>Active case finding</td>
<td>TB screening at TB screening services or at home, work, or school, that is, the blue and orange pathways in Figure 1. If the target group is TB contacts, this can also be referred to contact tracing or contact investigation.</td>
</tr>
<tr>
<td>Intensified case finding</td>
<td>TB screening of a clinical risk group, for example, people living with HIV (ie, the gray pathway in Figure 1).</td>
</tr>
</tbody>
</table>

<sup>a</sup>DR-TB: drug-resistant tuberculosis.
<sup>b</sup>MDR-TB: multidrug-resistant tuberculosis.
<sup>c</sup>XDR-TB: extensively drug-resistant tuberculosis.
<sup>d</sup>TB: tuberculosis.
<sup>e</sup>CXR: chest radiography.
**Results**

**Overview of the Available Literature**

We screened 3646 titles and abstracts and 236 full-text articles. We identified 6 systematic reviews and 61 primary studies (Figure 2) for inclusion. We divided primary studies into 6 different categories (themes) and described each category in more detail below. Table 2 gives an overview of the categories and references to further detail.
Figure 2. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

Table 2. Overview of categories into which included studies were divided.

<table>
<thead>
<tr>
<th>Type of study</th>
<th>Articles</th>
<th>Further detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic reviews</td>
<td>n=6</td>
<td>Table 3</td>
</tr>
<tr>
<td><strong>Primary studies (N=61)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close- or household-contact investigations</td>
<td>n=34 (56%)</td>
<td>Multimedia Appendix 3 and Table 4</td>
</tr>
<tr>
<td>Outbreak investigations</td>
<td>n=17 (28%)</td>
<td>Table 5</td>
</tr>
<tr>
<td>Airline contact investigations</td>
<td>n=3 (5%)</td>
<td></td>
</tr>
<tr>
<td>Epidemiological analyses</td>
<td>n=5 (8%)</td>
<td></td>
</tr>
<tr>
<td>Public-private partnership program</td>
<td>n=1 (2%)</td>
<td></td>
</tr>
<tr>
<td>E-registry program</td>
<td>n=1 (2%)</td>
<td></td>
</tr>
</tbody>
</table>

**Systematic Reviews**

We identified 6 systematic reviews. Outcomes were descriptive and none of the reviews identified any randomized controlled trials. In 5 of the 6 reviews, the date of the last search was more than 5 years ago (Table 3). In reviews with the yield of TB disease as an outcome, the denominator was reported as the number of contacts evaluated or screened; however, specific definitions for “evaluated” or “screened” were not reported and the yield for a specific screening or diagnostic strategy is unknown.
Table 3. Overview of the included systematic reviews.

<table>
<thead>
<tr>
<th>Review</th>
<th>Primary outcome</th>
<th>Date of last search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abubakar [17]</td>
<td>Number of contacts screened, and number of individuals with TB(^a) infection and TB disease identified</td>
<td>November 2009</td>
</tr>
<tr>
<td>Fox et al [18]</td>
<td>Yield of TB disease and TB infection for both DS-TB(^b) and DR-TB(^c) source cases</td>
<td>October 2011</td>
</tr>
<tr>
<td>Kodama et al [20]</td>
<td>Relative risk ratio of TB disease in DS-TB contacts compared with DR-TB contacts</td>
<td>Not reported</td>
</tr>
<tr>
<td>Svadzian et al [21]</td>
<td>Proportion of cases from those evaluated through universal testing (all individuals in the study received DST(^d)) and those evaluated through selective testing (only the high-risk group received DST)</td>
<td>June 2019</td>
</tr>
<tr>
<td>Chiang et al [22]</td>
<td>Percentage of secondary cases whose <em>Mycobacterium tuberculosis</em> strains were resistant to the same drugs as strains from the index cases</td>
<td>July 2018</td>
</tr>
</tbody>
</table>

\(^a\)TB: tuberculosis.  
\(^b\)DS-TB: drug-susceptible tuberculosis.  
\(^c\)DR-TB: drug-resistant tuberculosis.  
\(^d\)DST: drug-susceptibility testing.

### Primary Studies

We identified 61 primary studies that were not included in any of the above reviews. Primary studies were descriptive and included cross-sectional studies, prospective studies, and retrospective reviews of program data. No trials were identified. Thirty-four studies were close-contact or household-contact investigations, 17 were outbreak investigations, 3 were airline contact investigations, 5 were epidemiological analyses, 1 described a private-public partnership program, and I assessed the feasibility and acceptability of an e-registry program (Table 2). Case-finding pathways were seldom described clearly, for example, whether contacts were invited for screening regardless of symptoms (Figure 1, blue pathway), whether all contacts were screened for TB at home (Figure 1, orange pathway), or whether those who experienced TB symptoms were invited for further tests (Figure 1, black dashed pathway).

### Close-Contact or Household-Contact Investigations

Countries where contact investigations were conducted included South Africa (n=6), India (n=5), Pakistan (n=4), Australia (n=2), the United States (n=1), Ethiopia (n=2), Myanmar (n=2), Thailand (n=1), France (n=1), Vietnam (n=1), Papua New Guinea (n=1), Armenia (n=1), the United Kingdom (n=1), Spain (n=1), South Korea (n=1), Oman (n=1), and Tajikistan (n=1). Two multicountry studies were conducted in Botswana, Brazil, Haiti, Kenya, Peru, South Africa, and Thailand. Data extraction from contact investigations was difficult due to incomplete reporting of relevant information, such as the total number of source cases or the number of cases tested for drug susceptibility (Table 4). Screening and diagnostic tools were not well reported and often lacked consistent or standardized use. Although investigations focused on contacts who had been exposed to DR-TB, drug-susceptibility testing (DST) was seldom reported. Case-finding pathways were also not clearly described. Some contacts were followed up over 1-2 years and some were only evaluated at baseline. Lack of or inconsistent reporting of this relevant data results in an unknown or inconsistent denominator when calculating the yield of screening the contacts of individuals with TB and makes it challenging to pool results or compare different case-finding strategies. There was also little consistency in the use of definitions. Source cases were often defined as “registered MDR-TB or extensively drug-resistant tuberculosis (XDR-TB) cases” without knowledge of how they were diagnosed. Several different definitions for “close contact” or “household contact” were reported. Some definitions were broad, for example, “people living with or having regular daily interaction with the MDR-TB source case” [23], while other definitions were more specific, for example, “a person who had shared the same enclosed living space for one or more nights a week, or for frequent or extended periods of time during the day, with the index patient during the 3 months before the current treatment episode began” [24-26]. The latter definition was used more often. See Multimedia Appendix 3 for more details.
Table 4. Data from DR-TB<sup>a</sup> contact investigation studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Source/Index cases</th>
<th>Contacts</th>
<th>Screening</th>
<th>TB&lt;sup&gt;b&lt;/sup&gt; diagnosis</th>
<th>DS-TB&lt;sup&gt;c&lt;/sup&gt; diagnosis</th>
<th>DR-TB diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total identified</td>
<td>Studied</td>
<td>Total identified</td>
<td>Positive screen</td>
<td>Evaluated</td>
<td>Total TB</td>
</tr>
<tr>
<td>Mohamma-di et al [27]</td>
<td>Not reported</td>
<td>13</td>
<td>Not reported</td>
<td>140</td>
<td>Not reported</td>
<td>0</td>
</tr>
<tr>
<td>Tuberculosis Research Centre, Indian Council of Medical Research [28]</td>
<td>Not reported</td>
<td>209 INH&lt;sup&gt;d&lt;/sup&gt; resistant at intake</td>
<td>779 at intake and 8358 over 15 years</td>
<td>Not reported</td>
<td>22 over 15 years of f/u&lt;sup&gt;e&lt;/sup&gt; and 260 per 100,000 person-years in INH-resistant HH&lt;sup&gt;f&lt;/sup&gt; contacts</td>
<td>18</td>
</tr>
<tr>
<td>Denholm et al [29]</td>
<td>47</td>
<td>47</td>
<td>570</td>
<td>49 LTBI&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Not reported</td>
<td>2</td>
</tr>
<tr>
<td>Seddon et al [23]</td>
<td>Not reported</td>
<td>Not reported</td>
<td>281</td>
<td>228</td>
<td>Not reported (102 LTBI)</td>
<td>15</td>
</tr>
<tr>
<td>Adler-Shohet et al [30]</td>
<td>1</td>
<td>1</td>
<td>Not reported</td>
<td>118</td>
<td>31 TST&lt;sup&gt;h&lt;/sup&gt; pos&lt;sup&gt;i&lt;/sup&gt; (21 initially and 10 at repeat)</td>
<td>0</td>
</tr>
<tr>
<td>Garcia-Prats et al [31]</td>
<td>1</td>
<td>1</td>
<td>38</td>
<td>34</td>
<td>None</td>
<td>Not reported</td>
</tr>
<tr>
<td>Titiyos et al [32]</td>
<td>508</td>
<td>508</td>
<td>155 family members in households of 29 symptomatic contacts</td>
<td>155</td>
<td>Unclear (29 symptomatic contacts were initially identified and evaluated)</td>
<td>16</td>
</tr>
<tr>
<td>Arnold et al [33]</td>
<td>1</td>
<td>1</td>
<td>35</td>
<td>33</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>García et al [34]</td>
<td>1</td>
<td>1</td>
<td>39</td>
<td>39</td>
<td>19 Mantoux pos of which 1 had CXR&lt;sup&gt;l&lt;/sup&gt; changes</td>
<td>Not reported</td>
</tr>
<tr>
<td>Javaid et al [35]</td>
<td>200</td>
<td>154</td>
<td>Not reported</td>
<td>610</td>
<td>218 symptoms, 51 AFB&lt;sup&gt;m&lt;/sup&gt;-positive, and Nr&lt;sup&gt;n&lt;/sup&gt; with abnormal CXR not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Fournier et al [36]</td>
<td>68</td>
<td>32</td>
<td>84</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Golla et al [37]</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>229</td>
<td>Not reported</td>
<td>226</td>
</tr>
<tr>
<td>Study Name</td>
<td>Source/Index cases</td>
<td>Contacts</td>
<td>Screening</td>
<td>TB&lt;sup&gt;b&lt;/sup&gt; diagnosis</td>
<td>DS-TB&lt;sup&gt;c&lt;/sup&gt; diagnosis</td>
<td>DR-TB diagnosis</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>--------------------</td>
<td>----------</td>
<td>-----------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Lee et al [38]</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>2 asymptomatic with minimal nodules in baseline chest CT&lt;sup&gt;ab&lt;/sup&gt; scan</td>
<td>0</td>
</tr>
<tr>
<td>Chatla et al [39]</td>
<td>1602</td>
<td>1602</td>
<td>4858</td>
<td>4771</td>
<td>793</td>
<td>19</td>
</tr>
<tr>
<td>Dayal et al [40]</td>
<td>Not reported</td>
<td>43</td>
<td>Not reported</td>
<td>100</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Hiruy et al [24]</td>
<td>111</td>
<td>111</td>
<td>340</td>
<td>331</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>Huerga et al [41]</td>
<td>265</td>
<td>111</td>
<td>198</td>
<td>Not reported</td>
<td>150 at baseline and 138 at f/u</td>
<td>Not reported</td>
</tr>
<tr>
<td>Boonthanapat et al [42]</td>
<td>91</td>
<td>43</td>
<td>174</td>
<td>70 had screening records</td>
<td>3 abnormal CXRs</td>
<td>Not reported</td>
</tr>
<tr>
<td>Hoang et al [43]</td>
<td>112</td>
<td>99</td>
<td>496</td>
<td>325 at baseline and 160 at f/u</td>
<td>48 at baseline and 27 at f/u</td>
<td>1 (no TB at f/u)</td>
</tr>
<tr>
<td>Honjepari et al [44]</td>
<td>67</td>
<td>67 total (only 25 DR-TB cases)</td>
<td>635 total and 23 DR-TB contacts</td>
<td>156</td>
<td>114</td>
<td>5</td>
</tr>
<tr>
<td>Kigozi et al [45]</td>
<td>Not reported</td>
<td>92</td>
<td>297 (only 6 contacts of MDR-TB cases)</td>
<td>259 (6)</td>
<td>102 (1)</td>
<td>48 (1)</td>
</tr>
<tr>
<td>Phylo et al [26]</td>
<td>556</td>
<td>556</td>
<td>1908</td>
<td>1134</td>
<td>344 presumptive TB</td>
<td>15</td>
</tr>
<tr>
<td>Gupta et al [46]</td>
<td>308</td>
<td>284</td>
<td>1016</td>
<td>1007</td>
<td>186 presumptive TB and 213 others (399 in total)</td>
<td>27 (6 bacteriologically and 21 clinically)</td>
</tr>
<tr>
<td>Kyaw et al [25]</td>
<td>Not reported</td>
<td>210</td>
<td>Not reported</td>
<td>620</td>
<td>228 signs and symptoms and 169/969 abnormal CXR</td>
<td>24 (7 bacteriologically and 17 clinically)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Briefly describes the screening protocol and findings. <sup>b</sup>Specific details about TB diagnosis. <sup>c</sup>Details related to DS-TB and DR-TB diagnosis.
<table>
<thead>
<tr>
<th>Study</th>
<th>Source/Index cases</th>
<th>Contacts</th>
<th>Screening</th>
<th>TB$^b$ diagnosis</th>
<th>DS-TB$^c$ diagnosis</th>
<th>DR-TB diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total identified</td>
<td>Studied</td>
<td>Total identified</td>
<td>Screened</td>
<td>Positive screen</td>
<td>Evaluated</td>
</tr>
<tr>
<td>Malik et al [47]</td>
<td>Not reported</td>
<td>100</td>
<td>800 (8 on TB treatment)</td>
<td>737</td>
<td>402 symptoms or &lt;18 years or DM$^p$ or HIV or low BMI$^q$</td>
<td>326</td>
</tr>
<tr>
<td>Paryani et al [48]</td>
<td>129</td>
<td>109</td>
<td>518 (22 with TB already on treatment and 2 diagnosed at baseline)</td>
<td>495 (400 entered f/u)</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Shadrach et al [49]</td>
<td>Not reported</td>
<td>87</td>
<td>Not reported</td>
<td>Not reported</td>
<td>285</td>
<td>271</td>
</tr>
<tr>
<td>van de Water et al [50]</td>
<td>284</td>
<td>284</td>
<td>959</td>
<td>Not reported</td>
<td>336</td>
<td>Not reported</td>
</tr>
<tr>
<td>Chang et al [51]</td>
<td>Not reported</td>
<td>55</td>
<td>247</td>
<td>215</td>
<td>8 abnormal CXR</td>
<td>Not reported</td>
</tr>
<tr>
<td>Kim et al [52]</td>
<td>305</td>
<td>152 RR-TB</td>
<td>1324</td>
<td>303 children &lt;15 years</td>
<td>69 symptoms</td>
<td>Not reported</td>
</tr>
<tr>
<td>Ahmed et al [53]</td>
<td>329</td>
<td>324</td>
<td>1911</td>
<td>1734 symptom screen, 281 Xpert, and 123 CXR only</td>
<td>Not reported</td>
<td>All contacts were eligible for Xpert regardless of symptoms</td>
</tr>
<tr>
<td>Ahmed and Dadlani [54]</td>
<td>470</td>
<td>100 MDR-TB and 370 DS-TB</td>
<td>830</td>
<td>830</td>
<td>218 symptoms</td>
<td>102 GeneXpert, 76 Smear microscopy, and 11 CXR</td>
</tr>
<tr>
<td>Apolisi et al [55]</td>
<td>Not reported</td>
<td>48</td>
<td>146</td>
<td>112</td>
<td>19 symptoms</td>
<td>55 CXR only, 19 CXR and sputum, and 1 sputum only</td>
</tr>
</tbody>
</table>
Outbreak Investigations

The Dictionary of Epidemiology defines an outbreak as “an epidemic limited to localized increase in the incidence of disease, e.g., village, town, or closed institution” [57]. In the included studies, it was not always reported whether the number of identified patients was more than expected over a particular period. It was not always clear if a study was indeed an investigation of an outbreak. Studies are summarized in Table 5. These were mostly in lower TB burden countries. They were all descriptive and focused on different aspects of an outbreak, for example, contact investigation and follow-up, preventive measures, and transmission chains.
### Table 5. Overview of outbreak investigations.

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Disease</th>
<th>Population</th>
<th>Cases that triggered a response</th>
<th>Focus of the paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valway et al [58]</td>
<td>United States (New York)</td>
<td>MDR-TB²</td>
<td>Inmates from a prison in Upstate New York</td>
<td>4 inmates from one prison were diagnosed in the summer of 1991</td>
<td>Transmission patterns and contact investigation results</td>
</tr>
<tr>
<td>Rizdon et al [59]</td>
<td>United States (California)</td>
<td>DR-TB²</td>
<td>California high-school students</td>
<td>4 students were diagnosed in spring 1993</td>
<td>Findings from an outbreak investigation</td>
</tr>
<tr>
<td>Breathnach et al [60]</td>
<td>United Kingdom (London)</td>
<td>MDR-TB</td>
<td>Patients who were HIV positive at St Thomas’ Hospital in London</td>
<td>8 patients were identified between 1995 and 1997</td>
<td>Epidemiology and control of the hospital outbreak</td>
</tr>
<tr>
<td>Holdsworth et al [61]</td>
<td>United Kingdom (London)</td>
<td>MDR-TB</td>
<td>Nosocomial outbreak at Guy’s and St Thomas’ NHS Trust</td>
<td>8 patients were identified in the summer of 1996</td>
<td>Management of public relations following the outbreak</td>
</tr>
<tr>
<td>Moro et al [62]</td>
<td>Italy</td>
<td>MDR-TB</td>
<td>Patients infected by HIV and hospitalized in an HIV ward in Milan, Italy</td>
<td>33 patients were diagnosed between October 1992 and February 1994</td>
<td>Risk factors for transmission and the effectiveness of infection control measures</td>
</tr>
<tr>
<td>Schmid et al [63]</td>
<td>Austria</td>
<td>MDR-TB</td>
<td>Refugees in Austria</td>
<td>In 2005-2006 the Austrian laboratory for TB² identified 4 MDR-TB cases with similar genotypes</td>
<td>The chain of transmission</td>
</tr>
<tr>
<td>Asghar et al [64]</td>
<td>United States (Florida)</td>
<td>M. tuberculosis resistant to isoniazid</td>
<td>HIV-positive, rock cocaine (crack) users who lived in low-income neighborhoods in Miami</td>
<td>18 cases with matching spoligotypes were identified between January 2004 and May 2005</td>
<td>Transmission patterns and recommendations for TB control in this population</td>
</tr>
<tr>
<td>Fred et al [65]</td>
<td>Federated States of Micronesia</td>
<td>MDR-TB</td>
<td>Cluster of patients with MDR-TB in Chuik State</td>
<td>A cluster of 5 patients were identified in May 2008</td>
<td>Contact tracing and control measures</td>
</tr>
<tr>
<td>Chee et al [66]</td>
<td>Singapore</td>
<td>MDR-TB</td>
<td>LAN² gaming centers</td>
<td>In 2012, 5 men who attended 2 LAN gaming centers were diagnosed with MDR-TB</td>
<td>Highlights gaming centers as potential hotspots for TB transmission and notes challenges when conducting contact-tracing investigations</td>
</tr>
<tr>
<td>Norheim et al [67]</td>
<td>Norway</td>
<td>Streptomycin resistant TB</td>
<td>Students attending training sessions at an educational institution in Oslo, Norway</td>
<td>3 students were identified within one week in April 2013</td>
<td>Transmission patterns linking data from contact tracing to data from WGS²</td>
</tr>
<tr>
<td>Ho et al [68]</td>
<td>Singapore</td>
<td>MDR-TB</td>
<td>Residents of an 11-storey apartment block</td>
<td>6 residents were identified between February 2012 and May 2016</td>
<td>The cluster investigation and results from mass screening</td>
</tr>
<tr>
<td>Popovici et al [69]</td>
<td>Romania</td>
<td>XDR-TB⁸</td>
<td>Foreign medical students at a Romanian university</td>
<td>A cluster of 3 patients was identified in October 2015</td>
<td>Results from contact investigation and the efforts to identify the source case</td>
</tr>
<tr>
<td>Zhang et al [70]</td>
<td>China</td>
<td>MDR-TB</td>
<td>School in Zhejiang Province</td>
<td>A student was diagnosed in May 2014</td>
<td>Results from classmate contact investigation</td>
</tr>
<tr>
<td>Li et al [71]</td>
<td>China</td>
<td>MDR-TB</td>
<td>Senior high school</td>
<td>A female student diagnosed in March 2020</td>
<td>Results from household, classmate, and faculty investigations</td>
</tr>
<tr>
<td>Kobayashi et al [72]</td>
<td>Japan</td>
<td>MDR-TB</td>
<td>Japanese language school in Tokyo</td>
<td>A student was diagnosed in September 2019</td>
<td>Results from analysis of outbreak cases</td>
</tr>
<tr>
<td>Wu et al [73]</td>
<td>China</td>
<td>RR-TB¹</td>
<td>Middle school in Jiangsu Province</td>
<td>Unclear. 12 patients were diagnosed with TB of whom 6 were RR-TB.</td>
<td>Describe characteristics and epidemiology of outbreak and suggestions for prevention and control of school TB</td>
</tr>
</tbody>
</table>
aMDR-TB: multidrug-resistant tuberculosis.
bDR-TB: drug-resistant tuberculosis.
cNHS: National Health Services.
dTB: tuberculosis.
eLAN: local area network.
fWGS: whole-genome sequencing.
gXDR-TB: extensively drug-resistant tuberculosis.
hRR-TB: rifampicin-resistant tuberculosis.

**Airline Contact Investigations**

Three studies described airline DR-TB contact investigation. An der Heiden et al [75] investigated passengers and crew members after exposure to an individual with XDR-TB. The response rate was 83%. No secondary TB cases were reported and 1 individual with TB infection, probably newly acquired, was identified. Kornylo-Duong et al [76] evaluated passenger contacts of individuals with MDR-TB. More than 65% were lost to follow-up. No secondary TB cases were reported. Eight contacts tested positive for latent tuberculosis infection (LTBI); however, it was unknown if these contacts were recent converters as they might have acquired LTBI from their countries of residence. Glasauer et al [77] analyzed international contact tracing notifications received by Germany from other countries, with a focus on air travel. The high variability in the completeness of contact tracing information made analyses a challenge.

**Epidemiological Analyses**

Nitta et al [78], Anderson et al [79], de Vries et al [80], and Suppli et al [81] described transmission patterns of DR-TB in Los Angeles County (1993-1998), the United Kingdom (2004-2007), and the Netherlands (2010-2019 and 2018-2019), respectively. Intervention strategies to find those with TB disease are mentioned, but not described in detail. Villa et al [82] described a cluster of 16 pre-XDR and XDR-TB cases in Italy between 2016 and 2020 as well as the role of whole-genome sequencing in TB surveillance.

**Public-Private Partnership Program**

Joloba et al [83] described a program to improve MDR-TB detection by improving access to rapid and reliable DST, redesigning the TB specimen transport network, and training health care workers in Uganda. This study enhanced the care-seeking pathway (Figure 1, green dashed pathway) to specifically improve MDR-TB diagnosis and no screening took place.

**E-registry Program**

Naker et al [84] described a qualitative study in Mongolia to assess the feasibility and acceptability of an e-registry tool to simplify the systematic screening of MDR-TB contacts. Of 42 contacts identified, screened, and evaluated, when these contacts were screened at home or invited for screening at a health facility; however, except for target group differences, for example, contacts of DR-TB source cases compared with DS-TB source cases, and DST when presumptive DR-TB is identified, we did not note any differences between DR-TB case-finding and DS-TB case-finding strategies. Information on factors that may influence the yield of TB disease, like the number of contacts identified, screened, and evaluated, when these contacts were evaluated (baseline or follow-up), and which screening and diagnostic tools used [85] were seldom reported in detail.

**Discussion**

**Key Findings**

This scoping review charts the existing literature on DR-TB case finding. More than 60% of identified studies described DR-TB contact investigations. Included studies were all descriptive and no trials were identified. There is a lack of primary studies for inclusion in systematic reviews assessing the effectiveness, cost-effectiveness, acceptability, and feasibility of different DR-TB case-finding strategies. Case-finding strategies were not always reported in enough detail to deduce the specific pathways in our systems-based logic model (Figure 1), for example, whether symptomatic contacts were invited to a TB diagnostic service or whether contacts were screened at home or invited for screening at a health facility; however, except for target group differences, for example, contacts of DR-TB source cases compared with DS-TB source cases, and DST when presumptive DR-TB is identified, we did not note any differences between DR-TB case-finding and DS-TB case-finding strategies. Information on factors that may influence the yield of TB disease, like the number of contacts identified, screened, and evaluated, when these contacts were evaluated (baseline or follow-up), and which screening and diagnostic tools used [85] were seldom reported in detail.

**Previous Work**

Although conclusions about the most effective DR-TB case-finding strategy cannot be drawn, several reviews looked at the possible effects of active TB case finding (screening pathways in Figure 1) compared with passive TB case finding (care-seeking pathways in Figure 1) in general. Two Cochrane reviews failed to identify any studies for inclusion. Fox et al [86] aimed to compare the diagnostic yield of TB disease between active case finding and passive case finding in TB contacts but did not identify any trials for inclusion in the review. Braganza Menezes et al [87] also could not identify trials for inclusion in a review aiming to assess the effectiveness of novel methods, for example, social network analysis, of contact tracing versus the current standard of care to identify individuals with TB infection or TB disease. A review by
Kranzer et al [88] included observational studies and concluded that screening compared with standard care increases the number of patients with TB disease found in the short term, but that it is unknown whether it impacts TB epidemiology. This finding was underpinned by another Cochrane review. Mhimbira et al [5] found that active case-finding strategies may result in increased case finding in the short term, but long-term outcomes were lacking. Except for the unknown effect of TB screening on TB epidemiology, the effect of screening on individual outcomes had been studied by Telisinghe et al [89] and Kranzer et al [88]. These reviews found limited patient outcome data and no difference in treatment outcomes between active and passive case findings. While it is known that screening may increase the number of identified cases, it is unclear if TB screening makes a difference to TB epidemiology and individual outcomes compared with passive case finding.

**Implications for Research**

There is a need for standardization of terminology, design, and reporting of DR-TB case-finding studies, especially contact investigation studies. Future research should focus on clear definitions, methodology, and detailed descriptions of all intervention components. There is also a need for well-conducted randomized controlled trials assessing the effect of active case finding on individual outcomes and long-term TB epidemiological outcomes.

**Strengths and Limitations**

The strengths of our review included a thorough search strategy and the use of a systems-based logic model. Nevertheless, there were some limitations. We searched several databases with no date or language restriction, but we did not translate all full-text articles. Studies that were not translated (n=4) are reported in the list of excluded studies (Multimedia Appendix 4) and from the translated abstracts it seems that similar study designs were found to those of included contact investigation studies. Our review excluded patients with TB symptoms seeking care, patients diagnosed with TB, and laboratory samples or laboratory isolates because the focus of this review was on active case finding. It should therefore be noted that studies that investigated strategies to improve identification of DR-TB after clinical identification of presumptive DR-TB cases, or studies that screened laboratory samples were not part of this review. Furthermore, for collating, summarizing, and reporting the results, we initially envisaged using our systems-based logic model as a framework to describe different case-finding strategies and resulting pathways. However, reporting was incomplete and inconsistent, and we were not able to describe pathways in detail. Nevertheless, the logic model guided our interpretation of whether a case-finding study involved screening or not. Finally, for contact investigation studies, we included studies that reported on the number of individuals with TB disease diagnosed, even if the study focused on LTBI testing and treatment. This might be a reason why the screening and diagnostic pathways were not always reported in detail. However, it is important to note that in contact investigation studies, the active case-finding component and the LTBI treatment component are both important aspects of early case finding and prevention.

**Conclusions**

Existing descriptive reviews can be updated, but there is a dearth of knowledge on the effectiveness, cost-effectiveness, acceptability, and feasibility of DR-TB case-finding strategies to inform policy and practice. There is also a need for standardization of terminology, design, and reporting of DR-TB case-finding studies, especially contact investigation studies, to decrease a large amount of research waste and increase the number of studies that could be synthesized and meta-analyzed in high-impact systematic reviews in the future [90].

**Acknowledgments**

We would like to thank Anel Schoonees and Vittoria Lutje, our information specialists, for their assistance with the search strategy. SSwV is supported by the Research, Evidence and Development Initiative (READ-It). READ-It (project number 300342-104) is funded by UK aid from the UK government; however, the views expressed do not necessarily reflect the UK government’s official policies. MC is supported jointly by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO) under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 program supported by the European Union. JAS is supported by a Clinician Scientist Fellowship jointly funded by the UK MRC and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement (MR/R007942/1). GH received financial assistance from the European Union (grant DCI-PANAF/2020/420-028), through the African Research Initiative for Scientific Excellence (ARISE), pilot program. ARISE is implemented by the African Academy of Sciences with support from the European Commission and the African Union Commission.

**Data Availability**

All data analyzed during this study are included in this published article and its Multimedia Appendices.

**Authors’ Contributions**

MC and SSwV conceived the study idea. SSwV, MN, LV, FWL, CCL, and MC performed title, abstract, and full-text screening. All authors assisted with data extraction and analyses. SSwV drafted the paper. All authors read and approved the final paper.
Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.

Multimedia Appendix 2
Detailed search strategy.

Multimedia Appendix 3
Characteristics of drug-resistant TB contact investigation studies.

Multimedia Appendix 4
List of excluded studies with reasons.

References


Telehealth Utilization and Associations in the United States During the Third Year of the COVID-19 Pandemic: Population-Based Survey Study in 2022

Jiyeong Kim¹, MPH, PhD; Zhuo Ran Cai¹, MD; Michael L Chen¹, BA; Sonia Onyeka¹, MD; Justin M Ko¹, MD, MBA; Eleni Linos¹, MD, MPH, DrPH
Stanford Center for Digital Health, School of Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Jiyeong Kim, MPH, PhD
Stanford Center for Digital Health
School of Medicine
Stanford University
3180 Porter Dr
Stanford, CA, 94305
United States
Phone: 1 6507245017
Email: jykim3@stanford.edu

Abstract

Background: The COVID-19 pandemic rapidly changed the landscape of clinical practice in the United States; telehealth became an essential mode of health care delivery, yet many components of telehealth use remain unknown years after the disease’s emergence.

Objective: We aim to comprehensively assess telehealth use and its associated factors in the United States.

Methods: This cross-sectional study used a nationally representative survey (Health Information National Trends Survey) administered to US adults (≥18 years) from March 2022 through November 2022. To assess telehealth adoption, perceptions of telehealth, satisfaction with telehealth, and the telehealth care purpose, we conducted weighted descriptive analyses. To identify the subpopulations with low adoption of telehealth, we developed a weighted multivariable logistic regression model.

Results: Among a total of 6252 survey participants, 39.3% (2517/6252) reported telehealth use in the past 12 months (video: 1110/6252, 17.8%; audio: 876/6252, 11.6%). The most prominent reason for not using telehealth was due to telehealth providers failing to offer this option (2200/3529, 63%). The most common reason for respondents not using offered telehealth services was a preference for in-person care (527/578, 84.4%). Primary motivations to use telehealth were providers’ recommendations (1716/2517, 72.7%) and convenience (1516/2517, 65.6%), mainly for acute minor illness (600/2397, 29.7%) and chronic condition management (583/2397, 21.4%), yet care purposes differed by age, race/ethnicity, and income. The satisfaction rate was predominately high, with no technical problems (1829/2517, 80.5%), comparable care quality to that of in-person care (1779/2517, 70.5%), and no privacy concerns (1958/2517, 83.7%). Younger individuals (odd ratios [ORs] 1.48-2.23; 18-64 years vs ≥75 years), women (OR 1.33, 95% CI 1.09-1.61), Hispanic individuals (OR 1.37, 95% CI 1.05-1.80; vs non-Hispanic White), those with more education (OR 1.72, 95% CI 1.03-2.87; at least a college graduate vs less than high school), unemployed individuals (OR 1.25, 95% CI 1.02-1.54), insured individuals (OR 1.83, 95% CI 1.25-2.69), or those with poor general health status (OR 1.66, 95% CI 1.30-2.13) had higher odds of using telehealth.

Conclusions: To our best knowledge, this is among the first studies to examine patient factors around telehealth use, including motivations to use, perceptions of, satisfaction with, and care purpose of telehealth, as well as sociodemographic factors associated with telehealth adoption using a nationally representative survey. The wide array of descriptive findings and identified associations will help providers and health systems understand the factors that drive patients toward or away from telehealth visits as the technology becomes more routinely available across the United States, providing future directions for telehealth use and telehealth research.

(JMIR Public Health Surveill 2024;10:e51279) doi:10.2196/51279
Introduction

Telehealth refers to health care delivered through digital devices (eg, computers, tablets, telephones, or smartphones) and typically includes communicating with the health care providers via live chat over the audio or video format or asynchronous messages through email or a patient portal [1]. During the early COVID-19 pandemic, with the unprecedented “Stay at Home Order” in March 2020 in the United States, in-person office visits were extremely restrained, while the demand of health care significantly increased due to the widespread infectious disease outbreak. Telehealth quickly served as an essential alternative to the limited in-person care [2]. Many private and public health insurance plans promptly started to cover virtual visits to respond to these emergency situations in an effort to enhance the availability of telehealth access [3,4]. The majority of health care sectors, including oncology, psychology, and surgery, promptly provided telehealth services to address the care needs while avoiding unnecessary in-person exposure to the infectious virus [5-7].

Previous studies analyzing sociodemographic characteristics have shown lower telehealth adoption in some populations in the United States. A study using one primary care network reported that older adults, non-Hispanic Whites, and low-income individuals were less likely to utilize telehealth visits [8]. Women and Medicaid beneficiaries had low odds of using telehealth among cancer survivors [9]. Moreover, low English proficiency was also related to low telehealth adoption among New York residents [10]. However, previous assessments were conducted in limited population groups (eg, older African American individuals) [11], at a single medical center or academic institution, with a specific medical specialty area (eg, oncology surgery), or in one geographic area (eg, New York, Los Angeles) and solely examined trends of telehealth utilization [12] or the impact of parity payment laws on telehealth use [13].

However, many components of telehealth use, such as telehealth availability, motivations for using telehealth, and patient perceptions of telehealth visits remain unknown. Second, an analysis of sociodemographic factors associated with low adoption of telehealth among individuals in the United States has not been conducted at the population level. Third, prior studies were conducted in the early phases of the pandemic, which may not be representative of late pandemic and current telehealth practices. Thus, this study aimed to comprehensively assess telehealth use in the United States, including adoption, motivations to and purposes of use, satisfaction, and perceptions, and identify sociodemographic associations with low adoption of telehealth during the third year of the COVID-19 pandemic using a nationwide database. Our findings will advance knowledge of recent telehealth use in the United States and contribute to preparing targeted approaches to enhance telehealth among those with low adoption of it. Recent evidence suggests that telehealth could contribute to enhancing health care access in some marginalized subgroups [14]. Hence, the knowledge and effort will be timely because telehealth has now been grounded as an essential part of health care delivery, being promoted from an emergency alternative during the early COVID-19 pandemic [15].

Methods

Data Source

We used a nationally representative survey (Health Information National Trends Survey [HINTS] 6, 2022) for the study [16]. HINTS is a publicly available source of self-reported cross-sectional data. The survey was administered to noninstitutionalized civilians (≥18 years) in the United States who were selected by a random sampling of stratified addresses. HINTS 6 was offered as a paper or online survey and collected from March 2022 through November 2022. With a total of 6252 respondents, the response rate was 28.1% [17]. We applied the full-sample weights to account for the household-level base weight, nonresponse, and the person-level initial weight [17]. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [18].

Outcome

To evaluate telehealth adoption, the following question was used: “A telehealth is a telephone or video appointment with a doctor or health professional. In the past 12 months, did you receive care from a doctor or health professional using telehealth?” The response options were “yes” (received by video, audio, or both) or “no” (not received; see Multimedia Appendix 1).

To assess motivation, those who answered “yes” to the telehealth adoption question were asked to indicate reasons why they chose telehealth, with answer options including (1) the health care provider recommended or required it, (2) I wanted advice about whether I needed in-person medical care, (3) I wanted to avoid possible infection at the office, (4) it was more convenient than going to the doctor, and (5) I could include family or caregivers in my appointment. Answer choices included “yes” or “no.” To examine satisfaction with their telehealth visits, participants used a Likert scale (strongly agree, somewhat agree, somewhat disagree, strongly disagree) to rate the following items: (1) I had technical problems with my telehealth visits, (2) the telehealth care was as good as in-person care, (3) I was concerned about the privacy of my telehealth visits, (2) the telehealth care was as good as in-person care, (3) I was concerned about the privacy of my telehealth visits. To assess the purpose of telehealth use, participants also used a Likert scale (strongly agree, somewhat agree, somewhat disagree, strongly disagree) to rate the following items: (1) annual visit, (2) minor illness/acute care, (3) managing chronic condition/disease, (4) medical emergency, (5) mental health, (6) other. For this study, the responses to the latter 2 questions were then coded as a binary variable: agree (strongly agree, somewhat agree) or disagree (somewhat disagree, strongly disagree).
To evaluate their perceptions, those who answered “no” to the telehealth adoption question were asked to indicate the reasons why they did not participate in telehealth visits, answering “yes” or “no” to options including (1) a preference for in-person care, (2) privacy concerns, or (3) difficulty with use.

Covariates
We selected sociodemographic characteristics based on the social determinants of health conceptual framework from Healthy People 2030 [19], which includes age, birth gender, race/ethnicity, household income (<US $20,000, US $20,000 to <$35,000, US $35,000 to <$50,000, US $50,000 to <$75,000, ≥ US $75,000), education, marital status, employment status, health insurance, and rurality of residence (metropolitan, micropolitan, small town, rural) [20,21]. Additionally, health status factors included general health (excellent, very good, good, fair, poor) and chronic medical conditions (cancer, diabetes, high blood pressure, heart disease, lung disease).

Statistical Analysis
We performed weighted descriptive analyses to calculate the frequency (n) and weighted percentage (%) with the SE to illustrate the sociodemographic and health status characteristics of the study population. Weighted descriptive analyses were also conducted to present telehealth adoption (computed by prevalence), by mode, sociodemographic characteristics, and health status characteristics to identify subgroups with higher-than-average adoption, as well as motivations for use, the care purpose, satisfaction with recent telehealth visits, and perception of telehealth prior to its use. Differences by group were assessed using the Wald chi-square test. To further explore differences in purpose and satisfaction with recent telehealth visits by age, birth gender, race/ethnicity, education, and employment status, weighted descriptive analyses with the Wald chi-square test were also performed. To examine the factors associated with telehealth use, we developed a multivariable weighted logistic regression model to obtain odds ratios (ORs) and 95% CIs for the sociodemographic and health status characteristics related to telehealth use. The logistic regression model was adjusted for age, birth gender, race/ethnicity, education, income, marital status, employment status, health insurance, number of health care office visits a year, and general health status, which were selected because these were potential confounders in this study (eg, changed covariate estimates by more than 10%) or previously known confounders for telehealth use [8,9,11].

The range of missingness was 1.7% to 11.4%, and covariates with any missing values were imputed. We applied the Hot deck imputation method, which was used to account for nonresponse by HINTS [16]. For all the descriptive and regression analyses, the imputed data were used, and the statistical significance was determined at P<.05 in SAS 9.4 (SAS Studio) [22,23].

Ethical Considerations
This was a secondary analysis of publicly available national survey data (HINTS). This study did not involve human subjects or identifiable information. Given that the data were deidentified, this study was deemed exempt from review by the Institutional Review Board of Stanford University.

Results
Study Population Characteristics
Table 1 presents the sociodemographic and health status characteristics of the study population in the third year of the COVID-19 pandemic in the United States. Of the survey respondents, 78.6% (4045/6252) of the respondents were less than 65 years old, 50.5% (3733/6252) were women, 61% (3615/6252) were non-Hispanic White, 72% (4703/6252) had some or more than a college education, 62.4% (3558/6252) had an income of at least US $50,000, slightly more than one-half were employed (2980/6252, 54.1%), slightly more than one-half were married (3234/6252, 56.1%), 85.7% (5393/6252) resided in a metropolitan area, 89.2% (5709/6252) had health insurance, 83.2% (5134/6252) were in generally excellent or good health, and 37.1% (2798/6252) reported high blood pressure.
Table 1. Sociodemographic and health status characteristics of 6252 US adults in the third year of the COVID-19 pandemic (Health Information National Trends Survey [HINTS] 6, 2022).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Respondents, n(^a)</th>
<th>Respondents, weighted % (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>979</td>
<td>26.2 (0.8)</td>
</tr>
<tr>
<td>35-49</td>
<td>1262</td>
<td>25.1 (0.9)</td>
</tr>
<tr>
<td>50-64</td>
<td>1804</td>
<td>27.3 (0.6)</td>
</tr>
<tr>
<td>65-74</td>
<td>1362</td>
<td>12.9 (0.1)</td>
</tr>
<tr>
<td>≥75</td>
<td>848</td>
<td>8.5 (0.04)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3733</td>
<td>50.5 (0.4)</td>
</tr>
<tr>
<td>Male</td>
<td>2519</td>
<td>49.5 (0.4)</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>3615</td>
<td>61.0 (0.4)</td>
</tr>
<tr>
<td>Non-Hispanic Black/African American</td>
<td>955</td>
<td>10.6 (0.2)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1124</td>
<td>17.3 (0.3)</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>343</td>
<td>6.0 (0.2)</td>
</tr>
<tr>
<td>Others</td>
<td>215</td>
<td>5.1 (0.2)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>406</td>
<td>6.5 (0.6)</td>
</tr>
<tr>
<td>High school graduate</td>
<td>1143</td>
<td>21.5 (0.8)</td>
</tr>
<tr>
<td>Some college</td>
<td>1851</td>
<td>39.7 (0.8)</td>
</tr>
<tr>
<td>At least a college education</td>
<td>2852</td>
<td>32.3 (0.3)</td>
</tr>
<tr>
<td><strong>Household income (US $)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20,000</td>
<td>1064</td>
<td>14.7 (0.9)</td>
</tr>
<tr>
<td>20,000 to &lt;35,000</td>
<td>814</td>
<td>11.5 (0.7)</td>
</tr>
<tr>
<td>35,000 to &lt;50,000</td>
<td>816</td>
<td>11.4 (0.7)</td>
</tr>
<tr>
<td>50,000 to &lt;75,000</td>
<td>1062</td>
<td>18.1 (0.9)</td>
</tr>
<tr>
<td>≥75,000</td>
<td>2496</td>
<td>44.3 (1.0)</td>
</tr>
<tr>
<td><strong>Employment</strong>(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>2980</td>
<td>54.1 (1.3)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>3272</td>
<td>45.9 (1.3)</td>
</tr>
<tr>
<td><strong>Marital status</strong>(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>3234</td>
<td>56.1 (0.5)</td>
</tr>
<tr>
<td>Not married</td>
<td>3018</td>
<td>43.9 (0.5)</td>
</tr>
<tr>
<td><strong>Rurality</strong>(^d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metropolitan</td>
<td>5393</td>
<td>85.7 (0.7)</td>
</tr>
<tr>
<td>Micropolitan</td>
<td>489</td>
<td>7.8 (0.5)</td>
</tr>
<tr>
<td>Small town</td>
<td>241</td>
<td>4.3 (0.6)</td>
</tr>
<tr>
<td>Rural</td>
<td>129</td>
<td>2.2 (0.3)</td>
</tr>
<tr>
<td><strong>Health insurance</strong>(^e)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5709</td>
<td>89.2 (0.2)</td>
</tr>
<tr>
<td>No</td>
<td>543</td>
<td>10.8 (0.2)</td>
</tr>
</tbody>
</table>
### Characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Respondents, n&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Respondents, weighted % (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General health status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent/good</td>
<td>5134</td>
<td>83.2 (1.0)</td>
</tr>
<tr>
<td>Fair/poor</td>
<td>1118</td>
<td>16.8 (1.0)</td>
</tr>
<tr>
<td><strong>Chronic medical condition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>1342</td>
<td>17.1 (0.7)</td>
</tr>
<tr>
<td>High blood pressure</td>
<td>2798</td>
<td>37.1 (0.9)</td>
</tr>
<tr>
<td>Heart disease</td>
<td>617</td>
<td>7.6 (0.5)</td>
</tr>
<tr>
<td>Lung disease</td>
<td>850</td>
<td>11.7 (0.5)</td>
</tr>
<tr>
<td>Cancer</td>
<td>937</td>
<td>10.2 (0.2)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Covariates with any missing values were imputed; missingness of covariates from the sample of 6252 adults: age: n=98, 1.6%; birth gender: n=410, 6.6%; race/ethnicity: n=687, 11%; education: n=404, 6.5%; income: n=732, 11.7%; employment status: n=390, 6.2%; marital status: n=415, 6.6%; health insurance: n=126, 2%; general health status: n=234, 3.7%; diabetes: n=252, 4%; high blood pressure: n=244, 3.9%; heart disease: n=238, 3.8%; lung disease: n=234, 3.7%; cancer: n=370, 5.9%.

**Telehealth Adoption by Mode by Sociodemographic Characteristics**

Overall, 39.3% (2517/6252) of the respondents had adopted telehealth (Table 2). Video-only visits (17.8%) were more prevalent, followed by audio-only visits (11.6%). Table 2 illustrates telehealth adoption by mode and the differences by sociodemographic and health status characteristics. Video visit adoption was higher among those aged 35 years to 49 years (<i>P</i>&lt;.001), women (<i>P</i>&lt;.001), the employed (<i>P</i>&lt;.001), the insured (<i>P</i>&lt;.001), those with at least a college education (<i>P</i>&lt;.001), and individuals with a high income (≥US $75,000; <i>P</i>&lt;.001). Audio visit adoption was higher among the oldest age group (≥75 years; <i>P</i>&lt;.001), Hispanic individuals (<i>P</i>=.02), those with a low income (<US $35,000; <i>P</i>&lt;.001), and unemployed respondents (<i>P</i>&lt;.001).
Table 2. Telehealth adoption by sociodemographic and health status characteristics of 6252 US adults in the third year of the COVID-19 pandemic (Health Information National Trends Survey [HINTS] 6, 2022).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Anya</th>
<th>Participants, weighted % (SE)</th>
<th>Video only, weighted % (SE)</th>
<th>Audio only, weighted % (SE)</th>
<th>Both video and audio, weighted % (SE)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All adopters (n=6046)</td>
<td>2517</td>
<td>39.3 (1.0)</td>
<td>17.8 (0.8)</td>
<td>11.6 (0.6)</td>
<td>8.4 (0.5)</td>
<td>—</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34 (n=916)</td>
<td>385</td>
<td>35.7 (2.6)</td>
<td>15.8 (2.3)</td>
<td>8.6 (1.5)</td>
<td>9.5 (1.5)c</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>35-49 (n=1226)</td>
<td>566c</td>
<td>45.1 (2.1)c</td>
<td>22.7 (1.7)c</td>
<td>12.2 (1.3)c</td>
<td>8.4 (1.2)</td>
<td>.009</td>
</tr>
<tr>
<td>50-64 (n=1755)</td>
<td>728</td>
<td>39.0 (2.1)</td>
<td>18.9 (1.4)c</td>
<td>10.7 (0.9)</td>
<td>8.6 (1.2)c</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>65-74 (n=1324)</td>
<td>525</td>
<td>37.7 (2.3)</td>
<td>15.5 (1.5)</td>
<td>14.7 (1.4)c</td>
<td>6.5 (0.8)</td>
<td>.001</td>
</tr>
<tr>
<td>≥75 (n=825)</td>
<td>307</td>
<td>36.2 (2.5)</td>
<td>9.6 (1.6)</td>
<td>17.9 (2.2)c</td>
<td>7.4 (1.5)</td>
<td>.02</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (n=3607)</td>
<td>1613c</td>
<td>44.1 (1.5)c</td>
<td>19.5 (1.0)c</td>
<td>13.2 (0.9)c</td>
<td>9.9 (0.8)c</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male (n=2439)</td>
<td>904</td>
<td>34.4 (1.5)</td>
<td>16.2 (1.2)</td>
<td>10.0 (0.9)</td>
<td>6.9 (1.0)</td>
<td>.37</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.09</td>
</tr>
<tr>
<td>Non-Hispanic White (n=3506)</td>
<td>1440</td>
<td>39.1 (1.4)</td>
<td>18.5 (1.1)c</td>
<td>11.2 (1.0)</td>
<td>8.2 (0.8)</td>
<td>.7</td>
</tr>
<tr>
<td>Non-Hispanic Black/African American (n=933)</td>
<td>372</td>
<td>35.4 (3.1)</td>
<td>15.0 (2.3)</td>
<td>10.6 (1.2)</td>
<td>8.0 (1.9)</td>
<td>.37</td>
</tr>
<tr>
<td>Hispanic (n=1073)</td>
<td>479c</td>
<td>39.8 (2.1)c</td>
<td>15.4 (1.7)</td>
<td>13.9 (1.2)c</td>
<td>8.9 (1.4)c</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Non-Hispanic Asian (n=327)</td>
<td>127</td>
<td>38.5 (5.6)</td>
<td>18.9 (4.1)c</td>
<td>11.2 (3.0)</td>
<td>8.0 (2.1)</td>
<td>.02</td>
</tr>
<tr>
<td>Others (n=207)</td>
<td>99c</td>
<td>49.5 (8.2)c</td>
<td>23.1 (9.2)c</td>
<td>11.8 (3.6)c</td>
<td>10.3 (3.2)c</td>
<td>.02</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Less than high school (n=389)</td>
<td>125</td>
<td>29.8 (4.9)</td>
<td>13.5 (4.3)</td>
<td>11.2 (1.6)</td>
<td>3.2 (1.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High school graduate (n=1097)</td>
<td>376</td>
<td>33.7 (1.9)</td>
<td>13.4 (1.7)</td>
<td>12.3 (1.6)c</td>
<td>6.9 (1.8)</td>
<td>.2</td>
</tr>
<tr>
<td>Some college (n=1758)</td>
<td>710</td>
<td>39.2 (1.7)</td>
<td>17.5 (1.5)</td>
<td>11.3 (1.0)</td>
<td>8.7 (1.1)c</td>
<td>.001</td>
</tr>
<tr>
<td>At least a college graduate (n=2802)</td>
<td>1306c</td>
<td>45.0 (1.6)c</td>
<td>22.1 (1.2)c</td>
<td>11.7 (0.8)c</td>
<td>10.1 (0.9)c</td>
<td>.01</td>
</tr>
<tr>
<td>Household income (US $)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>&lt;20,000 (n=1012)</td>
<td>402</td>
<td>36.1 (2.8)</td>
<td>13.6 (1.9)</td>
<td>12.9 (1.3)c</td>
<td>7.8 (1.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>20,000 to &lt;35,000 (n=775)</td>
<td>281</td>
<td>36.2 (2.7)</td>
<td>12.7 (2.7)</td>
<td>14.8 (2.0)c</td>
<td>6.8 (1.5)</td>
<td>.09</td>
</tr>
<tr>
<td>35,000 to &lt;50,000 (n=772)</td>
<td>311</td>
<td>37.9 (3.5)</td>
<td>15.9 (2.2)</td>
<td>10.9 (1.3)</td>
<td>10.0 (2.7)c</td>
<td>.02</td>
</tr>
<tr>
<td>50,000 to &lt;75,000 (n=1044)</td>
<td>412</td>
<td>34.6 (2.4)</td>
<td>13.6 (1.8)</td>
<td>11.1 (1.4)</td>
<td>8.2 (1.5)</td>
<td>.001</td>
</tr>
<tr>
<td>≥75,000 (n=2443)</td>
<td>1111c</td>
<td>43.2 (1.9)c</td>
<td>22.6 (1.5)c</td>
<td>10.9 (1.2)</td>
<td>8.7 (0.8)c</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Married (n=3126)</td>
<td>1350d</td>
<td>41.7 (1.2)c</td>
<td>19.6 (1.1)c</td>
<td>12.0 (0.8)</td>
<td>8.7 (0.8)c</td>
<td>.06</td>
</tr>
<tr>
<td>Unmarried (n=2920)</td>
<td>1167</td>
<td>36.2 (1.8)</td>
<td>15.5 (1.5)</td>
<td>11.2 (0.8)</td>
<td>8.1 (0.9)</td>
<td>.001</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08</td>
</tr>
<tr>
<td>Employed (n=2901)</td>
<td>1189</td>
<td>37.5 (1.5)</td>
<td>18.2 (1.2)c</td>
<td>9.3 (0.7)</td>
<td>8.8 (0.8)c</td>
<td>.09</td>
</tr>
<tr>
<td>Unemployed (n=3145)</td>
<td>1328c</td>
<td>41.4 (1.4)c</td>
<td>17.4 (1.1)</td>
<td>14.5 (1.0)c</td>
<td>8.0 (0.7)</td>
<td>.02</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Any(^a)</td>
<td>Video only, weighted % (SE)</td>
<td>Audio only, weighted % (SE)</td>
<td>Both video and audio, weighted % (SE)</td>
<td>P value</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>--------------------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td><strong>Participants, weighted %</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rurality of residence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>Metropolitan (n=5223)</td>
<td>2254(^c)</td>
<td>39.9 (1.0)(^c)</td>
<td>17.6 (0.8)</td>
<td>12.2 (0.7)(^c)</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>Micropolitan (n=470)</td>
<td>152</td>
<td>34.4 (4.6)</td>
<td>16.9 (4.1)</td>
<td>9.4 (1.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small town (n=230)</td>
<td>80(^c)</td>
<td>40.6 (9.0)(^c)</td>
<td>26.7 (10.2)(^c)</td>
<td>6.2 (2.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural (n=123)</td>
<td>31</td>
<td>31.4 (9.3)</td>
<td>12.8 (3.7)</td>
<td>7.2 (2.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Health insurance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Yes (n=5542)</td>
<td>2374(^c)</td>
<td>41.5 (1.1)(^c)</td>
<td>19.0 (0.8)(^c)</td>
<td>11.8 (0.7)(^c)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>No (n=504)</td>
<td>143</td>
<td>21.2 (3.1)</td>
<td>7.8 (1.4)</td>
<td>10.3 (2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of office visits per year</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>None (n=683)</td>
<td>109</td>
<td>11.9 (1.8)</td>
<td>4.3 (1.2)</td>
<td>5.9 (1.3)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>1-4 (n=3832)</td>
<td>1530</td>
<td>38.7 (1.4)</td>
<td>18.0 (1.1)(^c)</td>
<td>12.0 (0.8)(^c)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>≥5 (n=1531)</td>
<td>878(^d)</td>
<td>58.3 (1.9)(^c)</td>
<td>25.8 (2.1)(^c)</td>
<td>14.3 (1.3)(^c)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td><strong>General health status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Excellent/good (n=4974)</td>
<td>2000</td>
<td>37.3 (1.1)</td>
<td>17.6 (0.9)</td>
<td>10.9 (0.7)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Fair/poor (n=1072)</td>
<td>517(^c)</td>
<td>49.2 (2.4)(^c)</td>
<td>18.8 (3.0)(^c)</td>
<td>15.4 (1.4)(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chronic medical condition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes (n=1296)</td>
<td>620(^c)</td>
<td>46.0 (2.5)(^c)</td>
<td>17.7 (2.2)</td>
<td>15.6 (1.4)(^c)</td>
<td>.005</td>
<td></td>
</tr>
<tr>
<td>High blood pressure (n=2710)</td>
<td>1203(^c)</td>
<td>43.3 (1.7)(^c)</td>
<td>16.7 (1.5)</td>
<td>14.2 (0.9)(^c)</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Heart disease (n=609)</td>
<td>306(^c)</td>
<td>49.5 (3.3)(^c)</td>
<td>18.2 (2.5)(^c)</td>
<td>19.3 (2.4)(^c)</td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>Lung disease (n=816)</td>
<td>432(^c)</td>
<td>50.4 (2.5)(^c)</td>
<td>17.6 (1.3)</td>
<td>16.5 (1.8)(^c)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Depression (n=1587)</td>
<td>913(^c)</td>
<td>56.7 (2.1)(^c)</td>
<td>24.6 (1.8)(^c)</td>
<td>15.4 (1.5)(^c)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>History of cancer (n=911)</td>
<td>407(^c)</td>
<td>47.8 (2.8)(^c)</td>
<td>20.5 (2.3)(^c)</td>
<td>13.6 (1.9)(^c)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)"Any" included video only (n=1110), audio only (n=876), and both (n=531).

\(^b\)Not applicable.

\(^c\)Telehealth adoption use was higher than the average.

**Telehealth Motivations, Satisfaction, Perceptions, and Purposes**

The primary reason for not using telehealth was the lack of an available telehealth option (2200/3529, 63%) or preference for in-person care if telehealth was offered (527/578, 84.4%; Figure 1). Provider's recommendation (1716/2517, 72.7%) and convenience (1516/2517, 65.6%) motivated people to use telehealth, and users were mostly satisfied with their telehealth visits, with no reported technical problems (1829/2517, 80.5%) or privacy concerns (1958/2517, 83.7%) and good care quality (1779/2517, 75%). Acute (600/2397, 29.7%) and chronic (583/2397, 21.4%) condition care were the most common purposes of telehealth use. However, the purpose of telehealth use differed by age (of those aged 18-34 years, 28.2% [99/381] used telehealth for mental health vs of those aged ≥75 years, 31.3% [93/292] used telehealth for chronic condition care), race/ethnicity (of non-Hispanic Asian individuals, 35.6% [41/124] used telehealth for acute care vs of individuals of other races/ethnicities, 29% [27/94] used telehealth for chronic condition care), and income (of those with an income <US $20,000, 27.5% [92/345] used telehealth for chronic condition care vs of those with an income ≥US $75,000, 34.9% [335/1085] used telehealth for acute care; Table 3 and Table S1 in Multimedia Appendix 2).
Figure 1. Study population flow chart (weighted %).

![Study population flow chart](image)

**Table 3.** Purpose for telehealth use by US adults in the third year of the COVID-19 pandemic (Health Information National Trends Survey [HINTS] 6, 2022).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Annual visit (n=485), weighted % (SE)</th>
<th>Acute minor illness (n=600), weighted % (SE)</th>
<th>Chronic medical condition management (n=585), weighted % (SE)</th>
<th>Medical emergency (n=46), weighted % (SE)</th>
<th>Mental health, behavioral, or substance abuse issues (n=364), weighted % (SE)</th>
<th>Other (n=346), weighted % (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All adopters (n=2397)</td>
<td>17.8 (1.6)</td>
<td>29.7 (1.9)</td>
<td>21.4 (1.4)</td>
<td>1.6 (0.3)</td>
<td>15.7 (1.3)</td>
<td>13.8 (1.0)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>19.5 (5.7)</td>
<td>28.0 (4.3)</td>
<td>9.3 (2.3)</td>
<td>2.3 (1.1)</td>
<td>28.2 (4.6)</td>
<td>12.7 (2.5)</td>
</tr>
<tr>
<td>35-49</td>
<td>11.3 (1.8)</td>
<td>36.5 (3.5)</td>
<td>19.6 (3.2)</td>
<td>2.0 (0.7)</td>
<td>17.1 (2.2)</td>
<td>13.5 (2.1)</td>
</tr>
<tr>
<td>50-64</td>
<td>17.0 (2.2)</td>
<td>30.0 (3.1)</td>
<td>27.9 (2.8)</td>
<td>0.4 (0.3)</td>
<td>11.2 (1.9)</td>
<td>13.5 (2.1)</td>
</tr>
<tr>
<td>65-74</td>
<td>26.2 (2.8)</td>
<td>22.6 (3.1)</td>
<td>29.4 (3.4)</td>
<td>1.2 (0.4)</td>
<td>6.4 (1.6)</td>
<td>14.2 (3.0)</td>
</tr>
<tr>
<td>≥75</td>
<td>28.6 (3.9)</td>
<td>17.4 (3.6)</td>
<td>31.3 (4.9)</td>
<td>2.6 (1.6)</td>
<td>0.8 (0.5)</td>
<td>19.3 (4.9)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>16.1 (1.6)</td>
<td>32.5 (2.3)</td>
<td>19.6 (1.6)</td>
<td>1.1 (0.3)</td>
<td>17.6 (1.8)</td>
<td>13.1 (1.5)</td>
</tr>
<tr>
<td>Non-Hispanic Black/African American</td>
<td>27.4 (4.2)</td>
<td>23.9 (6.6)</td>
<td>25.9 (5.5)</td>
<td>2.0 (0.9)</td>
<td>6.8 (1.6)</td>
<td>14.0 (3.9)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>13.9 (1.8)</td>
<td>28.0 (3.4)</td>
<td>22.3 (3.2)</td>
<td>3.5 (1.3)</td>
<td>13.1 (2.4)</td>
<td>19.2 (2.7)</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>18.8 (4.6)</td>
<td>35.6 (6.3)</td>
<td>22.4 (6.5)</td>
<td>1.6 (1.3)</td>
<td>14.8 (6.2)</td>
<td>6.8 (2.6)</td>
</tr>
<tr>
<td>Others</td>
<td>29.6 (19.4)</td>
<td>9.8 (4.4)</td>
<td>28.8 (11.1)</td>
<td>0.8 (0.7)</td>
<td>19.0 (7.6)</td>
<td>12.0 (5.2)</td>
</tr>
<tr>
<td>Household income (US $)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20,000</td>
<td>22.0 (3.6)</td>
<td>17.6 (4.8)</td>
<td>27.5 (5.7)</td>
<td>3.0 (1.2)</td>
<td>14.6 (3.1)</td>
<td>15.3 (3.3)</td>
</tr>
<tr>
<td>20,000 to &lt;35,000</td>
<td>19.9 (3.4)</td>
<td>19.8 (6.8)</td>
<td>29.6 (5.0)</td>
<td>2.2 (1.0)</td>
<td>14.8 (3.2)</td>
<td>13.6 (4.5)</td>
</tr>
<tr>
<td>35,000 to &lt;50,000</td>
<td>15.4 (3.2)</td>
<td>31.5 (5.4)</td>
<td>20.2 (3.2)</td>
<td>2.4 (1.9)</td>
<td>20.8 (7.8)</td>
<td>9.7 (1.8)</td>
</tr>
<tr>
<td>50,000 to &lt;75,000</td>
<td>15.2 (3.1)</td>
<td>27.1 (4.1)</td>
<td>25.1 (4.2)</td>
<td>0.6 (0.5)</td>
<td>16.8 (3.3)</td>
<td>15.3 (2.5)</td>
</tr>
<tr>
<td>≥75,000</td>
<td>17.7 (2.6)</td>
<td>34.9 (2.9)</td>
<td>17.5 (1.8)</td>
<td>1.2 (0.4)</td>
<td>14.8 (1.7)</td>
<td>13.9 (1.4)</td>
</tr>
</tbody>
</table>
Factors Associated With Telehealth Use

Individuals who were younger (ORs 1.48-2.23, 18-64 vs ≥75 years), were female (OR 1.33, 95% CI 1.09-1.61), were Hispanic (OR 1.37, 95% CI 1.05-1.80; vs non-Hispanic Whites), or had at least a college education (OR 1.72, 95% CI 1.03-2.87; vs less than high school) had higher odds of using telehealth (Table 4). Unemployed (OR 1.25, 95% CI 1.02-1.54) or insured (OR 1.83, 95% CI 1.25-2.69) adults were more likely to use telehealth. Those who had 1 or more health care office visits in 1 year or had fair or poor general health had higher odds of using telehealth.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Results, adjusted OR&lt;sup&gt;ab&lt;/sup&gt; (95% CI)</th>
<th>P value</th>
<th>Results, unadjusted OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>1.59 (1.15-2.19)</td>
<td>.005</td>
<td>0.98 (0.74-1.29)</td>
<td>.87</td>
</tr>
<tr>
<td>35-49</td>
<td>2.23 (1.60-3.09)</td>
<td>&lt;.001</td>
<td>1.45 (1.11-1.89)</td>
<td>.007</td>
</tr>
<tr>
<td>50-64</td>
<td>1.48 (1.10-1.99)</td>
<td>.01</td>
<td>1.13 (0.87-1.46)</td>
<td>.38</td>
</tr>
<tr>
<td>65-74</td>
<td>1.18 (0.86-1.63)</td>
<td>.23</td>
<td>1.06 (0.79-1.43)</td>
<td>.69</td>
</tr>
<tr>
<td>≥75</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.33 (1.09-1.61)</td>
<td>.004</td>
<td>1.51 (1.26-1.80)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic Black/African American</td>
<td>0.90 (0.67-1.22)</td>
<td>.51</td>
<td>0.85 (0.65-1.11)</td>
<td>.24</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.37 (1.05-1.80)</td>
<td>.02</td>
<td>1.03 (0.82-1.29)</td>
<td>.82</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>1.08 (0.71-1.65)</td>
<td>.72</td>
<td>0.97 (0.61-1.55)</td>
<td>.91</td>
</tr>
<tr>
<td>Others</td>
<td>1.49 (0.88-2.54)</td>
<td>.14</td>
<td>1.52 (0.86-2.71)</td>
<td>.15</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>High school graduate</td>
<td>1.16 (0.70-1.92)</td>
<td>.56</td>
<td>1.93 (1.23-3.03)</td>
<td>.005</td>
</tr>
<tr>
<td>Some college</td>
<td>1.41 (0.84-2.37)</td>
<td>.19</td>
<td>1.19 (0.75-1.89)</td>
<td>.45</td>
</tr>
<tr>
<td>At least a college graduate</td>
<td>1.72 (1.03-2.87)</td>
<td>.04</td>
<td>1.52 (0.95-2.42)</td>
<td>.08</td>
</tr>
<tr>
<td>Household income (US $)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20,000</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>20,000 to &lt;35,000</td>
<td>0.98 (0.70-1.38)</td>
<td>.90</td>
<td>1.00 (0.73-1.38)</td>
<td>.98</td>
</tr>
<tr>
<td>35,000 to &lt;50,000</td>
<td>1.13 (0.74-1.72)</td>
<td>.57</td>
<td>1.08 (0.73-1.59)</td>
<td>.70</td>
</tr>
<tr>
<td>50,000 to &lt;75,000</td>
<td>0.92 (0.67-1.26)</td>
<td>.60</td>
<td>0.94 (0.70-1.26)</td>
<td>.66</td>
</tr>
<tr>
<td>≥75,000</td>
<td>1.19 (0.85-1.67)</td>
<td>.32</td>
<td>1.34 (1.01-1.78)</td>
<td>.04</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>1.25 (1.02-1.54)</td>
<td>.03</td>
<td>1.18 (0.998-1.38)</td>
<td>.053</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1.15 (0.96-1.38)</td>
<td>.14</td>
<td>1.26 (1.05-1.52)</td>
<td>.02</td>
</tr>
<tr>
<td>Unmarried</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Rurality of residence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metropolitan</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Micropolitan</td>
<td>0.78 (0.55-1.11)</td>
<td>.17</td>
<td>0.79 (0.56-1.13)</td>
<td>.19</td>
</tr>
<tr>
<td>Small town</td>
<td>0.96 (0.56-1.64)</td>
<td>.88</td>
<td>0.69 (0.37-1.28)</td>
<td>.24</td>
</tr>
<tr>
<td>Rural</td>
<td>0.70 (0.33-1.47)</td>
<td>.34</td>
<td>1.03 (0.50-2.11)</td>
<td>.93</td>
</tr>
<tr>
<td>Health insurance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.83 (1.25-2.69)</td>
<td>.002</td>
<td>2.63 (1.84-3.76)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Adjusted for all variables in the table.
<sup>b</sup>Reference category.
Discussion

Principal Findings

Our study used nationally representative data to study modern telehealth practices observed in the third year of the COVID-19 pandemic. We comprehensively investigated features of telehealth use including adoption, motivation for use, purposes of use, respondent satisfaction, telehealth perceptions, and associated sociodemographic and health status characteristics of US survey participants. Approximately 2 in 5 adults received telehealth care, motivated primarily by providers’ recommendations and convenience. Care for both acute and chronic conditions was received via telehealth, with high overall patient satisfaction. Additionally, we identified factors associated with telehealth adoption, including younger age, female gender, Hispanic ethnicity, higher education status, being unemployed and insured, previously visiting a health care office, and poor health status. The wide array of descriptive findings and identified associations will help providers and health systems understand the factors that drive patients toward or away from telehealth visits. We illustrated the importance of the role of health policy to support those with low telehealth adoption and of integrating patients’ perspectives [24], as well as how telehealth adoption may improve inequities in health care access for socioeconomically marginalized groups [14].

In the United States, 39.3% of the population adopted telehealth during the third year of COVID-19. Video-only visits were more prevalent than audio-only visits, yet the mode used differed by sociodemographic characteristics. The main users of video visits were younger, educated, employed, insured, or high income–earning (≥US $75,000) adults, while the primary users of audio visits were older (≥75 years), of Hispanic ethnicity, unemployed, or low income–earning (<US $35,000), which is aligned with previous findings [8]. Given that video visits require technology-enabling environments (eg, broadband Internet access, digital device) and health technology literacy, our findings highlight the existence of inequities in video visit access. Although video and audio visits could serve different care needs, it is apparent that efforts will be necessary to improve the accessibility of video visits for disadvantaged groups.

In this study, the biggest barrier to telehealth adoption was a lack of accessibility despite its recent popularity. Although telehealth users’ satisfaction was generally high, preference for in-person visits was still the main reason for not scheduling an available telehealth service. Interestingly, the ability of telehealth visits to include family members in patient visits was a major motivation for its utilization, highlighting the potential role of telehealth in improving the active engagement of patients and caregivers in clinical care, which could improve disease management [25]. Moreover, approximately one-third of participants mentioned that they used telehealth as a means of advice for in-person visits. This suggests that telehealth could help patients make informed health decisions and foster self-care practices, as telehealth could create an enabling environment for individuals to educate themselves using quality guidance for self-care [26]. As a preliminary visit, telehealth could reduce unnecessary office visits, potentially contributing to relieving health professionals’ burden and enhancing health care efficiency in addition to saving patients’ time and effort required to make a visit [27]. Further analyses on the cost-effectiveness of telehealth are suggested to examine the economic and efficiency impact of telehealth on the health care system.

Although acute minor illness care (eg, fever, sinus infection) was the most common purpose of a telehealth visit, telehealth use for chronic condition management (eg, diabetes, high blood pressure) was also high. The findings reveal that the scope of telehealth currently goes beyond acute care and is broadly used throughout multiple practice settings, which aligns with previous findings [28]. We observed the drivers behind telehealth use differed by age, race/ethnicity, and income. Notably, the youngest subgroup used telehealth mainly for mental health care. Non-Hispanic Black/African Americans, other ethnic groups, those with the lowest income, and older adults used televisits mainly for chronic conditions, while non-Hispanic White individuals, non-Hispanic Asian individuals, and those with the highest income used it for acute minor illnesses. The findings contribute to our understanding of how telehealth can improve different care needs for different subpopulations. Need-based targeted promotions could be considered to meet the observed care needs, including telepsychiatry for younger individuals or teleendocrinology and telecardiology for those with a low income or of older age. As the technology matures, we expect the features, implementation, and delivery of telehealth to vary based on specialty, care setting, and acuity. Our findings suggest that future development of telehealth tools may vary depending on the practice setting, including its use for behavior modification [29,30].
Younger individuals (18-64 years) were more likely to use telehealth than the oldest individuals (≥75 years), which aligns with the findings of a recent meta-analysis on telehealth use in cancer care [31]. Previously, older individuals had lower odds of using tablets and smartphones to communicate with providers [32]. Older adults’ low health technology literacy or hearing issues were identified as potential inhibiting factors for the use of health technology, including telehealth [33-35]. In this study, older adults (≥65 years) had higher prevalences of preferring in-person visits and difficulty using telehealth as reasons not to schedule available telehealth visits than their younger counterparts. Our findings suggest that targeted patient education, particularly in the context of health technology literacy or tailored care, could ensure smooth care delivery [35,36].

Women had 30% higher odds of using telehealth than men, which is aligned with existing evidence that women are more likely to make primary care visits [37]. However, during the early phase of the pandemic (March 2020-May 2021), men were nearly twice as likely to use telehealth than women in a surgical oncology center [9]. A notable observation from our study was that the reasons for telehealth visits differed by gender; there was a higher prevalence of telehealth use for mental health care among women (19% vs 12% for men; \( P=0.03 \)). Given that women had higher odds of poor mental health during the pandemic [38], we may need to further examine if telehealth could be a promising tool to address women’s mental health care needs.

Hispanic individuals were more likely to use telehealth than non-Hispanic White individuals. In contrast, during the early pandemic, Hispanic individuals were less likely to use telehealth than in-person office visits [10], and Hispanic cancer survivors were less likely to use email or the Internet to communicate with health providers compared with White cancer survivors [32]. One possible reason for the discrepancy in the findings from previous studies may be that Hispanic individuals were more impacted by COVID-19 infection than individuals of other race/ethnic groups during the early pandemic [39,40], which might have led them to using more hospital visits and less telehealth [10]. Another potential scenario for the high likelihood of telehealth use among Hispanic individuals could be that they may have health issues that could be appropriately handled through telehealth visits. Given that Hispanic individuals had a higher prevalence of other reasons (19%) for recent telehealth visits than non-Hispanic White or Black individuals (13%), further study is warranted to assess if there were unmet health care needs that could be categorized as other reasons in the Hispanic population and if telehealth can sustainably satisfy those [41].

Highly educated individuals (at least a college graduate) had increased odds of using telehealth than the least educated individuals (less than high school). Similarly, a lower education level was a known predictor of low adoption of health technology [42] and technology-based patient-provider communications during the early pandemic [32,43]. Although it is possible that other factors could be at play in this association, it is also possible that less-educated individuals’ concerns about the privacy of telehealth visits might have limited active telehealth use, as we witnessed in this study. Additional steps to ensure telehealth users’ privacy is protected may need to be taken, with continued efforts to secure health information safety.

Insured individuals had 80% higher odds of using telehealth than uninsured counterparts. Given that the major private health plans and Centers for Medicare and Medicaid Services started to cover telehealth since early in the pandemic [3], this is not surprising. Previously, an association was found between a lack of health insurance and low adoption of tablets or smartphones to communicate with providers [32]. Further discussions may need to be initiated on how to support and help uninsured individuals’ low telehealth use. Moreover, unemployed individuals were more likely to use telehealth than employed counterparts. Although there is limited evidence pertaining to this association in the literature, perhaps one interpretation could be that unemployed individuals had more health issues for which they could receive care through telehealth. We observed that unemployed individuals had poor health conditions and used telehealth visits for chronic condition care compared with their employed counterparts. On the other hand, it was notable that employed adults used telehealth primarily for acute minor illnesses than unemployed adults. Although further study is warranted to examine the purpose of telehealth visits by employment status to better understand the dynamics, our findings suggest that a tailored approach to enhance working individuals’ telehealth access and providing more telehealth options outside of regular office hours could be considered.

Individuals who had health care office visits at least once were significantly more likely to use telehealth than those without any office visits yearly, which aligned with the findings of a previous study that examined the relationship between frequent office visits and communicating through electronic health records with providers [32]. This association could also be interpreted as having more health issues and a poorer health condition. For example, individuals with frequent office visits (≥5 times a year) had higher prevalences of poor general health and of using telehealth for chronic condition care compared with those without health office visits. Our findings indicate that active telehealth users were those who were frequent in-person office visitors, rather than those who had been away from health care. Future assessments may need to focus on whether telehealth could accommodate individuals with long-term, consistent care needs as that could potentially reduce the burden on health care professionals particularly in the areas suffering from shortages of the health care labor force.

Limitations
This study had some limitations. First, the temporality and causality of associations cannot be confirmed due to its cross-sectional design. Second, selection bias could be possible given the low response rate of 28.1%. However, HINTS is considered a high-quality national survey, full sample weights were applied to be representative, and imputation was conducted to minimize the bias from nonresponse. Third, as the HINTS is a self-report survey, reporting bias could also be present (eg, general health status is subjective). Fourth, although our data reflect telehealth use in the middle of the pandemic (data collected from March 2022 to November 2022), there may be...
changes in telehealth adoption and the factors associated with low adoption during the postemergency pandemic era. Hence, we suggest future studies to assess if there are any additional changes during the postemergency pandemic era starting in later 2023 and onward. Fifth, we were not able to look at asynchronous messaging as a mechanism of telecare and teledelivery due to data unavailability. Given its substantial use in practice, future studies are warranted to assess its trends and use [44]. Last, we did not consider health technology–related factors, including a technology-enabling environment (eg, digital device ownership, Internet connectivity) and health technology literacy, which are likely associated with telehealth use [45]. We would suggest further investigations into multiple factors at play for telehealth adoption in the future.

Conclusions
Our findings from a nationally representative study of modern telehealth practices show that nearly 2 of 5 individuals in the United States used telehealth, motivated by providers’ recommendations and mainly for acute and chronic condition care, resulting in a positive experience. We identified that an age between 18 years and 64 years, female gender, Hispanic ethnicity, higher education, being unemployed, being insured, frequent health office visits, and poor health status were associated with telehealth use. Future research is warranted to assess the inhibiting factors for those with low telehealth adoption especially where telehealth could be an alternate for in-person care. Furthermore, we recommend evaluating if telehealth satisfies the care needs of those with several health comorbidities. These findings help providers and health systems understand the factors that drive patients to or away from telehealth visits as the technology becomes more routinely available across the United States beyond the needs of the pandemic, providing future directions for telehealth use and telehealth research.

Acknowledgments
Funding for this article was provided by the National Institutes of Health: Mid-career Investigator Award in Patient-Oriented Research (K24AR075060), Research Project Grant (R01AR082109). The funders of this study had no role in study design, data collection and analysis, data interpretation, reporting the findings, or decision to submit the manuscript.

Data Availability
The data for this study are publicly available at the National Cancer Institute website through the standard data access procedure [46].

Authors’ Contributions
All authors contributed to the study design and conduct (acquiring and analyzing data, drafting and revising the manuscript). JK and EL contributed to study conception and study design. JK acquired and curated the data and performed the formal analysis. JK and EL designed the methodology and performed the investigation. JK, ZRC, and MLC wrote the original manuscript, and JK, ZRC, MLC, SO, JMK, and EL conducted several rounds of review and revision. EL supervised the study. All authors critically reviewed and approved the final manuscript before submission. JK and EL had full access to all the data and guarantee the integrity of the work.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Telehealth questionnaires.
[DOCX File, 18 KB - publichealth_v10i1e51279_app1.docx ]

Multimedia Appendix 2
Care purpose of recent telehealth visits by sociodemographic characteristics.
[DOCX File, 20 KB - publichealth_v10i1e51279_app2.docx ]

References


A Profile of Influenza Vaccine Coverage for 2019-2020: Database Study of the English Primary Care Sentinel Cohort

Uy Hoang¹, PhD; Gayathri Delanerolle¹, MSc; Xuejuan Fan¹, PhD; Carole Aspden¹; Rachel Byford¹, BA; Mansoor Ashraf², MBBS; Mendel Haag³, PhD; William Elson¹, MBBS; Meredith Leston¹, MSc; Sneha Anand¹, PhD; Filipa Ferreira¹, PhD; Mark Joy¹, PhD; Richard Hobbs¹, MA; Simon de Lusignan¹, MD

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
²CSL Seqirus, Maidenhead, United Kingdom
³CSL Seqirus, Amsterdam, Netherlands

Abstract

Background: Innovation in seasonal influenza vaccine development has resulted in a wider range of formulations becoming available. Understanding vaccine coverage across populations including the timing of administration is important when evaluating vaccine benefits and risks.

Objective: This study aims to report the representativeness, uptake of influenza vaccines, different formulations of influenza vaccines, and timing of administration within the English Primary Care Sentinel Cohort (PCSC).

Methods: We used the PCSC of the Oxford-Royal College of General Practitioners Research and Surveillance Centre. We included patients of all ages registered with PCSC member general practices, reporting influenza vaccine coverage between September 1, 2019, and January 29, 2020. We identified influenza vaccination recipients and characterized them by age, clinical risk groups, and vaccine type. We reported the date of influenza vaccination within the PCSC by International Standard Organization (ISO) week. The representativeness of the PCSC population was compared with population data provided by the Office for National Statistics. PCSC influenza vaccine coverage was compared with published UK Health Security Agency’s national data.

Results: The PCSC comprised 7,010,627 people from 693 general practices. The study population included a greater proportion of people aged 18-49 years (2,982,390/7,010,627, 42.5%; 95% CI 42.5%-42.6%) compared with the Office for National Statistics 2019 midyear population estimates (23,219,730/56,286,961, 41.3%; 95% CI 4.12%-41.3%; P<.001). People who are more deprived were underrepresented and those in the least deprived quintile were overrepresented. Within the study population, 24.7% (1,731,062/7,010,627; 95% CI 24.7%-24.7%) of people of all ages received an influenza vaccine compared with 24.2% (14,468,665/59,764,928; 95% CI 24.2%-24.2%; P<.001) in national data. The highest coverage was in people aged ≥65 years (913,695/1,264,700, 72.3%; 95% CI 72.2%-72.3%). The proportion of people in risk groups who received an influenza vaccine was also higher; for example, 69.8% (284,280/407,228; 95% CI 69.7%-70%) of people with diabetes in the PCSC received an influenza vaccine compared with 61.2% (983,727/1,607,996; 95% CI 61.1%-61.3%; P<.001) in national data. In the PCSC, vaccine type and brand information were available for 71.8% (358,365/498,923; 95% CI 71.7%-72%) of people aged 16-64 years and 81.9% (748,312/913,695; 95% CI 81.8%-82%) of people aged ≥65 years, compared with 23.6% (696,880/2,900,000) and 17.8% (1,385,888/7,700,000), respectively, of the same age groups in national data. Vaccination commenced during ISO week 35, continued until ISO week 3, and peaked during ISO week 41. The in-week peak in vaccination administration was on Saturdays.

Conclusions: The PCSC’s sociodemographic profile was similar to the national population and captured more data about risk groups, vaccine brands, and batches. This may reflect higher data quality. Its capabilities included reporting precise dates of administration. The PCSC is suitable for undertaking studies of influenza vaccine coverage.
Introduction

The range of influenza vaccines available for clinical use has grown; there are now live attenuated influenza vaccines and recombinant vaccines in addition to the long-established inactivated influenza vaccine [1,2]. Live attenuated influenza vaccines have also been widely introduced to children via intranasal administration [3,4]. There have also been many new formulations of inactivated influenza vaccine (IIV). IIV are now often quadrivalent (active against 2 influenza B lineages, as well as 2 influenza A subtypes) rather than trivalent (which has a single B lineage), and their effectiveness may be increased by the addition of an adjuvant or higher antigen dosages [4]. Cell-based manufacture of IIV has been introduced in addition to traditional egg-based methods. Recombinant influenza vaccines are also being introduced, although these formulations are currently less widely used [5,6]. The available evidence indicates additional benefits from the use of newer and enhanced influenza vaccine formulations [7].

The Joint Committee on Vaccination and Immunization (JCVI) provides UK-wide advice on vaccination policy, including for seasonal influenza vaccine. The JCVI has progressively updated its recommendations as new vaccines have been introduced. These include the use of adjuvanted trivalent inactivated influenza vaccine (aIIV3) and high-dose trivalent inactivated influenza vaccine (HD-IIV3) for use in the 2018-2019 season and the use of cell-based quadrivalent inactivated influenza vaccine (IIV4c) vaccine for the 2019-2020 influenza season [8]. IIV4c was approved for adults and children from 9 years of age in at-risk groups and was available for use starting in the 2019-2020 season in the United Kingdom [9]. For the 2021-2022 season, a non-egg–based quadrivalent vaccine based on recombinant technology was added to the recommendations put forward by the JCVI for adults aged between 18 and 65 years [7]. For adults aged 65 years or older, 2 vaccine formulations were recommended by the JCVI in the 2021-2022 season, namely aIIV4, which replaced the trivalent formulation; aIIV3; and high-dose quadrivalent inactivated influenza vaccine (HD-IIV4) [7].

Additionally, the JCVI advised that the relative effectiveness of different enhanced influenza vaccine formulations should be assessed. Comparative influenza vaccine effectiveness data are required, preferably from the same country over multiple seasons and with laboratory-confirmed influenza end points [7,10].

In the context of an ever-increasing range of influenza vaccine formulations, we report influenza vaccine coverage in the English Primary Care Sentinel Cohort (PCSC) in the 2019-2020 winter season. We describe the representativeness of the PCSC compared with the national population, the uptake of influenza vaccination within this cohort (differentiated by formulation and also compared with national data), and the timing of vaccine administration across age bands.

Methods

Study Design and Setting

In this study, we compare the PCSC with national data for population representativeness and vaccine uptake by formulation and provide a descriptive report by week of vaccine uptake. These elements were included because vaccine studies can either be analyzed at the national population level or run in representative samples. It is important to know if there is sufficient representation of the different influenza vaccine formulations and to understand the timings of vaccine administration for studies of adverse events of interest or those of vaccine effectiveness compared with known dates of cocirculating strains of influenza [11].

This study was conducted during the 2019-2020 influenza season. The 2020-2021 season was not used as it was considered by authors as atypical due to the COVID-19 pandemic and associated changes in vaccination schedules (inclusion of persons aged 50 to 64 years and an absence of circulating influenza as a result of stay-at-home mandates) [12]. This study was conducted using the PCSC of the Oxford Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC). At the time of this study, the RSC was, and remains, one of the largest and longest-established primary care sentinel networks in Europe [13,14].

At the time of this study, the RSC had been recording respiratory conditions, including influenza and notifiable diseases, and had provided data to support clinical and public health research (most notably on vaccine effectiveness) for over 50 years [3,13-16]. Data from the RSC were, and still are, hosted within the Oxford RCGP Clinical Informatics Digital Hub, a trusted research environment [17]. The RSC as a whole comprises over 1800 general practices in England, and the total current registered list includes over 19 million people—approximately 32% of the national population (Figure 1).
The RSC adapted and grew over the course of the COVID-19 pandemic when it was divided into 2 divisions [18]. First, the PCSC is the long-standing sentinel group of general practices (n=693) that conduct virology sampling (n=240) and serosurveillance (n=273) [19,20]. This group of practices has a long history of receiving feedback about data quality, historically at an annual visit and, in recent years, through dashboards, webinars, and web-based visits [21,22]. They were recruited to be nationally representative [13]. The second group, the Syndromic Surveillance General Practices (SSGPs), was recruited to support the UK Health Security Agency’s (UKHSA; formerly Public Health England) syndromic surveillance [23]. These practices were recruited to complement other data sources and were not nationally representative.

Representativeness of the PCSC, Including Risk Groups

The study population included participants of all ages in England who were registered with a general practitioner (GP) in the PCSC between September 1, 2019, and January 29, 2020. We describe the demographic characteristics of the study population in terms of the following:

- Age, including the following subgroups: ≤1 year, ≥2 to ≤3 years, ≥4 to ≤17 years, ≥18 to ≤49 years, ≥50 to ≤64 years, and ≥65 years.
- Sex: male, female, and missing.
- Race: Asian, Black, White, mixed, and missing or other, using an ontology to maximize identification [24].
- Socioeconomic status using the quintile of the Index of Multiple Deprivation (IMD) as a proxy (a nationally available measure assigned based on postcode). This provides an overall relative measure of deprivation for each Lower Layer Super Output Area (LSOA). An LSOA is a small area with an average population of 1500 people. The overall deprivation scores are ranked for all LSOAs within a country and can be divided into 5 groups (quintiles), where quintile 1 represents the most deprived LSOAs and quintile 5 represents the least deprived LSOAs. The IMD is a score based on the area as a whole and not everyone within an LSOA necessarily experiences the same level or type of deprivation [25,26].
- Geographic location is defined according to National Health Service (NHS) region [27].
- Risk factors for influenza as defined by the Chief Medical Officer, include chronic pulmonary disease, asthma, coronary health disease, diabetes, liver disease, chronic kidney disease, neurological conditions, immunosuppression, asplenia, learning difficulties, severe mental illness, and obesity [28].

To evaluate the representativeness of the PCSC in terms of the demographic characteristics above, we undertook a quantitative comparison of the age, sex, and IMD profile of the PCSC against published data for 2019-2020 from the Office for National Statistics (ONS) for England [29]. We compared the ethnic profile of the PCSC against ONS ethnic population estimates extrapolated from the 2011 census [30]. We also compared the broad geographic distribution of the PCSC population in terms of NHS regions with the data from the ONS on the geographical distribution of the population by government office regions [27,31]. Last, we compared the distribution of people with risk factors for influenza defined by the Chief Medical Officer in the PCSC with data from the Quality and Outcomes Framework 2019-2020 primary care disease prevalence figures published by NHS Digital [28,32]. We made 1 near-match using the...
Influenza Vaccine Coverage
The exposure of interest was influenza vaccination. Influenza vaccination was either recorded as a clinical term for vaccine administration or as a prescription. A large number of doses of influenza vaccines are administered by pharmacies and other community health care workers, although the majority of these vaccines are recorded in the primary care computerized medical record (CMR). Where vaccination was recorded as a clinical term in the CMR, it may not have been specific to the formulation, brand, or batch; when vaccination was prescribed, it was brand and batch specific. Prescriptions were only issued for adult and high-risk children who were vaccinated in their general practice [34]. A list of Systematized Nomenclature of Medicine clinical terms used to identify influenza vaccinations from the CMR of patients in the PCSC is listed in Table S1 in Multimedia Appendix 1 [27,35].

Vaccine coverage in the study population was described by providing the count and proportion (expressed as a percentage) of the PCSC vaccinated against influenza overall and by the following:

- Number and percentage of vaccinated individuals.
- Number and percentage of vaccinated individuals by type of vaccine.
- Timing of vaccination by week and day of the week.
- Number and percentage of vaccinated individuals by geographic location.
- Risk factors for influenza as defined by the Chief Medical Officer [28].

We could not precisely match the geographic locations. Our nearest match was to add East and West Midlands government regions to be equivalent to our Midlands region; we also combined Yorkshire and the Humber and the North East to be equivalent to our North East and Yorkshire regions.

To evaluate the representativeness of influenza vaccine coverage in the PCSC, a quantitative comparison was made to the influenza vaccine coverage for the winter season—September 1, 2019, to January 29, 2020—reported by the UKHSA [35]. We compared the influenza coverage in the PCSC according to the following inclusive age groups reported in national data by the UKHSA: 6 months to 1 year, 2–4 years, 5–15 years, 16–64 years, and ≥65 years. For influenza coverage by type of vaccine, coverage in the PCSC was compared with national data from the UKHSA in the following age groups only: 16–64 years and ≥65 years. We again made some near-matches for at-risk groups: coronary heart disease is compared with chronic heart disease, liver disease with chronic liver disease, neurological disease with chronic neurological disease, and obesity with morbid obesity.

Influenza Vaccination Timing
In order to evaluate the timing of influenza vaccinations given in the PCSC, we calculated the vaccine coverage overall in the study population by week and day of the week and recorded this information in line with its relevant International Standard Organization (ISO) week number.

Statistical Methods
For categorical measures, we calculated the frequency and percentage of total study participants observed in each category. For continuous and count variables, we presented the mean, SD, median, and range.

For proportion counts, we calculated 95% CI using the normal approximation to the binomial calculation [36]. Paired 2-tailed t tests were used to compare populations [37].

Statistical analyses were undertaken using R (version 3.5.1; R Core Team) [38].

Ethical Considerations
This study was classified as a service evaluation (measuring what standard of care this service achieved) by the Medical Research Council/Health Research Authority decision tool [39], so it did not require formal ethical approval. It was reviewed by the RCGP Approval Committee on January 8, 2021.

Results
PCSC Population and Comparison With English National Population
A total of 7,010,627 people (male: n=3,488,789, 49.8%; female: n=3,521,838, 50.2%) were registered with a primary care practice from the PCSC that submitted data for this study (n=693) between September 1, 2019, and January 29, 2020. The study population had a greater proportion of people aged 18–49 years (2,982,390/7,010,627, 42.5%; 95% CI 42.5%–42.6%) compared with the 2019 midyear population estimates from ONS (23,219,730/56,286,961, 41.3%; 95% CI 41.2%–41.3%; P<.001; Table 1) [29]. Populations in the most deprived quintile, namely those in IMD quintile 1, were underrepresented in this cohort, and those in the least deprived quintile, namely those in IMD quintile 5, were overrepresented.
Table 1. Demographic profile of the Primary Care Sentinel Cohort compared with the English national population.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Study population, 2019-2020 (n=7,010,627), n (%)</th>
<th>2019 ONS\textsuperscript{a} midyear population estimates for England [29] (n=56,286,961), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall population</td>
<td>7,010,627 (100)</td>
<td>56,286,961 (100)</td>
</tr>
<tr>
<td><strong>Age range (years, inclusive)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>158,484 (2.3)</td>
<td>1,262,914 (2.2)</td>
</tr>
<tr>
<td>2-3</td>
<td>157,857 (2.3)</td>
<td>1,351,731 (2.4)</td>
</tr>
<tr>
<td>4-17</td>
<td>1,126,575 (16.1)</td>
<td>9,408,923 (16.7)</td>
</tr>
<tr>
<td>18-49</td>
<td>2,982,390 (42.5)</td>
<td>23,219,730 (41.3)</td>
</tr>
<tr>
<td>50-64</td>
<td>1,320,621 (18.8)</td>
<td>10,689,947 (19)</td>
</tr>
<tr>
<td>≥65</td>
<td>1,264,700 (18)</td>
<td>10,353,716 (18.4)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3,488,789 (49.8)</td>
<td>27,827,831 (49.4)</td>
</tr>
<tr>
<td>Female</td>
<td>3,521,838 (50.2)</td>
<td>28,459,130 (50.6)</td>
</tr>
<tr>
<td><strong>Race\textsuperscript{b}</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>496,330 (7.1)</td>
<td>4,143,403 (7.8)</td>
</tr>
<tr>
<td>Black</td>
<td>210,501 (3)</td>
<td>1,846,614 (3.5)</td>
</tr>
<tr>
<td>White</td>
<td>4,674,668 (66.7)</td>
<td>45,281,142 (85.4)</td>
</tr>
<tr>
<td>Mixed</td>
<td>114,846 (1.6)</td>
<td>1,192,879 (2.3)</td>
</tr>
<tr>
<td>Missing or other</td>
<td>1,514,282 (21.6)</td>
<td>548,418 (1)</td>
</tr>
<tr>
<td><strong>IMD\textsuperscript{c} quintile</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (most deprived)</td>
<td>1,245,558 (17.8)</td>
<td>11,267,059 (20)</td>
</tr>
<tr>
<td>2</td>
<td>1,353,953 (19.3)</td>
<td>11,576,973 (20.6)</td>
</tr>
<tr>
<td>3</td>
<td>1,409,593 (20.1)</td>
<td>11,424,153 (20.3)</td>
</tr>
<tr>
<td>4</td>
<td>1,451,445 (20.7)</td>
<td>11,117,694 (20.3)</td>
</tr>
<tr>
<td>5 (least deprived)</td>
<td>1,548,132 (22.1)</td>
<td>10,901,082 (19.4)</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London</td>
<td>1,025,002 (14.6)</td>
<td>8,961,989 (15.9)</td>
</tr>
<tr>
<td>East of England</td>
<td>12,122 (7.3)</td>
<td>6,236,072 (11.1)</td>
</tr>
<tr>
<td>Midlands</td>
<td>1,000,656 (14.3)</td>
<td>10,769,965 (19.1)</td>
</tr>
<tr>
<td>North East and Yorkshire</td>
<td>823,102 (11.7)</td>
<td>8,172,908 (14.5)</td>
</tr>
<tr>
<td>North West</td>
<td>1,077,808 (15.4)</td>
<td>7,341,196 (13)</td>
</tr>
<tr>
<td>South East</td>
<td>1,339,701 (19.1)</td>
<td>9,180,135 (16.3)</td>
</tr>
<tr>
<td>South West</td>
<td>1,232,236 (17.6)</td>
<td>5,624,696 (10)</td>
</tr>
<tr>
<td><strong>Risk factor\textsuperscript{d}</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>647,526 (9.2)</td>
<td>1,891,019 (3.1)\textsuperscript{f}</td>
</tr>
<tr>
<td>Asplenia</td>
<td>36,199 (0.5)</td>
<td>N/A\textsuperscript{f}</td>
</tr>
<tr>
<td>Asthma</td>
<td>488,596 (7)</td>
<td>3,916,150 (6.5)\textsuperscript{f}</td>
</tr>
<tr>
<td>COPD\textsuperscript{e}</td>
<td>205,107 (2.9)</td>
<td>1,170,786 (1.9)\textsuperscript{f}</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>267,091 (3.8)</td>
<td>1,949,865 (4.1)\textsuperscript{h}</td>
</tr>
<tr>
<td>Liver disease</td>
<td>107,984 (1.5)</td>
<td>N/A</td>
</tr>
<tr>
<td>Diabetes</td>
<td>407,228 (5.8)</td>
<td>3,455,176 (7.1)\textsuperscript{h}</td>
</tr>
</tbody>
</table>
Influenza Vaccine Coverage in the PCSC Compared With National Data

A total of 1,731,062 (24.7%; 95% CI 24.7%-24.7%) out of 7,010,627 people within the PCSC received an influenza vaccine (male: 790,660/1,731,062, 22.7%; female: 940,402/1,731,062, 26.7%). We subdivided vaccine coverage in the PCSC population by sociodemographic characteristics (Table 2 and Table S2 in Multimedia Appendix 1) and calculated the proportion of the PCSC population who were vaccinated in each subpopulation stratum. For example, 72.3% (913,695/1,264,700; 95% CI 72.2%-72.3%) of people aged 65 years or older in the PCSC were vaccinated, whereas the national data showed that 72.4% (7,621,505/10,523,854) of individuals aged 65 years or older received seasonal influenza vaccine [35], with a coverage of 24.2% (14,468,665/59,764,928; 95% CI 24.2%-24.2%; \( P < .001 \)). National coverage data only provided information on limited age bands, making direct comparisons between cohorts challenging (see Table S2 in Multimedia Appendix 1). Coverage of risk groups was slightly higher in our study than in comparable national data; for example, 69.8% (284,280/407,228; 95% CI 69.7%-70%) of people with diabetes in our study received an influenza vaccine whereas 61.2% (983,727/1,607,996; 95% CI 61.1%-61.3%; \( P < .001 \)) of people with diabetes received an influenza vaccine according to national data (Table 3).
Table 2. Influenza vaccine coverage in strata of age, sex, race, Index of Multiple Deprivation, and region in the Primary Care Sentinel Cohort.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Vaccinated, n/N (%)</th>
<th>Vaccinated (%), 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>1216/158,484 (0.8)</td>
<td>0.7-0.8</td>
</tr>
<tr>
<td>2-3</td>
<td>63,532/157,857 (40.3)</td>
<td>40.0-40.5</td>
</tr>
<tr>
<td>4-17</td>
<td>259,902/1,126,575 (23.1)</td>
<td>23.1-23.2</td>
</tr>
<tr>
<td>18-49</td>
<td>216,040/2,982,390 (7.2)</td>
<td>7.2-7.3</td>
</tr>
<tr>
<td>50-64</td>
<td>276,677/1,320,621 (21)</td>
<td>20.9-21</td>
</tr>
<tr>
<td>≥65</td>
<td>913,695/1,264,700 (72.3)</td>
<td>72.2-72.3</td>
</tr>
<tr>
<td>All ages</td>
<td>1,731,062/7,010,627 (24.7)</td>
<td>24.7-24.7</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>790,660/3,488,789 (22.7)</td>
<td>22.6-22.7</td>
</tr>
<tr>
<td>Female</td>
<td>940,402/3,521,838 (26.7)</td>
<td>26.7-26.8</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>82,293/496,330 (16.6)</td>
<td>16.5-16.7</td>
</tr>
<tr>
<td>Black</td>
<td>29,884/210,501 (14.2)</td>
<td>14.1-14.4</td>
</tr>
<tr>
<td>White</td>
<td>1,325,892/4,674,668 (28.4)</td>
<td>28.3-28.4</td>
</tr>
<tr>
<td>Mixed</td>
<td>16,797/114,846 (14.6)</td>
<td>14.4-14.8</td>
</tr>
<tr>
<td>Missing or other</td>
<td>276,196/1,514,282 (18.2)</td>
<td>18.2-18.3</td>
</tr>
<tr>
<td><strong>IMDb quintile</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (most deprived)</td>
<td>246,536/1,245,558 (19.8)</td>
<td>19.7-19.9</td>
</tr>
<tr>
<td>2</td>
<td>297,114/1,353,953 (21.9)</td>
<td>21.9-22</td>
</tr>
<tr>
<td>3</td>
<td>351,594/1,409,593 (24.9)</td>
<td>24.9-25</td>
</tr>
<tr>
<td>4</td>
<td>390,403/1,451,445 (26.9)</td>
<td>26.8-27</td>
</tr>
<tr>
<td>5 (least deprived)</td>
<td>444,876/1,548,132 (28.7)</td>
<td>28.7-28.8</td>
</tr>
<tr>
<td><strong>NHSb region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>London</td>
<td>148,930/1,025,002 (14.5)</td>
<td>14.5-14.6</td>
</tr>
<tr>
<td>East of England</td>
<td>131,758/1,122,122 (25.7)</td>
<td>25.6-25.9</td>
</tr>
<tr>
<td>Midlands</td>
<td>256,769/1,000,656 (25.7)</td>
<td>25.6-25.8</td>
</tr>
<tr>
<td>North East and Yorkshire</td>
<td>226,792/823,102 (27.6)</td>
<td>27.5-27.7</td>
</tr>
<tr>
<td>North West</td>
<td>277,648/1,077,808 (25.8)</td>
<td>25.7-25.8</td>
</tr>
<tr>
<td>South East</td>
<td>337,225/1,339,701 (25.2)</td>
<td>25.1-25.3</td>
</tr>
<tr>
<td>South West</td>
<td>351,940/1,232,236 (28.6)</td>
<td>28.5-28.6</td>
</tr>
</tbody>
</table>

*aIMD: Index of Multiple Deprivation.  
bNHS: National Health Service.*
Table 3. Proportion of risk groups who are vaccinated in the Primary Care Sentinel Cohort (PCSC) compared with national data.

<table>
<thead>
<tr>
<th>Risk group</th>
<th>PCSC, 2019-2020, n/N (%)</th>
<th>National data [35], n (%)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary heart disease vs chronic heart disease</td>
<td>419,417/647,526 (64.8)</td>
<td>432,939/1,042,670 (41.1)</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Asplenia</td>
<td>19,895/36,199 (55)</td>
<td>127,437/379,520 (33.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Asthma</td>
<td>262,158/488,596 (53.7)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Chronic pulmonary disease vs chronic respiratory disease</td>
<td>148,268/205,107 (72.3)</td>
<td>1,533,903/3,108,241 (48.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>199,248/267,091 (74.6)</td>
<td>175,415/342,661 (50.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Liver disease vs chronic liver disease</td>
<td>52,271/107,984 (48.4)</td>
<td>212,048/562,410 (37.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Diabetes</td>
<td>284,280/407,228 (69.8)</td>
<td>994,675/1,607,996 (61.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Immunosuppression</td>
<td>87,748/136,583 (64.2)</td>
<td>188,198/423,273 (44)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Neurological disease vs chronic neurological disease</td>
<td>214,685/366,963 (58.5)</td>
<td>378,549/883,590 (42.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Severe mental illness</td>
<td>21,509/66,319 (32.4)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Learning difficulties</td>
<td>16,219/35,327 (45.9)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Obesity vs morbid obesity</td>
<td>70,142/162,397 (43.2)</td>
<td>167,050/532,494 (30.9)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup>Population comparisons are not provided where we used a near-match, for example, coronary heart disease with chronic heart disease.

<sup>b</sup>N/A: not applicable.

In the PCSC, vaccine type and brand information were available for 71.8% (358,365/498,923; 95% CI 71.7%-72%) of influenza vaccines administered to people aged 16 to 64 years (Table 4 and Table S3 in Multimedia Appendix 1) and 81.9% (748,312/913,695; 95% CI 81.8%-82%) of those administered to people aged 65 years or older (Table 5 and Table S3 in Multimedia Appendix 1). This compares with national data where only 23.6% (696,880/2,900,000) of seasonal influenza vaccinations for people aged 16 to 64 years (Table 4) and 17.8% (1,385,888/7,700,000) for those aged 65 years or older were coded with a defined vaccine type in 2019-2020 (Table 5) [35].

Table 4. Type of influenza vaccine given to people aged 16-64 years who received an influenza vaccine in the Primary Care Sentinel Cohort (PCSC) and national data (as available).

<table>
<thead>
<tr>
<th>Vaccine type</th>
<th>Vaccine recipients in PCSC 2019-2020 (n=498,923), n (%)</th>
<th>Vaccine recipients in national data [35] (n=2,900,000), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIV4&lt;sup&gt;a&lt;/sup&gt; (specified)</td>
<td>9281 (1.9)</td>
<td>564,000 (19.1)</td>
</tr>
<tr>
<td>IIV4&lt;sup&gt;b&lt;/sup&gt;</td>
<td>307,617 (61.7)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>IIV4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>41,233 (8.3)</td>
<td>132,000 (4.5)</td>
</tr>
<tr>
<td>IIV3&lt;sup&gt;d&lt;/sup&gt;</td>
<td>234 (0)</td>
<td>40 (0)</td>
</tr>
<tr>
<td>Missing brand or vaccine type</td>
<td>140,558 (28.2)</td>
<td>2,200,000 (76.4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>IIV4: quadrivalent inactivated influenza vaccine.

<sup>b</sup>IIV4: egg-based quadrivalent inactivated influenza vaccine.

<sup>c</sup>N/A: Not available.

<sup>d</sup>IIV4: cell-based quadrivalent inactivated influenza vaccine.

<sup>e</sup>IIV3: nonadjuvanted (standard) trivalent inactivated influenza vaccine.
Table 5. Type of influenza vaccine given to people aged 65 years or older who received an influenza vaccine in the Primary Care Sentinel Cohort (PCSC) and national data (as available).

<table>
<thead>
<tr>
<th>Vaccine type</th>
<th>Vaccine recipients in PCSC 2019-2020 (n=913,695), n (%)</th>
<th>Vaccine recipients in national data [35] (n=7,700,000), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIV4(^a) (unspecified)</td>
<td>2288 (0.3)</td>
<td>19,000 (0.2)</td>
</tr>
<tr>
<td>IIV4(^b)</td>
<td>4331 (0.5)</td>
<td>N/A(^c)</td>
</tr>
<tr>
<td>IIV4(^d)</td>
<td>59,563 (6.5)</td>
<td>251,000 (3.2)</td>
</tr>
<tr>
<td>aIIV3(^e)</td>
<td>680,961 (74.5)</td>
<td>1,100,000 (14.3)</td>
</tr>
<tr>
<td>IIV3(^f) (unspecified)</td>
<td>1169 (0.1)</td>
<td>3000 (0)</td>
</tr>
<tr>
<td>Missing brand or vaccine type</td>
<td>165,383 (18.1)</td>
<td>6,400,000 (82.2)</td>
</tr>
</tbody>
</table>

\(^a\)IIV4: quadrivalent inactivated influenza vaccine.
\(^b\)IIV4\(^e\): egg-based quadrivalent inactivated influenza vaccine.
\(^c\)N/A: Not available.
\(^d\)IIV4\(^c\): cell-based quadrivalent inactivated influenza vaccine.
\(^e\)aIIV3: adjuvanted trivalent inactivated influenza vaccine.
\(^f\)IIV3: nonadjuvanted (standard) trivalent inactivated influenza vaccine.

Timing of Influenza Vaccination in PCSC

During the 2019-2020 influenza season in the PCSC, influenza vaccination started in ISO week 35 and continued until ISO week 3. Peak of vaccinations occurred in ISO week 41 (Figures 2 and 3). Weekly peaks in vaccination coverage occurred on Saturdays, especially among those aged 65 years or older. People aged 65 years or older received their influenza vaccinations earlier in the season than other age groups, with a peak around ISO week 40.

Figure 2. Timing of influenza vaccination across all ages in the Primary Care Sentinel Cohort. ISO: International Standard Organization.
Discussion

Principal Findings

The population of patients in the English national PCSC is broadly comparable to the general population in terms of demographic features. However, there is a greater proportion of people aged 18-49 years in the PCSC, and people who are more deprived are underrepresented and those in the least deprived quintile overrepresented. It is possible this may result in denominator inflation in the younger half of the working-age population for any influenza vaccine coverage estimates.

Overall, influenza vaccine coverage is comparable in the PCSC to national data, although coverage of risk groups was higher in the PCSC than in comparable published national data. Influenza vaccine type was recorded for a higher proportion of vaccines recorded in the PCSC than available from national data, although there are still a substantial number of vaccinations in the PCSC for which no influenza vaccine type or brand information was recorded.

The timing of influenza vaccination during the 2019-2020 season showed that influenza vaccination started in ISO week 35, with people aged 65 years or older receiving their vaccinations earlier in the season than other age groups. Weekly peaks in vaccination coverage were seen on Saturdays, especially among those aged 65 years or older.

Implications of the Findings

Our study demonstrated the representativeness of the PCSC in describing vaccine exposure, including timing, vaccine type, and brand-specific exposure. However, although we took steps to avoid denominator inflation, it is possible that our findings may be affected by an inflated denominator in the younger half of the working-age population (age 18 to 49 years). Denominator inflation may be a result of systematic differences in PCSC practices versus those practices outside the PCSC that have not been possible to quantify in this study, including opening times, accessibility, range of services offered, and patient satisfaction.

PCSC data about individual vaccine exposure, if combined with data on infection status, could be used to estimate vaccine effectiveness. The timing of vaccination in relation to the start of the season would allow comparison of individual vaccine exposure with circulating influenza virus strains. Where a larger population is required for a study, we could also include the RSC’s SSGPs. As the aggregated PCSC and SSGP population is over 19 million (32% of the national population; Figure 1), their combined representativeness would support a national study for which a larger sample size was needed. The RSC could also link individual exposure data to clinical records or conduct enhanced surveillance to monitor adverse events of interest.

The PCSC data did not capture the influenza vaccine type and brand for all vaccinations. Steps should be taken to improve data quality to enable comparative influenza vaccine effectiveness studies to be conducted. Of note for researchers looking to collect further information at the time of vaccination, many general practices run their influenza vaccination clinics on Saturdays, particularly for older people.

Strengths and Limitations

Strengths of the study included its large sample size from a well-established sentinel network that includes primary care practitioners who have been recording information on influenza incidence and vaccinations for over 50 years [14]. Additionally, given the completeness of our data on influenza vaccine type, we were able to include a wider range of influenza vaccine types in the PCSC than comparable national data [35]. However, the limited amount of comparable national data also curtailed our quantitative comparison of influenza vaccine coverage. It was especially difficult to compare the proportions of those vaccinated within at-risk groups as there was no detailed...
information available in national data. We have presented the proportion of people in at-risk groups vaccinated against influenza as a proportion of all the people in the at-risk group who were registered that year. For example, 284,280 people with diabetes received an influenza vaccine versus 407,228 people with diabetes. Where we identified a higher proportion of risk groups, we think this likely indicated better data quality; however, this may be a statistical artifact of differences in disease prevalence. Finally, we could not precisely match some risk groups.

The COVID-19 pandemic has resulted in interruptions of health services since early 2020. The indirect effects this had on routine vaccinations in primary care may have also affected the results of this study. However, the influenza vaccination programs had largely been completed by early 2020, and we felt confident in including 2019-2020 in our analysis. That said, this limited us to including only 1 year of data in our analysis. Furthermore, a large number of doses of influenza vaccines are administered by pharmacies and other community health care workers [40]. Although the majority of these vaccines are recorded in the CMR, communications from pharmacists to GPs around vaccination records have been described as suboptimal, especially with regard to data on manufacturer or batch number [34]. This may have affected the completeness and accuracy of our data, specifically, incomplete data about influenza vaccinations provided by pharmacies and other community health care workers would serve to underestimate the vaccine coverage rate we present in this paper.

Comparison With Prior Work
Other authors have reported strengths and limitations in the personal demographics service (described in detail in our method) and NHS number as a unique identifier [41]. Although a national identifier is overall a great strength, we need to conduct further studies to explore whether there is any denominator inflation in the PCSC database.

Past work codifying COVID-19 vaccination records in primary care settings has shown a high level of completeness of vaccine brand recording for this vaccination program. The National Immunization Management Service has recorded 99.6% of vaccine administration details and electronically transmits brand information to every individual’s GP record on a daily basis [42]. This compares with between 43.8% and 81.9% of influenza vaccine records having vaccine type and brand information accessible in the PCSC, depending on age.

Influenza vaccine effectiveness assessment is increasingly stratified by vaccine type or even vaccine brand to enable comparisons of different formulations that help to inform public health decision-making [43,44].

The United Kingdom is a possible candidate country to source comparative data for the newer influenza vaccine formulations, as it has historically had high annual influenza vaccine coverage for those aged 65 years or older. Between the 2004-2005 and 2019-2020 seasons, average coverage ranged from 71% to 75%; these figures are relatively high compared with the rest of Europe [45,46]. The World Health Organization has set a vaccination coverage target of at least 75% in the older adult population and among risk groups [47]. Influenza vaccine coverage among at-risk groups <65 years of age, excluding pregnant women, in England has been lower—approximately 49% between the 2007-2008 and 2016-2017 seasons [48,49].

Conclusions
The English PCSC was broadly representative of the national population. It included high-quality routine data, in terms of sociodemographic risk groups, recording of risk groups, and influenza uptake compared with national data sources. However, although data quality was reported to be good, there were significant gaps in vaccine type and brand. Methods implemented to capture COVID-19 vaccine data and encode it within general practice CMRs should now be implemented for influenza vaccines. The PCSC of the RSC, and the UK health care ecosystem itself, provide favorable environments for conducting influenza vaccine benefit-risk studies.

Acknowledgments
The authors thank Oxford Royal College of General Practitioners Research and Surveillance Centre member practices who shared data as well as patients and their parents or caregivers for not opting out of data sharing; EMIS, TPP System One, In Practice Systems, and Wellbeing for their collaboration with pseudonymized data extraction; and James Mansi, formerly from Seqirus, for his initial help and advice. The manuscript was also reviewed by Seqirus prior to submission.

Authors’ Contributions
SL, UH, and MA were involved in the planning of this study. All other authors were involved in the conduct and reporting of this work.

Conflicts of Interest
SL is the Director of the Oxford Royal College of General Practitioners Research and Surveillance Centre. SL has undertaken projects funded by AstraZeneca GSK, Takeda, and Seqirus, and has been a member of Advisory Boards for AstraZeneca Seqirus and Sanofi. MA and MH are employees of Seqirus Ltd.

Multimedia Appendix 1
References


38. R FAQ. CRAN. 2018 Aug 05. URL: https://cran.r-project.org/doc/FAQ/FAQ.html#What-is-R_003F [accessed 2024-03-09]


Abbreviations

- aIIV3: adjuvanted trivalent inactivated influenza vaccine
- CMR: computerized medical record
- GP: general practitioner
- HD-IIV3: high-dose trivalent inactivated influenza vaccine
- HD-IIV4: high-dose quadrivalent inactivated influenza vaccine
- IIV: inactivated influenza vaccine
- IIV4c: cell-based quadrivalent inactivated influenza vaccine
- IMD: Index of Multiple Deprivation
- ISO: International Standard Organization
- JCVI: Joint Committee on Vaccination and Immunization
- LSOA: Lower Layer Super Output Area
- NHS: National Health Service
- ONS: Office for National Statistics
- PCSC: Primary Care Sentinel Cohort
- RCGP: Royal College of General Practitioners
- RSC: Research and Surveillance Centre
- SSGPs: Syndromic Surveillance General Practices
- UKHSA: UK Health Security Agency

Please cite as:


A Profile of Influenza Vaccine Coverage for 2019-2020: Database Study of the English Primary Care Sentinel Cohort

JMIR Public Health Surveill 2024;10:e39297
URL: https://publichealth.jmir.org/2024/1/e39297
doi:10.2196/39297
PMID:38787605
Rapid Surveillance Report

Trend of Mortality Due to Congenital Anomalies in Children Younger Than 5 Years in Eastern China, 2012-2021: Surveillance Data Analysis

Wen-Hong Dong¹,², PhD; Jun-Xia Guo¹,², MSc; Lei Wang¹,², PhD; Shuang-Shuang Zheng¹,², MSc; Bing-Quan Zhu¹,², MD, PhD; Jie Shao¹,², MD

¹Department of Child Health Care, Children’s Hospital, Zhejiang University School of Medicine, Hangzhou, China
²National Clinical Research Center for Child Health, Hangzhou, China

Corresponding Author:
Bing-Quan Zhu, MD, PhD
Department of Child Health Care
Children’s Hospital
Zhejiang University School of Medicine
No. 3333 Binseng Road
Binjiang District
Hangzhou, 310052
China
Phone: 86 571 88873597
Email: ispring2003@163.com

Abstract

Background: As one of the leading causes of child mortality, deaths due to congenital anomalies (CAs) have been a prominent obstacle to meet Sustainable Development Goal 3.2.

Objective: We conducted this study to understand the death burden and trend of under-5 CA mortality (CAMR) in Zhejiang, one of the provinces with the best medical services and public health foundations in Eastern China.

Methods: We used data retrieved from the under-5 mortality surveillance system in Zhejiang from 2012 to 2021. CAMR by sex, residence, and age group for each year was calculated and standardized according to 2020 National Population Census sex- and residence-specific live birth data in China. Poisson regression models were used to estimate the annual average change rate (AACR) of CAMR and to obtain the rate ratio between subgroups after adjusting for sex, residence, and age group when appropriate.

Results: From 2012 to 2021, a total of 1753 children died from CAs, and the standardized CAMR declined from 121.2 to 62.6 per 100,000 live births with an AACR of –9% (95% CI –10.7% to –7.2%; P<.001). The declining trend was also observed in female and male children, urban and rural children, and neonates and older infants, and the AACRs were –9.7%, –8.5%, –8.5%, –9.2%, –12%, and –6.3%, respectively (all P<.001). However, no significant reduction was observed in children aged 1-4 years (P=.22). Generally, the CAMR rate ratios for male versus female children, rural versus urban children, older infants versus neonates, and older children versus neonates were 1.18 (95% CI 1.08-1.30; P<.001), 1.20 (95% CI 1.08-1.32; P=.001), 0.66 (95% CI 0.59-0.73; P<.001), and 0.20 (95% CI 0.17-0.24; P<.001), respectively. Among all broad CA groups, circulatory system malformations, mainly deaths caused by congenital heart diseases, accounted for 49.4% (866/1753) of deaths and ranked first across all years, although it declined yearly with an AACR of –9.8% (P<.001). Deaths due to chromosomal abnormalities tended to grow in recent years, although the AACR was not significant (P=.90).

Conclusions: CAMR reduced annually, with cardiovascular malformations ranking first across all years in Zhejiang, China. Future research and practices should focus more on the prevention, early detection, long-term management of CAs and comprehensive support for families with children with CAs to improve their survival chances.

(JMIR Public Health Surveill 2024;10:e53860) doi:10.2196/53860

KEYWORDS
under-five years; congenital anomalies; mortality; death cause; rank

https://publichealth.jmir.org/2024/1/e53860 JMir Public Health Surveill 2024 | vol. 10 | e53860 | p.426 (page number not for citation purposes)
Introduction

As a diverse range of structural, functional, and metabolic malformations present at birth, congenital anomalies (CAs) pose a great burden to the families involved [1]. According to the World Health Organization’s latest estimate [2], more than 240,000 newborns die from CAs in the first 28 days of life every year worldwide, and another 170,000 newborns who survive then die from CAs in the following 5 years. Although the world has witnessed encouraging reductions in CA burdens with regard to both mortality and disability-adjusted life-years [1], mortality due to CAs still constitutes 9.4% of all deaths for children younger than 5 years and ranked fourth among all causes globally [3]. For regions such as southern sub-Saharan Africa, the under-5 mortality (U5MR) due to CAs showed no sign of decline; instead, it had an annual increase of 0.17% from 1990 to 2019 [4]. In China’s case, we have also observed an ongoing decline of both U5MR and CA-related mortality [5]. However, deaths caused by CAs remain one of the leading causes of child mortality [6], as observed in other countries [7].

Even if children with CAs survive from an early death, the long-term physical disability or functional disorder still leaves a great burden to both the family and society. It was estimated that the cost of birth defect–related hospitalizations in the United States was US $22.9 billion per year [8]. Although the total medical costs regarding all birth defects is not available in China, the median expenditure of surgery for congenital heart disease (CHD) in China was reported to be CN ¥67,867 (US $9839) in 2020 [9]. For newborns, the median cost was as high as CN ¥144,380 (US $20,931), which was 5.24 times of the median yearly per capita disposable income of the Chinese population in 2020 [9].

According to data issued by the National Bureau of Statistics of China, Zhejiang, a province located in Eastern China, ranked fourth among all 34 administrative regions regarding regional total gross domestic product (GDP) and eighth in GDP per capita in 2022. It has the best medical resources and public health services among the 28 provinces and autonomous regions in the country, excluding the other 6 administrative regions including Beijing, Shanghai, and Hong Kong due to the incomparable size of the targeted service population [10]. However, with the epidemiological transition of the death-cause pattern from infectious diseases to noncommunicable diseases across the whole society [11], little is known about the survival burden of CAs in children younger than 5 years in Zhejiang. Therefore, we conducted this study to understand the trend and burden of under-5 CA mortality (CAMR) in this more economically advanced province in China.

Methods

In this descriptive study, data from a long-existing mortality surveillance system in Zhejiang were used.

Mortality Surveillance

Mortality surveillance for children younger than 5 years in Zhejiang started in the early 1990s and is carried out by health workers from 3 levels (village, county, and city) of child health care institutes, mainly maternal and child health care hospitals and community health care centers. Out of the 90 administrative districts and counties, 30 were randomly chosen, and the vital status of children younger than 5 years is routinely monitored in these districts and counties according to the protocol designed by the National Maternal and Child Surveillance Office, as reported elsewhere [5].

Briefly, all deaths that occur in the first 5 years after birth are recorded in a structured, paper-based death report card (DRC) by a physician in a hospital or a community health care worker, depending on whether the death event involves medical treatments in a hospital. Information on the children’s demographics, obstetric status, hospitalization, and death is recorded in the DRC and submitted to the provincial-level surveillance office seasonally, together with a quarterly report including residence- and sex-specific number of live births and deaths in each village.

Quality Control

To ensure the accuracy of the children’s survival status and the information in the DRC, cross-checking with the local maternal delivery system, vaccination information system, household registration and cancellation system in the Bureau of Public Security, etc., is performed by village-, county-, and city-level health workers quarterly, semiannually, and semiannually, respectively. Provincial-level staff perform a field quality control investigation every year. To minimize information error, 2 experienced physicians were invited to review all DRCs in the data preparation process independently, and any disagreement were settled by a third senior physician.

CA Groups

Deaths due to CAs (ie, CAs as the underlying cause of death) were coded according to the International Classification of Diseases, 10th Revision (ICD-10) and recorded in the DRC. In this study, we combined them into 10 broad groups, which were nervous system malformations (Q90-Q97); congenital malformations of the eye, ear, face, and neck (Q10-Q18); circulatory system anomalies including CHD (Q20-Q28); respiratory system malformations (Q30-Q34); cleft lip and cleft palate (Q35-Q37); other digestive system malformations (Q38-Q45); urogenital malformations (Q50-Q64); musculoskeletal malformations (Q65-Q79); chromosomal abnormalities including Down syndrome (Q90-Q99); and other congenital anomalies (Q80-Q89).

Statistical Analysis

Three age groups were generated during analysis, that is, neonates (aged 0-27 days), older infants (aged 28-364 days), and older children (aged 1-4 years). In the analysis, year-, age-, sex-, and residence-specific CAMRs were calculated as the number of CA deaths in the corresponding year, age group, sex (male or female), and residential region (urban or rural) divided by the number of live births in that specific group. Urban and rural regions were defined according to the fifth digit of the 12-digit administrative code issued by National Bureau of Statistics, for which 1 represents an urban region and all other digits represent a rural region. All mortality data were then standardized by using sex- and residence-specific live birth data.

https://publichealth.jmir.org/2024/1/e53860

JMIR Public Health Surveil 2024 | vol. 10 | e53860 | p.427
(page number not for citation purposes)
from 2020 National Population Census of China. The average annual change rates (AACRs) of CAMR and the rate ratios (RRs) of CAMR for male versus female children, rural versus urban children, older infants versus neonates, and older children versus neonates were analyzed using the Poisson regression model after adjusting for age, sex, and residence when appropriate. All analyses were conducted in SAS (version 9.2; SAS Institute), and a 2-sided \( P < .05 \) was considered significant.

**Ethical Considerations**

The Medical Ethics Committee of Children’s Hospital, Zhejiang University School of Medicine approved this study (2023-IRB-0248-P-01). Since all mortality data were deidentified from the government-established death surveillance system, informed consent was waived by the Medical Ethics Committee.

**Results**

From 2012 to 2021, a total of 1753 deaths caused by CAs were reported and the proportion of CA deaths stabilized around 20% of all deaths (Multimedia Appendix 1). The standardized CAMR in Zhejiang dropped from 121.2 per 100,000 live births in 2012 to 62.6 per 100,000 live births in 2021, representing a 48.3% (58.6/121.2) decline in total with a 9% annual reduction on average (\( P < .001 \); Table 1). Compared with female children, the CAMR of male children was generally higher in each year of the last decade, and the adjusted RR in total was 1.18 (95% CI 1.08-1.30; \( P < .001 \)). Generally, the AACRs for female and male children were –9.7% (95% CI –12.3% to –7.0%; \( P < .001 \)) and –8.5% (95% CI –11.1% to –6.1%; \( P < .001 \)), respectively.

Table 2 shows that relative to children from urban areas, rural children had a 20% higher CAMR (RR 1.20, 95% CI 1.08-1.32; \( P = .001 \)). Moreover, rural children had a higher AACR than their urban counterparts and the AACRs for urban and rural children were –8.5% (95% CI –11.6% to –5.4%; \( P < .001 \)) and –9.2% (95% CI –11.3% to –7.1%; \( P < .001 \)), respectively.

CAMR declined annually by 12% (95% CI 9.6%-14.5%; \( P < .001 \)) for neonates and by 6.3% (95% CI 3.4%-9.2%; \( P < .001 \)) for older infants, but not in older children (\( P = .22 \); Table 3). Generally, neonates had the highest standardized CAMR compared to older infants and older children (47.2 versus 31.0 and 9.5 per 100,000 live births; Table 3). However, the CAMR gaps across age groups narrowed annually. For older infants, the RR versus neonates turned insignificant since 2018 (all \( P > .05 \)) and increased from 0.48 (95% CI 0.37-0.63; \( P < .001 \)) in 2012 to 1.03 (95% CI 0.65-1.62; \( P = .91 \)) in 2021. For older children, the RR versus neonates grew from 0.17 (95% CI 0.11-0.25) in 2012 to 0.35 (95% CI 0.19-0.66) in 2021 (\( P < .001 \) in all years).

### Table 1. Standardized under-5 mortality caused by congenital anomalies in Zhejiang from 2012 to 2021 by sex.

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Deaths, n</th>
<th>Mortality Deaths, n</th>
<th>Sex</th>
<th>Female Deaths, n</th>
<th>Mortality</th>
<th>Male Deaths, n</th>
<th>Mortality</th>
<th>RR (95% CI)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1753</td>
<td>87.7</td>
<td></td>
<td>757</td>
<td>79.9</td>
<td>992</td>
<td>94.7</td>
<td>1.18 (1.08-1.30)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2012</td>
<td>272</td>
<td>121.2</td>
<td></td>
<td>114</td>
<td>108.5</td>
<td>158</td>
<td>132.7</td>
<td>1.22 (0.96-1.56)</td>
<td>.10</td>
</tr>
<tr>
<td>2013</td>
<td>258</td>
<td>119.3</td>
<td></td>
<td>121</td>
<td>118.3</td>
<td>137</td>
<td>120.1</td>
<td>1.02 (0.79-1.30)</td>
<td>.91</td>
</tr>
<tr>
<td>2014</td>
<td>198</td>
<td>93.8</td>
<td></td>
<td>92</td>
<td>91.6</td>
<td>106</td>
<td>95.8</td>
<td>1.05 (0.79-1.39)</td>
<td>.74</td>
</tr>
<tr>
<td>2015</td>
<td>220</td>
<td>104.8</td>
<td></td>
<td>78</td>
<td>79.1</td>
<td>141</td>
<td>128.0</td>
<td>1.64 (1.25-2.17)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2016</td>
<td>184</td>
<td>87.2</td>
<td></td>
<td>86</td>
<td>86.3</td>
<td>97</td>
<td>88.0</td>
<td>1.02 (0.76-1.36)</td>
<td>.91</td>
</tr>
<tr>
<td>2017</td>
<td>180</td>
<td>75.3</td>
<td></td>
<td>83</td>
<td>72.8</td>
<td>96</td>
<td>77.6</td>
<td>1.06 (0.79-1.42)</td>
<td>.71</td>
</tr>
<tr>
<td>2018</td>
<td>145</td>
<td>72.1</td>
<td></td>
<td>58</td>
<td>60.3</td>
<td>87</td>
<td>82.6</td>
<td>1.36 (0.98-1.90)</td>
<td>.07</td>
</tr>
<tr>
<td>2019</td>
<td>124</td>
<td>66.5</td>
<td></td>
<td>54</td>
<td>61.5</td>
<td>69</td>
<td>71.1</td>
<td>1.17 (0.82-1.67)</td>
<td>.39</td>
</tr>
<tr>
<td>2020</td>
<td>84</td>
<td>52.7</td>
<td></td>
<td>37</td>
<td>48.8</td>
<td>47</td>
<td>56.3</td>
<td>1.15 (0.75-1.77)</td>
<td>.52</td>
</tr>
<tr>
<td>2021</td>
<td>88</td>
<td>62.6</td>
<td></td>
<td>34</td>
<td>50.6</td>
<td>54</td>
<td>73.3</td>
<td>1.45 (0.94-2.22)</td>
<td>.09</td>
</tr>
</tbody>
</table>

- \( a \) The annual average change rate (AACR) of under-five mortality caused by congenital anomalies in total from 2012 to 2021 was –9.0% (95% CI –10.7% to –7.2%; \( P < .001 \)).
- \( b \) Four children with unknown sex between 2012 and 2021 were not included in the analysis.
- \( c \) RR: rate ratio.
- \( d \) RR of under-5 mortality due to congenital anomalies for male versus female children after adjusting for age group and residence.
- \( e \) The unit of standardized mortality was per 100,000 live births.
- \( f \) The AACR of under-five mortality caused by congenital anomalies in female children from 2012 to 2021 was –9.7% (95% CI –12.3% to –7.0%; \( P < .001 \)).
- \( g \) The AACR of under-five mortality caused by congenital anomalies in male children from 2012 to 2021 was –8.5% (95% CI –11.1% to –6.1%; \( P < .001 \)).
### Table 2. Standardized under-5 mortality caused by congenital anomalies in Zhejiang from 2012 to 2021 by residence.

<table>
<thead>
<tr>
<th>Year</th>
<th>Residence</th>
<th></th>
<th></th>
<th>RR&lt;sup&gt;b&lt;/sup&gt; (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Deaths, n</td>
<td>Mortality&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Rural&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Deaths, n</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>537</td>
<td>78.3</td>
<td>1216</td>
<td>93.3</td>
<td>1.20 (1.08-1.32)</td>
</tr>
<tr>
<td>2012</td>
<td>81</td>
<td>101.8</td>
<td>191</td>
<td>132.9</td>
<td>1.30 (1.01-1.69)</td>
</tr>
<tr>
<td>2013</td>
<td>75</td>
<td>98.9</td>
<td>183</td>
<td>131.5</td>
<td>1.33 (1.02-1.74)</td>
</tr>
<tr>
<td>2014</td>
<td>73</td>
<td>100.7</td>
<td>125</td>
<td>89.6</td>
<td>0.89 (0.67-1.19)</td>
</tr>
<tr>
<td>2015</td>
<td>72</td>
<td>103.0</td>
<td>148</td>
<td>105.9</td>
<td>1.03 (0.77-1.36)</td>
</tr>
<tr>
<td>2016</td>
<td>48</td>
<td>68.1</td>
<td>136</td>
<td>98.6</td>
<td>1.45 (1.04-2.01)</td>
</tr>
<tr>
<td>2017</td>
<td>54</td>
<td>66.0</td>
<td>126</td>
<td>80.9</td>
<td>1.23 (0.89-1.69)</td>
</tr>
<tr>
<td>2018</td>
<td>37</td>
<td>54.8</td>
<td>108</td>
<td>82.4</td>
<td>1.50 (1.04-2.19)</td>
</tr>
<tr>
<td>2019</td>
<td>36</td>
<td>57.1</td>
<td>88</td>
<td>72.2</td>
<td>1.26 (0.86-1.86)</td>
</tr>
<tr>
<td>2020</td>
<td>30</td>
<td>54.8</td>
<td>54</td>
<td>51.5</td>
<td>0.94 (0.60-1.47)</td>
</tr>
<tr>
<td>2021</td>
<td>31</td>
<td>60.7</td>
<td>57</td>
<td>63.7</td>
<td>1.05 (0.68-1.62)</td>
</tr>
</tbody>
</table>

<sup>a</sup>RR: rate ratio.

<sup>b</sup>RR of under-5 mortality due to congenital anomalies for rural versus urban children after adjusting for sex and age group.

<sup>c</sup>The annual average change rate (AACR) of under-five mortality caused by congenital anomalies in urban children from 2012 to 2021 was –8.5% (95% CI –11.6% to –5.4%; P<.001).

<sup>d</sup>The AACR of under-five mortality caused by congenital anomalies in rural children from 2012 to 2021 was –9.2% (95% CI –11.3% to –7.1%; P<.001).

<sup>e</sup>The unit of standardized mortality was per 100,000 live births.

### Table 3. Standardized under-5 mortality caused by congenital anomalies in Zhejiang from 2012 to 2021 by age group.

<table>
<thead>
<tr>
<th>Year</th>
<th>Age group</th>
<th></th>
<th></th>
<th>RR&lt;sup&gt;d&lt;/sup&gt; (95% CI)</th>
<th>P value</th>
<th>RR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neonates&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Deaths, n</td>
<td>Mortality&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Older infants&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Deaths, n</td>
<td>Mortality</td>
<td>Older children&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Total</td>
<td>944</td>
<td>47.2</td>
<td>618</td>
<td>31.0</td>
<td>9.5</td>
<td>0.66 (0.59-0.73)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2012</td>
<td>165</td>
<td>73.9</td>
<td>79</td>
<td>35.0</td>
<td>28</td>
<td>0.48 (0.37-0.63)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2013</td>
<td>139</td>
<td>64.4</td>
<td>95</td>
<td>43.8</td>
<td>24</td>
<td>0.68 (0.53-0.89)</td>
<td>.004</td>
</tr>
<tr>
<td>2014</td>
<td>117</td>
<td>55.6</td>
<td>65</td>
<td>30.7</td>
<td>16</td>
<td>0.56 (0.41-0.75)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2015</td>
<td>118</td>
<td>56.4</td>
<td>78</td>
<td>37.1</td>
<td>24</td>
<td>0.67 (0.50-0.89)</td>
<td>.006</td>
</tr>
<tr>
<td>2016</td>
<td>99</td>
<td>46.7</td>
<td>67</td>
<td>31.8</td>
<td>18</td>
<td>0.68 (0.50-0.93)</td>
<td>.02</td>
</tr>
<tr>
<td>2017</td>
<td>102</td>
<td>42.4</td>
<td>63</td>
<td>26.8</td>
<td>15</td>
<td>0.62 (0.46-0.85)</td>
<td>.003</td>
</tr>
<tr>
<td>2018</td>
<td>69</td>
<td>34.4</td>
<td>54</td>
<td>27.0</td>
<td>22</td>
<td>0.78 (0.55-1.12)</td>
<td>.18</td>
</tr>
<tr>
<td>2019</td>
<td>54</td>
<td>28.8</td>
<td>50</td>
<td>27.0</td>
<td>20</td>
<td>0.94 (0.64-1.39)</td>
<td>.77</td>
</tr>
<tr>
<td>2020</td>
<td>44</td>
<td>27.3</td>
<td>29</td>
<td>18.5</td>
<td>11</td>
<td>0.66 (0.41-1.05)</td>
<td>.08</td>
</tr>
<tr>
<td>2021</td>
<td>37</td>
<td>26.3</td>
<td>38</td>
<td>27.1</td>
<td>13</td>
<td>0.92</td>
<td>1.03 (0.65-1.62)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The annual average change rate (AACR) of under-five mortality caused by congenital anomalies in neonates from 2012 to 2021 was –12% (95% CI –14.5% to –9.6%; P<.001).

<sup>b</sup>The AACR of under-five mortality caused by congenital anomalies in older infants from 2012 to 2021 was –6.3% (95% CI –9.2% to –3.4%; P<.001).

<sup>c</sup>The AACR of under-five mortality caused by congenital anomalies in older children from 2012 to 2021 was –3.2% (95% CI –8.3% to 2.0%; P=.22).

<sup>d</sup>RR: rate ratio.

<sup>e</sup>RR of under-5 mortality due to congenital anomalies adjusted for sex and residence.

<sup>f</sup>The unit of standardized mortality was per 100,000 live births.

Although a 9.8% (95% CI 7.3%-12.3%; P<.001) annual reduction was witnessed in circulatory system anomalies and more than half of the decline (33.1/58.6 per 100,000 live births, 56.5%) in the CAMR was contributed by the reduction in CHD
mortality, it still ranked first across all years in the last decade. Apart from CHD, the CAMRs reduced annually for all digestive ($P<.001$), other ($P=.004$), neural ($P=.004$), and respiratory ($P=.001$) anomalies, whereas CAMRs for the other 4 CA groups remained stable (all $P>.05$). Among the 10 CA groups, the rank of causes of death attributed to chromosomal abnormalities changed the most: from the eighth in 2012 to the third in 2021, with a 0.4% (95% CI: −6.7% to 7.6%; $P=.90$) insignificant annual increase. Details of the mortality for the 10 CA groups and a few common CA subtypes every 3 years from 2012 to 2021 are depicted in Figure 1.

Figure 1. Ranking of standardized under-5 mortality caused by 10 broad congenital anomaly groups and certain common subtypes in Zhejiang from 2012 to 2021. All data in this figure represent standardized under-5 mortalities due to certain congenital anomaly groups or subtypes (unit: per 100,000 live births). Different shades of red and blue indicating the rank of mortality caused by the broad category of congenital malformations: the darker the red is, the higher the rank; and the lighter the blue is, the lower the rank. AACR: annual average change rate; CHD: congenital heart disease; NTD: neural tube defect.

<table>
<thead>
<tr>
<th>Congenital anomalies</th>
<th>Rank</th>
<th>Total 2012</th>
<th>Total 2015</th>
<th>Total 2018</th>
<th>Total 2021</th>
<th>AACR (%)</th>
<th>95% CI</th>
<th>$P_{\text{trend value}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td>87.7</td>
<td>121.2</td>
<td>104.8</td>
<td>72.1</td>
<td>62.6</td>
<td>−9.0</td>
<td>(−10.7 to −7.2)</td>
</tr>
<tr>
<td>Circulatory system malformations</td>
<td>1</td>
<td>43.3</td>
<td>60.8</td>
<td>48.3</td>
<td>37.0</td>
<td>29.1</td>
<td>−9.8</td>
<td>(−12.3 to −7.3)</td>
</tr>
<tr>
<td>CHD</td>
<td></td>
<td>42.6</td>
<td>60.8</td>
<td>48.3</td>
<td>37.0</td>
<td>27.7</td>
<td>−10.4</td>
<td>(−12.9 to −7.9)</td>
</tr>
<tr>
<td>Other digestive system malformations</td>
<td>2</td>
<td>14.3</td>
<td>12.9</td>
<td>23.0</td>
<td>9.3</td>
<td>5.7</td>
<td>−8.5</td>
<td>(−12.8 to −4.1)</td>
</tr>
<tr>
<td>Esophageal atresia</td>
<td></td>
<td>2.3</td>
<td>1.8</td>
<td>3.7</td>
<td>1.9</td>
<td>1.4</td>
<td>−3.5</td>
<td>(−14.0 to 7.0)</td>
</tr>
<tr>
<td>Biliary atresia</td>
<td></td>
<td>4.0</td>
<td>2.7</td>
<td>7.2</td>
<td>2.9</td>
<td>2.2</td>
<td>−5.7</td>
<td>(−13.6 to 2.3)</td>
</tr>
<tr>
<td>Anal atresia</td>
<td></td>
<td>2.5</td>
<td>1.8</td>
<td>4.4</td>
<td>0.5</td>
<td>0.0</td>
<td>−18.1</td>
<td>(−29.2 to −7.0)</td>
</tr>
<tr>
<td>Other anomalies</td>
<td>3</td>
<td>6.7</td>
<td>9.4</td>
<td>8.1</td>
<td>5.5</td>
<td>4.3</td>
<td>−9.3</td>
<td>(−15.6 to −3.0)</td>
</tr>
<tr>
<td>Multiple system–involved malformations</td>
<td></td>
<td>4.1</td>
<td>5.0</td>
<td>4.3</td>
<td>2.9</td>
<td>2.1</td>
<td>−5.4</td>
<td>(−13.4 to 2.5)</td>
</tr>
<tr>
<td>Musculoskeletal malformations</td>
<td>4</td>
<td>5.3</td>
<td>8.5</td>
<td>4.9</td>
<td>3.9</td>
<td>7.8</td>
<td>−5.6</td>
<td>(−12.6 to 1.4)</td>
</tr>
<tr>
<td>Nervous system malformations</td>
<td>5</td>
<td>5.1</td>
<td>9.3</td>
<td>6.2</td>
<td>4.5</td>
<td>3.5</td>
<td>−10.5</td>
<td>(−17.8 to −3.3)</td>
</tr>
<tr>
<td>NTDs</td>
<td></td>
<td>1.6</td>
<td>3.2</td>
<td>2.8</td>
<td>0.5</td>
<td>0.7</td>
<td>−17.8</td>
<td>(−31.5 to −4.1)</td>
</tr>
<tr>
<td>Congenital hydrocephalus</td>
<td></td>
<td>2.1</td>
<td>4.4</td>
<td>2.0</td>
<td>2.5</td>
<td>2.1</td>
<td>−7.2</td>
<td>(−18.5 to 4.0)</td>
</tr>
<tr>
<td>Chromosomal abnormalities</td>
<td>6</td>
<td>4.9</td>
<td>3.1</td>
<td>5.1</td>
<td>7.3</td>
<td>6.4</td>
<td>0.4</td>
<td>(−6.7 to 7.6)</td>
</tr>
<tr>
<td>Down syndrome</td>
<td></td>
<td>2.6</td>
<td>2.6</td>
<td>3.2</td>
<td>4.4</td>
<td>2.1</td>
<td>−6.7</td>
<td>(−16.7 to 3.3)</td>
</tr>
<tr>
<td>Respiratory system malformations</td>
<td>7</td>
<td>4.1</td>
<td>8.1</td>
<td>7.4</td>
<td>2.9</td>
<td>2.8</td>
<td>−14.2</td>
<td>(−22.6 to −5.9)</td>
</tr>
<tr>
<td>Cleft lip and cleft palate</td>
<td>8</td>
<td>2.7</td>
<td>6.7</td>
<td>1.4</td>
<td>1.1</td>
<td>0.7</td>
<td>−14.9</td>
<td>(−25.2 to −4.6)</td>
</tr>
<tr>
<td>Urogenital malformations</td>
<td>9</td>
<td>0.7</td>
<td>1.4</td>
<td>0.0</td>
<td>0.0</td>
<td>1.4</td>
<td>1.7</td>
<td>(−17.9 to 21.2)</td>
</tr>
<tr>
<td>Congenital malformations of the eye, ear, face, and neck</td>
<td>10</td>
<td>0.5</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>−7.7</td>
<td>(−30.6 to 15.2)</td>
</tr>
</tbody>
</table>

Multimedia Appendix 2 shows that 944 (53.9%) and 1562 (89.1%) out of 1753 deaths were caused by CAs occurred during the neonatal and infantile period, respectively. Among all common CAs, some were more fatal, with more than 70% of deaths occurring in the first 7 days of life, such as neural tube defects (NTDs; 25/32, 78%) and multiple systems–involved malformations (60/81, 74%). Some were less fatal, with better survival chances over 1 year, such as Down syndrome (17/52, 33%) and biliary atresia (23/81, 28%).
Discussion

Principal Findings

In this retrospective study, we observed an annual 9% reduction of CAMR in Zhejiang, and the change was significant in male and female children, urban and rural children, and neonatal and older infants. Generally, CAMRs were higher in male versus female children, rural versus urban children, and neonatal versus older infants. Among all CAs, some are more fatal, with ≥70% of deaths occurring in the first 7 days of life, such as NTDs (78%) and multiple systems-involved malformations (74%), whereas some are less fatal, with more survival chances over 1 year, such as Down syndrome (33%) and biliary atresia (28%). Circulatory system malformations, mainly CHDs, caused the maximum deaths in children younger than 5 years. U5MR caused by chromosomal abnormalities, although insignificant, tended to grow in recent years.

Comparison With Prior Work

The CAMR we observed in Zhejiang was not only lower than most provinces in China [6] but also lower than that in high-income countries such as the United States [12] and European countries [7]. The rather low CAMR and an ongoing declining trend can be explained by several reasons. First, there has been a series of nationally covered maternal and children health programs with stable financial investment from the Chinese government [13]. Among them, some aimed at reducing the prevalence of certain CAs and adverse birth outcomes, such as providing free folic acid tablets to women of reproductive age [14], whereas others aimed at improving the survival and health of children, such as providing antenatal and postnatal health care and health management for children aged 0-6 years [13].

Second, the rapid development and universal application of prenatal and newborn screening is another potential reason for the continuous decline of CAMR in Zhejiang. In 2000, only 7.6% children with Down syndrome were identified prenatally in China [15]. From 2013 to 2017 in Zhejiang, 91%, 96.9% and 100% of Down syndrome, encephalocele, and anencephaly were identified through prenatal screening, respectively [16]. Along with the extensive application of prenatal screening was a ≥90% rate of early termination of pregnancy, leading to less live births with severe CAs [17]. Newborn screening, a well-established public health measure to reduce early infant deaths, is widely used in Zhejiang, and more than 400,000 newborns are screened every year [18].

Third, with the development of postpartum surgical and repair techniques, the survival chances of children with CAs such as CHD, cleft lip, and cleft palate improved greatly [19,20]. For example, although the birth prevalence of CHD increased continuously worldwide [21] as well as in China [22], mortality due to CHD decreased gradually [23]. According to a hospital-based study, in-hospital mortality for children who underwent CHD surgeries during the first month after delivery reduced from 16.4% in the period from 2004 to 2008 to 5% in the period from 2014 to 2018 [24].

Similar to previous studies [6,25], we identified a larger CA death burden in male children than in female children. Both biological and genetic factors may play a role. First, male children are more likely to be born prematurely, which signifies lower survival chances [26]. Second, the X chromosome carries more genes responsible for immune function, leaving male children with CAs more vulnerable to infections [27]. The mortality differences across age groups are mainly related to the type and severity of CAs. As mentioned above, most severe CAs such as NTDs would experience pregnancy termination, and less than half of live births with NTDs would survive over 1 year [28]. For CAs that are less fatal such as Down syndrome, the life expectancy has increased dramatically and reached 60 years in 2002 [29].

The high death burden caused by CHD in our study was prevalent globally [4]. Although great progress has been made, the elimination of CHD remains a great challenge. On the one hand, as a public health issue driven mainly by genetics, it is difficult to remove the genetic predisposition and other environmental contributors that once existed [30]. On the other hand, timely diagnosis and treatment is still an issue to be solved. In Western countries such as Norway, only 13% of infants were discharged without the identification of severe CHDs [31], whereas in China, this number was 52.5% and can be as high as 71% for neonates with asymptomatic critical CHD [32]. In addition, mortality from complex CHD is still higher than other CHDs [33]. Besides, the high medical cost and parental worries for children’s future quality of life might also affect parents’ decision-making on whether to take further treatment or not.

In our study, we also noticed an increase in U5MR for chromosomal issues. Although insignificant, several reasons may provide an explanation. First, the proportion of pregnancies with advanced maternal age, a well-established risk factor of chromosomal abnormalities in the offspring [34], almost doubled after the termination of the One-Child Policy [35]. Second, new genetic techniques promote the detection and correct categorization of chromosomal issues [36]. Third, although detectable, systematic treatment strategies for most abnormalities are still absent [37]. Considering the factors mentioned above, higher prevalence and mortality of chromosomal abnormalities are projected in the future, and more attention should be paid to this issue.

Limitations

There are mainly 3 limitations in our study. First, the miscoding of death cause might exist. However, we have invited 2 experienced physicians to independently review all DRCs in the data preparation process to minimize this type of error in addition to routine quality control measures. Second, detailed information regarding each live birth with CAs and their survival status in the next 5 years was not collected in our work, limiting further analysis and comparisons on survival chances for different CA subtypes and the associated factors. Third, the analysis for CAs as a contributory cause of study was necessary but not feasible due to the lack of information in this study. It was reported that only 70% of deaths in infants with CAs...
recorded a CA as the underlying cause of death [38], meaning that the actual burden of CAs is greater than what we found.

**Conclusion**

In our study, we observed an ongoing decline in the CAMR in Eastern China, with cardiovascular malformations ranking first across all years. Despite the remarkable achievements, we still face great challenges, and we hereby make the following suggestions: (1) researchers should perform more well-designed studies to understand CAs’ risk factors and take measure accordingly to lower CA incidence, similar to what have been achieved with NTDs; (2) we must strengthen in-service education and training of medical staffs to improve their skills and promote “early detection, early diagnosis and early treatment”; (3) more sensitive and population-wide screening techniques such as echocardiography for early identification of CHD should be developed and applied to the public for free to lower undiagnosed CA cases [39]; and (4) we should provide easily accessible educational, emotional, and financial support to families with children with CAs, as reported by Wray and colleagues [40], to increase families’ capabilities and confidence in caring for children with CAs and improve their survival possibilities.

**Acknowledgments**

This study was funded by National Nature and Science Foundation (81773440). We thank all physicians and health care workers from village-, county-, and city-level hospitals and institutions for their hard work across the years.

**Data Availability**

The data sets generated and/or analyzed during this study are not publicly available due to safety and privacy concerns but are available from the corresponding author on reasonable request.

**Authors’ Contributions**

WHD and BQZ contributed to study concept, study design, and the drafting of the manuscript. WHD, JXG, and SSZ contributed to the acquisition of data. WHD, JXG, and LW contributed to statistical analysis and the interpretation of data. BQZ and JS contributed to critical revision of the manuscript for important intellectual content, funding acquisition, and study supervision. All authors contributed to the paper and approved the submitted version.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Standardized under-5 mortality and percentages caused by congenital anomalies in Zhejiang from 2012 to 2021.

[PNG File, 185 KB - publichealth_v10i1e53860_app1.png]

**Multimedia Appendix 2**

Age distribution of deaths due to 10 broad congenital anomaly groups and certain common subtypes for children younger than 5 years in Zhejiang from 2012 to 2021.

[PNG File, 245 KB - publichealth_v10i1e53860_app2.png]

**References**


Abbreviations

AACR: average annual change rate
CA: congenital anomaly
CAMR: under-5 congenital anomaly mortality
CHD: congenital heart disease
DRC: death report card
GDP: gross domestic product
ICD-10: International Classification of Diseases, 10th Revision
NTD: neural tube defect
RR: rate ratio
U5MR: under-5 mortality
Original Paper

Updated Surveillance Metrics and History of the COVID-19 Pandemic (2020-2023) in the Middle East and North Africa: Longitudinal Trend Analysis

Alan G Soetikno, BS; Alexander L Lundberg, MSc, PhD; Egon A Ozer, MD, PhD; Scott A Wu, BA; Sarah B Welch, MPH; Maryann Mason, PhD; Yingxuan Liu, MA; Robert J Havey, MD; Robert L Murphy, MD; Claudia Hawkins, MD, MPH; Charles B Moss, PhD; Lori Ann Post, PhD

Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
Buehler Center for Health Policy and Economics, Robert J. Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States
Department of Emergency Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
Department of Medicine, Division of Infectious Diseases, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
Center for Pathogen Genomics and Microbial Evolution, Robert J. Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States
Robert J. Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States
Department of Medicine, General Internal Medicine and Geriatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
Center for Global Communicable and Emerging Infectious Diseases, Robert J. Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States
Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States

Corresponding Author:
Lori Ann Post, PhD
Buehler Center for Health Policy and Economics
Robert J. Havey, MD Institute for Global Health
Northwestern University
420 E. Superior
Chicago, IL, 60611
United States
Phone: 1 3125031706
Email: lori.post@northwestern.edu

Abstract

Background: This study updates the COVID-19 pandemic surveillance in the Middle East and North Africa (MENA) we first conducted in 2020 with 2 additional years of data for the region.

Objective: The objective of this study is to determine whether the MENA region meets the criteria for moving from a pandemic to endemic. In doing so, this study considers pandemic trends, dynamic and genomic surveillance methods, and region-specific historical context for the pandemic. These considerations continue through the World Health Organization (WHO) declaration of the end of the public health emergency for the COVID-19 pandemic on May 5, 2023.

Methods: In addition to updates to traditional surveillance data and dynamic panel estimates from the original study by Post et al, this study used data on sequenced SARS-CoV-2 variants from the Global Initiative on Sharing All Influenza Data (GISAIID) to identify the appearance and duration of variants of concern. We used Nextclade nomenclature to collect clade designations from sequences and Pangolin nomenclature for lineage designations of SARS-CoV-2. Finally, we conducted a 1-sided t test to determine whether regional weekly speed of COVID-19 spread was greater than an outbreak threshold of 10. We ran the test iteratively with 6 months of data from September 4, 2020, to May 12, 2023.

Results: The speed of COVID-19 spread for the region had remained below the outbreak threshold for 7 continuous months by the time of the WHO declaration. Acceleration and jerk were also low and stable. Although the 1- and 7-day persistence coefficients remained statistically significant and positive, the weekly shift parameters suggested the coefficients had most recently turned negative, meaning the clustering effect of new COVID-19 cases became even smaller in the 2 weeks around the WHO declaration.
From December 2021 onward, Omicron was the predominant variant of concern in sequenced viral samples. The rolling t test of the speed of spread equal to 10 became entirely insignificant from October 2022 onward.

Conclusions: The COVID-19 pandemic had far-reaching effects on MENA, impacting health care systems, economies, and social well-being. Although COVID-19 continues to circulate in the MENA region, the rate of transmission remained well below the threshold of an outbreak for over 1 year ahead of the WHO declaration. COVID-19 is endemic in the region and no longer reaches the threshold of the pandemic definition. Both standard and enhanced surveillance metrics confirm that the pandemic had transitioned to endemic by the time of the WHO declaration.

Introduction

COVID-19, the disease caused by the virus SARS-CoV-2, was first detected in Wuhan, China, in the fall of 2019 [1-6]. The Middle East and North Africa (MENA) reported their first 9 COVID-19 cases in the United Arab Emirates and Egypt between January 29, 2020, and February 16, 2020 [7]. Our research team first conducted an epidemiological assessment of the pandemic in MENA 1 year into the pandemic [8]. This study provides 2 additional years of updated surveillance and analysis for the region.

We adopt the World Bank's definition of MENA, which is based on economic development and geographical proximity, encompassing Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, the United Arab Emirates, Yemen, Algeria, Djibouti, Egypt, Libya, Morocco, and Tunisia [9].

The World Health Organization (WHO) and Director-General Ghebreyesus declared the end of COVID-19 as a public health emergency of international concern on May 5, 2023 [10-12] based on the recommendation of the COVID-19 Emergency Committee [12]. We compared how the pandemic, as experienced in the MENA region, progressed before and after the declaration.

Epidemiological terms, such as pandemic, epidemic, outbreak, and endemic, are used to describe the occurrence and spread of diseases [13,14]. The distinctions between these terms lie in their scope, geographic extent, and severity. An epidemic refers to a sudden increase in the number of disease cases in a specific population at the subnational level. If the epidemic spreads across several countries or continents, it becomes a pandemic. An outbreak, on the other hand, describes a sudden increase in a concentrated setting, usually involving a more limited geographic area than an epidemic. Endemic refers to the constant presence of a disease in a particular geographic region or population, with no sudden increases in case volume [15,16]. Field epidemiology defines these terms based on transmission metrics and geographical distribution. All public health surveillance data suffer from incomplete case ascertainment, meaning that not all cases of a disease may be reported or captured in the surveillance system. Despite this limitation, public health surveillance is valuable because it provides a near real-time view of disease trends within a population. Surveillance is crucial for leaders to respond effectively to health threats, implement control measures, and allocate resources where they are needed most. In sum, although the data may not capture every single case, they serve as a critical tool for monitoring and tracking the spread of diseases and health events in a population [17].

Since conducting and publishing our initial research on the pandemic in MENA, much more has been learned about the SARS-CoV-2 virus and its transmission and prevention [18-24]. Additionally, disease control measures and their effects on medical, social, and economic well-being have been further studied in the interval period, with new emphasis on clear guidelines and effective communication strategies [18,19]. In parallel, there was increased focus on risk factors for COVID-19 infection and mortality, ranging from personal risk factors such as age and medical comorbidities to air pollution, climate, and population density [18,25-27]. The vast breadth and depth of new knowledge related to COVID-19 and public health gained in the past 2 years underscore the need for new analysis with updated data and historical context.

Public health surveillance is the “ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning and evaluation of public health practice” [28]. Surveillance explains the burden of disease with transmission and death rates [29-43] and allows us to compare that burden between geographical regions and to understand which regions are most impacted. The impact can be measured through standardized population rates regarding how many people contract or die from a disease.

However, traditional surveillance carries several limitations that this study addressed. Traditional surveillance provides a snapshot of what has already happened [29-43], meaning surveillance is static. In the middle of a burgeoning pandemic, policymakers and public health practitioners also need to understand what is about to happen. Is an outbreak increasing? Will growth switch from linear to exponential? Are more people dying from that particular condition in one place than another? To inform health policy and practice, knowledge of what is about to happen is often more valuable than knowledge of what did happen. To that end, we developed enhanced surveillance metrics that reflect the dynamics of a pandemic and can inform...
imminent growth and, most importantly, where along the epidemiological outbreak curve a particular region is situated. We also included dynamic metrics about the speed of the pandemic spread at the national, regional, and global levels. We measured how acceleration of speed this week compared with last week, as well as how novel infections last week predicted new cases this week. We can think of the latter measure as the echoing forward of cases. These metrics were tested and validated in prior research [8,44-54]. For the purpose of this study, standard surveillance metrics explain what has already happened in MENA, while enhanced surveillance metrics speak to what is about to happen or where along an epidemiological curve a country may sit. We used both types of metrics to analyze the possible end to the pandemic.

This study had 3 objectives. First, we aimed to measure whether there was a pandemic expansion or contraction in the MENA region at the time the WHO declared the end of the COVID-19 pandemic as a public health emergency of international concern on May 5, 2023. At both the region and country levels, we used advanced surveillance and analytical techniques to describe the status of the pandemic in a 2-week window around the WHO declaration. From a public health perspective, we need to know whether the rate of new COVID-19 cases was increasing, decreasing, or stable from week to week and if any changes in the transmission rate indicated an acceleration or deceleration of the pandemic. Statistical insignificance is significant—it signals an epidemiological “end” to the pandemic if the rate of new cases is 0 (or very low) and stable, meaning the number of new cases is neither accelerating nor decelerating.

Second, we used dynamic and genomic surveillance methods to describe the history of the pandemic in the region and situate the time window around the WHO declaration. We included the ratio of COVID-19 deaths to the number of transmissions as a proxy for the population-level mortality risk from infection. We also included a historical record of genomic surveillance from sequenced viral specimens to identify the appearance and spread of variants of concern in the region.

Third, we aimed to provide historical context for the course of the pandemic in the MENA region. We addressed several questions. How did countries respond to the pandemic? How did the region fare in terms of disease burden? What social, economic, and political factors shaped the course of COVID-19 in the region? This context can provide important lessons for disease prevention and mitigation in future pandemics.

**Methods**

**Data Source**

This study conducted trend analyses with longitudinal COVID-19 data for the MENA region from 2020 through 2023. Data on COVID-19 transmission and death were sourced from Our World in Data (OWID) [55]. OWID compiles data on COVID-19 cases and mortality from multiple sources, including individual websites, statistical reports, and press releases. The MENA region was defined by the World Bank’s economic and geographical metrics [9], and the data comprised an unbalanced panel of 20 countries and territories from September 4, 2020, to May 12, 2023. Because several countries around the world switched from daily to weekly reports at various points in 2023, we used a cubic spline to interpolate daily new cases and deaths when any country had 4 consecutive periods of non-zero new cases interspersed by 6 days of zero new cases. A cubic spline is a statistical function used to assess the “smoothness” of data points and estimate missing or unclear data on a line given surrounding trends.

To identify the appearance and duration of variants of concern, we used data on sequenced SARS-CoV-2 variants from the Global Initiative on Sharing All Influenza Data (GISAID), which is an effective and trusted online resource for sharing genetic, clinical, and epidemiological COVID-19 data [56-60]. We used Nextclade nomenclature [61] to collect clade designations from sequences and Pangolin nomenclature for lineage designations of SARS-CoV-2 [62,63]. Nextclade nomenclature is an open-source tool for viral genome analysis, mutation identification, clade assignment, and phylogenetic mapping. Pangolin nomenclature is an open-source tool (Phylogenetic Assignment of Named Global Outbreak Lineages) used to track the transmission and spread of SARS-CoV-2 and its lineages. Metadata for the study period were collected on June 22, 2023. To avoid low frequency or potentially erroneous samples, the data set was further filtered to exclude months with fewer than 100 available samples, variant groups with fewer than 5 samples in a month, and variant groups representing less than 0.5% of the total samples in a month. The final data set consisted of 184,386 total samples available on GISAID [56-59].

**Measures**

This study provides updates of traditional surveillance data and dynamic panel estimates from the original study by Post et al [8,52,53,64-66]. The “speed” of spread of the pandemic is the rate of new COVID-19 cases per 100,000 population. Novel metrics go beyond speed to add acceleration, jerk, and 1- and 7-day persistence measures. Acceleration is the difference in speed from one week or day to the next. Acceleration identifies whether the number of new cases is increasing (positive acceleration), decreasing (negative acceleration), or at a stable inflection point (zero). “Jerk” is the change in acceleration from one time interval to the next, and its name is adopted from physics nomenclature. A positive jerk can indicate explosive growth in the spread of a disease. Finally, 1- and 7-day persistence measures capture the impact of the 1- and 7-day lag of speed on current speed. These measures capture the echo-forward effect of COVID-19 cases on future cases either 1 or 7 days later. They are derived from coefficient estimates on lagged transmission rates in an Arellano-Bond dynamic panel data model [67]. The model follows the general form of:

$$y_{it} = \rho y_{it-1} + \beta x_{it} + \alpha_i + u_{it}(1)$$

where the dependent variable is the rate of COVID-19 transmissions, the independent variables include weekend and recent week indicators, $\alpha_i$ denotes country fixed effects, and $u_{it}$ is the idiosyncratic error term. Please see the initial study for more details [8].

We further analyzed the potential “statistical end” to the pandemic with a 1-sided $t$ test for whether the mean of speed
of spread, defined as the rate of new COVID-19 cases per 100,000 population in a given time period, was equal to or greater than the outbreak threshold of 10. We ran the test on a rolling 6-month window over weekly speed for the region, and we plotted the P values from the test over time. All statistical analyses were conducted with the *plm* package (version 2.6-2) in R (version 4.2.1) [64,65].

**Ethical Considerations**

This study followed the guidelines of the World Medical Association’s Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects [68,69]. This research relied on publicly available data with no private, identifiable information. Thus, institutional review board review was unsolicited.

**Results**

Table 1 presents the dynamic panel estimates for the most recent time window. The Wald test for the regression was significant (P<.001), and the Sargan test failed to reject the validity of the overidentification restrictions (P=.99). Although the 1- and 7-day lag coefficients were statistically significant, suggesting a cluster effect in which cases on a given day impact cases 1 day and 7 days later, the coefficients were moderate in magnitude (0.310 and 0.586, respectively). Furthermore, the shift parameters for either of the 2 most recent weeks were both significant and negative, meaning the clustering effect had become smaller—in fact, negative—in the 2 weeks around May 5, 2023.

Static surveillance metrics for the weeks of April 28, 2023, and May 5, 2023, are provided in Table S1 of the Multimedia Appendix 1. Every country had a small number of new COVID-19 cases. The highest rate of new cases per 100,000 population was 8.90 in Qatar for the week of April 28, 2023, still considered a low transmission rate by the Centers for Disease Control and Prevention (CDC) [70]. This rate falls just below the informal outbreak threshold of 10 cases per week per 100,000 population [8,44-54]. Specifically, a “Low” transmission is considered no more than 10 cases per 100,000 people per week. “Moderate” transmission is 10 to 50 cases per 100,000 people per week, and “Substantial” transmission is 50 to 100 cases per 100,000 people per week [70,71].

The weekly transmission rate in Qatar also fell to 0 the following week. For the same week, no other country had a speed greater than 2.

Comparisons in Table S1 (Multimedia Appendix 1) demonstrate little to no change in surveillance metrics before and after the WHO declared an end to the COVID-19 emergency. Without question, Iran and Israel had the most cases of COVID-19 transmissions and deaths, but this rank is a function of population size. Thus, a better measure is the number of transmissions and deaths, but this rank is a function of population size. Thus, a better measure is the number of cases per 100,000 people. Iran reported 0.02 deaths per 100,000 population. When we controlled for a risk of death given the number of COVID-19 transmissions, we found that Egypt had the highest conditional death rate of 0.05 deaths per case. The next highest rate was 0.03 deaths per case in both Algeria and Tunisia.

Table 2 contains enhanced dynamic surveillance metrics for the 2 weeks before and after May 5, 2023. Speed of spread was low for every country except Qatar in the week of April 28, 2023, and acceleration was 0 or negative for almost every country. During this time, Qatar still remained well below the outbreak threshold (see Table S1 in Multimedia Appendix 1). Although positive, acceleration was small. The 7-day persistence effect on speed of spread was also very small in magnitude for the week of April 28, 2023, in every country, and the persistence effect fell to 0 or negative for every country the following week. These metrics suggest the pandemic may have indeed ended for the region. We note that the figures in Table 2 are not calculated as day-over-day averages across the week, as they are in Table S1 in Multimedia Appendix 1. Thus, the magnitudes of speed may not exactly match those in Table S1 in Multimedia Appendix 1.

Table 3 compares the 7-day persistence effect on speed for the 5 countries with the highest persistence for the weeks of April 28, 2023, and May 5, 2023. In each case, the effect had become either 0 or negative by the second week. Again, these metrics indicate that COVID-19 was well controlled in the region overall.

Figure 1 plots the regional speed of spread, acceleration, jerk, and 7-day persistence metrics from September 4, 2020, to May 12, 2023. The dashed grey line denotes the informal CDC outbreak threshold of speed equal to 10. The region experienced 3 outbreaks over the course of the pandemic. The first was brief, reaching a peak speed of only 13 in April 2021. The second saw a peak speed of 19 in August 2021. However, these 2 outbreaks can largely be considered 1, as speed dipped only slightly below the outbreak threshold of 10 between them. The third outbreak was the largest, with a peak speed of 33 in February 2022.

Figure 2 plots variant groups as a proportion of all viral specimens collected and sequenced in the region (and made available through GISAID) each month. The first 1 to 2 outbreaks referenced in the previous paragraph occurred just around the appearance of the Delta variant. The last outbreak was driven by the Omicron variant. MENA, like much of the rest of the world, saw a surge in cases amid the heightened transmissibility of Omicron [73]. Still, the outbreak was much smaller than in several other regions of the world, such as North America, Europe, and East Asia and the Pacific, which each saw peak speeds of over 200 amid Omicron outbreaks.

Another potential indication of the end to the pandemic was the continued dominance of the Omicron variant. Although the region saw a mixture of the ancestral, Alpha, Beta, and Delta variants prior to the arrival of Omicron in November 2021, viral sequences have almost exclusively returned as Omicron and its subvariants ever since.

Figure 3 plots the P values from a series of 1-sided t tests to determine whether speed for the region was equal to or greater than the threshold outbreak of 10. These tests were conducted...
on a rolling 6-month window of weekly regional speed. The dashed grey line denotes the least restrictive conventional significance level threshold of $\alpha$=.10. The only time the test rejected the null hypothesis in favor of the alternative occurred around the 6-month period ending in mid-August 2021. This period marked the extended, intermittent outbreak driven by (presumably) the Delta variant. From then on, the test failed to reject the null hypothesis, although $P$ values did drop again somewhat around the later Omicron outbreak. The test statistic became consistently insignificant from approximately October 2022 onward. This more recent lack of statistical significance is consistent with the end of the pandemic in the region, as the test clearly failed to reject the null hypothesis of outbreak threshold speed.

With the historical context of enhanced surveillance metrics, the MENA region appeared to be at the end stage of the pandemic. Speed had not been this low for this long since the start of the pandemic. Furthermore, speed remained well below outbreak status for over 1 year ahead of the WHO declaration. Figure 4 provides a timeline of the onset of COVID-19 in MENA as well as vaccination programs and major events that likely created additional challenges to disease control, such as the arrival of new variants of concern.

Table 1. Arellano-Bond dynamic panel data modeling from Equation (1) of the number of daily infections reported by country, April 28, 2023, through May 12, 2023.a,b

<table>
<thead>
<tr>
<th>Variable</th>
<th>Valuec</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-day persistence coefficient</td>
<td>0.310</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>7-day persistence coefficient</td>
<td>0.586</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Shift parameter week of April 28, 2023</td>
<td>-0.752</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Shift parameter week of May 5, 2023</td>
<td>-0.304</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Weekend</td>
<td>0.300</td>
<td>.17</td>
</tr>
</tbody>
</table>

aWald test: $\chi^2_6=12567.77, P<2.22^{-16}$.  
bSargan: $\chi^2_{540}=13, P= .99$.  
cContains estimates from the model in Equation (1).

Table 2. Novel surveillance metrics in Middle East and North Africa (MENA) for the weeks of April 28, 2023, and May 5, 2023.

<table>
<thead>
<tr>
<th>Country</th>
<th>April 28, 2023</th>
<th>May 5, 2023</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Speeda</td>
<td>Accelerationb</td>
</tr>
<tr>
<td>Algeria</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>Egypt</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Iran</td>
<td>0.22</td>
<td>0.06</td>
</tr>
<tr>
<td>Israel</td>
<td>2.39</td>
<td>0.05</td>
</tr>
<tr>
<td>Kuwait</td>
<td>0.20</td>
<td>0</td>
</tr>
<tr>
<td>Lebanon</td>
<td>1.17</td>
<td>0.27</td>
</tr>
<tr>
<td>Libya</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Morocco</td>
<td>0.13</td>
<td>0</td>
</tr>
<tr>
<td>Oman</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Qatar</td>
<td>16.52</td>
<td>1.27</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>0.50</td>
<td>-0.04</td>
</tr>
<tr>
<td>Tunisia</td>
<td>2.09</td>
<td>0.06</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>2.21</td>
<td>0.01</td>
</tr>
</tbody>
</table>

aNew COVID-19 cases per 100,000 population.  
bDifference in speed from one week to the next.  
cChange in acceleration from one week to the next.  
dThe impact of the 1- and 7-day lags of speed on current speed.
Table 3. Top 5 countries ranked by 7-day persistence in the Middle East and North Africa (MENA) for the weeks of April 28, 2023, and May 5, 2023.

<table>
<thead>
<tr>
<th>Week and country</th>
<th>7-day persistence&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Week of April 28, 2023</strong></td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td>0.19</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>0.15</td>
</tr>
<tr>
<td>Tunisia</td>
<td>0.12</td>
</tr>
<tr>
<td>Lebanon</td>
<td>0.10</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Week of May 5, 2023</strong></td>
<td></td>
</tr>
<tr>
<td>Libya</td>
<td>0</td>
</tr>
<tr>
<td>Oman</td>
<td>0</td>
</tr>
<tr>
<td>Egypt</td>
<td>0</td>
</tr>
<tr>
<td>Algeria</td>
<td>0</td>
</tr>
<tr>
<td>Morocco</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

<sup>a</sup>7-day persistence was estimated using Equation (1).

Figure 1. Timeline of speed, acceleration, jerk, and 7-day persistence in the Middle East and North Africa from 2020 to 2023.
**Figure 2.** Variant of concern (VOC) groups as a proportion of all sequenced SARS-CoV-2 specimens over time in the Middle East and North Africa from May 2020 through May 2023.

**Figure 3.** Rolling $t$ test of weekly speed equal to 10 over a 6-month window in the Middle East and North Africa from 2020 through 2023.
**Discussion**

**Principal Findings**

The principal finding to the stated objective of this study is that the MENA region meets the 3 criteria for moving from a pandemic to an endemic. First, the transmission rates before and after the WHO declaration of the end of the COVID-19 public health emergency of international concern on May 5, 2023, were well below the threshold of an outbreak. Second, the transmission and death rates approached 0 and remained extremely small and stable. Finally, the pandemic in MENA is no longer propagating forward. Our findings therefore signal that COVID-19 has changed from a pandemic to an endemic in MENA. Nonetheless, COVID-19 had profound effects on the region’s economy, political, and public health institutions since the pandemic began in spring 2020.
The MENA region saw several waves of infection over the course of the pandemic, with significant spikes driven by new SARS-CoV-2 variants and the rollback of COVID-19 restrictions. The region saw significant disparities in how effectively countries were able to enforce quarantine rules and contain the virus. During an initial wave in 2020, MENA countries with the economic, social, and political resources to institute quarantines were quick to respond to the pandemic and had substantially lower infection and fatality rates compared with countries in Europe and North America. This response allowed them to reopen their economies quickly. MENA governments similarly responded to later waves by reinitiating quarantine rules in focused areas. For example, Saudi Arabia, Kuwait, and Qatar each saw the number of COVID-19 cases drastically increase in May 2020 after restrictions were eased for Ramadan. These governments quickly reinforced strict lockdowns to control the spread of the virus [74].

Other MENA countries, which were already struggling with economic and sociopolitical instability, faced heightened crises as COVID-19 strained their already fragile societies. In Syria, which has grappled with internal conflict since 2011, COVID-19 created a health care catastrophe as the country’s already strained health care system was pushed to the breaking point. According to the WHO, only 50% of hospitals in Syria were fully functioning, and 25% were partially functioning at the start of the pandemic due to shortages of staff and medical equipment from years of war [75]. Refugee camps, which lacked proper medical staff and equipment, created grounds for the virus to spread quickly [76]. Yemen, which has dealt with a long and protracted civil conflict, faced its own crisis in the pandemic. Trade restrictions due to COVID-19 led to more than 50% of Yemen’s population facing food insecurity and hunger [77-79].

In economic impacts, the MENA region saw a 3.5% decline in gross domestic product (GDP) in 2020 driven by COVID-19 containment measures. However, as vaccines became more prevalent and countries began easing COVID restrictions, economies rebounded, with GDP growth of 4.5% in 2021 [80]. Some countries, such as Saudi Arabia, Kuwait, and the United Arab Emirates, also faced a secondary economic shock from a sharp decline in oil prices in 2020 [81].

Many countries in the region approved large fiscal packages to support struggling economic sectors. These packages typically included direct cash support for affected workers, tax relief for individuals and corporations, investment in strained health care systems, and economic investment in specific sectors. Saudi Arabia announced a SR 70 billion (US $18.6 billion; 2.7% of GDP) private support package, which suspended government tax payments and provided additional liquidity to the private sector. Kuwait allocated KD 500 million (US $1.6 billion; 1.5% of GDP) toward economic stimulus, which included the provision of full unemployment benefits for Kuwaiti nationals [81]. Many MENA countries began the pandemic with underfunded health care systems and had to drastically increase their health care expenditure and invest in personal protective equipment, ventilators, and other medical supplies. Egypt allocated £5 billion (US $103 million) toward the purchase of necessary medical supplies and funds for additional medical staff [81]. Overall, these health and social welfare packages required MENA countries to borrow substantial amounts of money, leading to increases in government debt [82].

Given prior experience with the Middle East respiratory syndrome (MERS), many countries in the region were quickly able to enact containment strategies. The backbone of these strategies consisted of masking, border closures, contact tracing, social distancing, quarantines, and lockdowns. However, there were significant disparities in how effective these measures were implemented given the large differences in economic resources, health care infrastructure, and the political stability between countries. The United Arab Emirates was one of the first countries to issue a COVID-19 alert before the WHO declared COVID-19 a public health emergency. Saudi Arabia, a deeply religious country, barred Muslims from conducting prayers inside mosques. Pilgrimages to holy sites such as Mecca and Medina were banned for foreigners [83]. Additionally, the United Arab Emirates and Saudi Arabia levied fines against any individuals or businesses that failed to adhere to mask mandate policies [84]. Saudi Arabia also built makeshift hospitals to increase their health care capacity and accommodate surges in COVID-19 cases [85]. Israel’s Ministry of Health implemented a containment strategy focused on early travel restrictions to countries reporting COVID-19 cases, and 14-day self-quarantine measures were imposed on anyone arriving from abroad [86]. More extensive lockdown measures were ultimately implemented, including the closure of schools, universities, and all nonessential businesses [87]. Another cornerstone of Israel’s containment strategy was the use of its extensive health care resources to test and quarantine patients at scale [88]. Israel also utilized mobile apps and spatial data collected from mobile phones to effectively perform contact tracing [88,89].

Some countries in the region lacked the health care resources to effectively fight the spread of the virus. At the start of the pandemic, Morocco had a hospital bed capacity of 1.1 beds for 1000 people, among the lowest in MENA [90]. Morocco therefore adopted an early, strict containment strategy to prevent straining their limited health care system. Policies included suspension of public events, suspension of international travel, and restrictions on intercity travel [18,21,25,90,91]. Additionally, the government created a “COVID-19 Fund,” which provided funding to increase hospital beds and intensive care unit capacity, purchase personal protective equipment, and increase testing capabilities [92]. Morocco experienced one of the lowest fatality rates in the region during the initial wave of infections [92].

Yemen faced difficulties implementing a comprehensive containment strategy given the violent civil conflict facing the country. Health infrastructure lacked the capability to properly test patients and track the virus [93]. In parts of the country controlled by the Houthis, faction, almost no COVID-19 restrictions were implemented [81,94].

As the pandemic continued, countries eventually relaxed their restrictions to stimulate the economy and only reinstated lockdowns when COVID-19 cases began to spike [22,95]. MENA countries focused on achieving herd immunity through widespread vaccination campaigns. The Pfizer-BioNTech
vaccine was the first vaccine approved in Saudi Arabia in mid-December 2020, with millions of vaccine doses being administered throughout the pandemic [96]. In July 2021, Saudi Arabia was also one of the world’s first countries to implement a large-scale vaccine mandate to combat a new surge in cases from the COVID-19 Delta variant. This mandate required proof of vaccination to enter public and private institutions, including schools, shops, malls, public transportation, and more [97]. The United Arab Emirates also mandated vaccinations for all people attending live events such as sports events and art exhibitions. Attendees also had to provide a negative polymerase chain reaction (PCR) test within 48 hours of the event [98]. Egypt announced that COVID-19 vaccinations were mandatory for people over the age of 18 years. The country banned all nonvaccinated people from entering public, governmental, and educational buildings to further incentivize vaccination [99]. In June 2021, Egypt also signed an agreement with Chinese vaccine manufacturer SinoVac to locally produce the company’s inactivated vaccine, CoronaVac [100,101].

There were large disparities in access to vaccines, with richer nations such as the United Arab Emirates and Saudi Arabia able to vaccinate their population much faster than less-resourced countries. For example, by early December 2021, the United Arab Emirates had one of the world’s highest vaccination rates at 90% of the population, while Yemen had just reached a rate of 1% [23,82].

One unique issue facing countries in the MENA region was the need for halal-certified vaccines. Many countries in the region have significant Muslim populations [102]. Islamic law requires that Muslims avoid using vaccines manufactured with certain types of ingredients (such as pork products) [102]. To avoid potential vaccine hesitancy, religious leaders in several MENA countries, including the United Arab Emirates and Egypt, declared vaccines halal and encouraged citizens to get vaccinated to prevent further spread of the virus [103].

Limitations

Limitations of our data analysis and resulting manuscript include the following. COVID-19 data had become less frequently reported around the world by the time the WHO declared an end to the COVID-19 Emergency of International Concern [104]. Additionally, more people began to use at-home tests as the pandemic evolved [105]. Because the enhanced surveillance metrics of speed, acceleration, jerk, and 7-day persistence are based on rates, not total counts, statistical bias caused by countries dropping in or out of the sample is mitigated, but to the extent that a nonincluded country is unrepresentative of the region in disease burden, the omission of a country or territory can still influence historical data comparisons. Viral specimen tests for variants of concern in GISAID are also dependent on testing and sequencing capacity, which varied by country across the region.

Conclusion

Overall, the COVID-19 pandemic had far-reaching effects on MENA, impacting health care systems, economies, and social well-being. Although the region fared better in disease burden than several others, many countries in the region continue to face challenges due to limited economic and health care resources [48,106-108]. Although COVID-19 continues to circulate in MENA, the rate of transmission remained well below the threshold of an outbreak for over 1 year ahead of the WHO declaration.

The concern about potential resurgences of the SARS-CoV-2 virus is valid [20,109-111]. As long as COVID-19 continues to spread and mutate, the possibility of new variants emerging remains. Variants could potentially be more transmissible, resistant to vaccines, or cause more severe illness. For example, 4 months after the WHO declared the end to the public health emergency, Omicron has further mutated into Omicron EG.5, incorrectly referred to as Eris and sensationalized in the media [112]. There is no evidence that this subclade results in more severe disease or death [113]. This underscores the importance, however, of continued vigilance, vaccination efforts, and global cooperation to control the spread of the virus [19,24,26,27,51,114-116].

Our findings underscore the importance of the following efforts on a country and regional basis: government investment in epidemiological surveillance and preparedness for future pandemics and public health emergencies; identification of barriers to access of key public health resources including care (including acute and preventative care [ie, vaccines]), education, and material goods and financial support; additional crisis management preparations in the social and economic spheres given their intersectionality with public health; regular assessment of social attitudes toward health crises, including both ongoing and potential future emergencies.

Acknowledgments

Research reported in this publication was generously supported by the Robert J Havey, MD Institute for Global Health’s Global Innovation Challenge Award at Northwestern University, Feinberg School of Medicine. This study was also supported by Feed the Future through the US Agency for International Development, under the terms of contract number 7200LAI800003 and Feed the Future Innovation Lab for Collaborative Research on Sustainable Intensification (grant number AID-OAA-L-14-00006). The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the US Agency for International Development or the authors’ organizations.

We gratefully acknowledge all data contributors (ie, the authors and their originating laboratories responsible for obtaining the specimens and their submitting laboratories for generating the genetic sequence and metadata and sharing via the Global Initiative on Sharing All Influenza Data [GISAID] Initiative, on which this research is based).
Generative artificial intelligence (AI) was not used to assist or perform any portion of manuscript creation, including data collection and analysis, drafting, or editing.

Data Availability
All data in this study are unrestricted and publicly available in the Our World In Data (OWID) and Global Initiative on Sharing All Influenza Data (GISAID) repositories described in the Methods section [55,56].

Conflicts of Interest
None declared.

Multimedia Appendix 1
Static surveillance metrics for the weeks of Apr 28, 2023 (Pre-Declaration) and May 5, 2023 (Post-Declaration).

References


56. Global Initiative on Sharing All Influenza Data (GISAID). URL: https://www.gisaid.org/ [accessed 2024-04-18]


77. Stone R. Yemen was facing the world's worst humanitarian crisis. Then the coronavirus hit. Science. 2020 May 28. URL: https://www.science.org/content/article/yemen-was-facing-worlds-worst-humanitarian-crisis-then-coronavirus-hit [accessed 2023-09-07]


Abbreviations

- **CDC**: Centers for Disease Control and Prevention
- **GDP**: gross domestic product
- **GISAID**: Global Initiative on Sharing All Influenza Data
- **MENA**: Middle East and North Africa
- **MERS**: Middle East respiratory syndrome
- **OWID**: Our World in Data
- **PCR**: polymerase chain reaction
- **WHO**: World Health Organization

©Alan G Soetikno, Alexander L Lundberg, Egon O Ozer, Scott A Wu, Sarah B Welch, Maryann Mason, Yingxuan Liu, Robert J Havey, Robert L Murphy, Claudia Hawkins, Charles B Moss, Lori Ann Post. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 12.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.

https://publichealth.jmir.org/2024/1/e53219

JMIR Public Health Surveill 2024 | vol. 10 | e53219 | p.451

(Paint number not for citation purposes)
Adherence to 24-Hour Movement Guidelines Among Chinese Older Adults: Prevalence, Correlates, and Associations With Physical and Mental Health Outcomes

Wei Liang1*, PhD; Yanping Wang2*, MEd; Qian Huang3, MEd; Borui Shang4, PhD; Ning Su1, PhD; Lin Zhou5, PhD; Ryan E Rhodes6, PhD; Julien Steven Baker2, PhD, DSC; Yanping Duan2, PhD

1School of Physical Education, Shenzhen University, Shenzhen, China
2Department of Sport, Physical Education and Health, Hong Kong Baptist University, Hong Kong, China (Hong Kong)
3Fitness and Health Lab, Hubei Institute of Sport Science, Wuhan, China
4Department of Social Sciences, Hebei Sports University, Shijiazhuang, China
5School of Physical Education, Hebei Normal University, Shijiazhuang, China
6School of Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada
* these authors contributed equally

Corresponding Author:
Yanping Duan, PhD
Department of Sport, Physical Education and Health
Hong Kong Baptist University
15 Baptist Road
Kowloon Tong
Hong Kong, 999077
China (Hong Kong)
Phone: 852 34113038
Fax: 852 34113357
Email: duanyp@hkbu.edu.hk

Abstract

Background: It is known that 24-hour movement behaviors, including physical activity (PA), sedentary behavior (SB), and sleep, are crucial components affecting older adults' health. Canadian 24-hour movement guidelines for older adults were launched in 2020, emphasizing the combined role of these 3 movement behaviors in promoting older adults' health. However, research on the prevalence and correlates of guideline adherence and its associations with health-related outcomes is limited, especially among Chinese older adults.

Objective: This study aimed to investigate the prevalence and correlates of meeting 24-hour movement guidelines among Chinese older adults. Furthermore, this study aimed to examine the associations of guideline adherence with older adults' physical and mental health outcomes.

Methods: Using a stratified cluster random sampling approach, a total of 4562 older adults (mean age 67.68 years, SD 5.03 years; female proportion: 2544/4562, 55.8%) were recruited from the latest provincial health surveillance of Hubei China from July 25 to November 19, 2020. Measures included demographics, movement behaviors (PA, SB, and sleep), BMI, waist circumference, waist-hip ratio (WHR), percentage body fat (PBF), systolic and diastolic blood pressure, physical fitness, depressive symptoms, and loneliness. Generalized linear mixed models were employed to examine the associations between variables using SPSS 28.0 (IBM Corp).

Results: Only 1.8% (83/4562) of participants met all 3 movement guidelines, while 32.1% (1466/4562), 3.4% (155/4562), and 66.4% (3031/4562) met the individual behavioral guidelines for PA, SB, and sleep, respectively. Participants who were older, were female, and lived in municipalities with lower economic levels were less likely to meet all 3 movement guidelines. Adhering to individual or combined movement guidelines was associated with greater physical fitness and lower values of BMI, waist circumference, WHR, PBF, depressive symptoms, and loneliness, with the exception of the relationship of SB+sleep guidelines with loneliness. Furthermore, only meeting SB guidelines or meeting both PA and SB guidelines was associated with lower systolic blood pressure.
Conclusions: This is the first study to investigate adherence to 24-hour movement guidelines among Chinese older adults with regard to prevalence, correlates, and associations with physical and mental health outcomes. The findings emphasize the urgent need for promoting healthy movement behaviors among Chinese older adults. Future interventions to improve older adults’ physical and mental health should involve enhancing their overall movement behaviors and should consider demographic differences.

(KEYWORDS: physical activity; sedentary behavior; sleep; cardiometabolic indicators; physical fitness; mental health; post–COVID-19 era; older adults; COVID-19; systolic blood pressure; diastolic blood pressure; depression; loneliness)

Introduction

The number of older adults (≥60 years) worldwide was estimated to be 1 billion in 2019, and this figure is expected to double by 2050, accounting for around 22% of the global population [1]. This demographic shift has been occurring at an unprecedented pace and may accelerate in the coming decades, and it will bring considerable challenges to worldwide societies, especially in developing countries [1,2]. As a vulnerable group, older adults have shown low levels of physical fitness and high levels of morbidity and mortality of infectious respiratory diseases, cardiometabolic diseases, and mental disorders (eg, late-life depression), which have worsened because of the outbreak and continuation of the COVID-19 pandemic [3-5]. The daily routines of older adults have been substantially altered [6]. They have been challenged by requirements to increase their time living at home, limits to physical and social connections with other family members and friends, temporary decreases or cessation of employment and recreational activities, loneliness, and fear of illness and death for themselves and others [3,6-8]. Therefore, promoting physical and mental health among older adults during the pandemic and beyond to achieve healthy aging is a public health and socioeconomic imperative globally [7].

It has been shown that 24-hour movement behaviors, including physical activity (PA), sedentary behavior (SB), and sleep, have prominent impacts on a wide range of physical and mental health outcomes among older adults [9-11]. For example, regular engagement in PA has been shown to be reliably associated with better health-related outcomes, such as BMI [12,13], percentage body fat (PBF) [12,13], waist circumference [14,15], waist-hip ratio (WHR) [15,16], systolic and diastolic blood pressures (SBP and DBP, respectively) [17], physical fitness [18,19], and depression and loneliness [20-22]. Similarly, adequate sleep duration has been shown to be associated with greater cardiorespiratory fitness [23,24] and decreased risks of metabolic diseases [25] and mental disorders [26] among older adults. In contrast, prolonged sedentary time has been shown to be a modifiable risk factor that negatively affects the physical and mental health of older adults [19,27-29].

Historically, most studies have focused only on the effect of one of these specific movement behaviors on health-related outcomes, whereas the interrelationship and joint effects of these movement behaviors were comparatively ignored [30,31]. In recent decades, an increasing collection of evidence has supported the combined effects on health among different populations, including preschool children [32], school children [33-36], adolescents [33,36], and adults and older adults [30,31]. Drawing on the theoretical underpinnings of time-use epidemiology and bolstered by recent empirical findings, the Canadian 24-hour Movement Guidelines for adults aged 18-64 years and those aged 65 years or older were established in 2020 [31]. These guidelines advocate for adults aged 18 years or older to engage in a minimum of 150 minutes of moderate-to-vigorous physical activity (MVPA) cumulatively per week. Additionally, they recommend limiting SB to no more than 8 hours daily, with recreational screen time not exceeding 3 hours, and maintaining a sleep duration of 7-9 hours for adults aged 18-64 years and 7-8 hours for those aged 65 years or older. Furthermore, the guidelines emphasize the importance of regular sleep and wake-up times.

The launch of 24-hour movement guidelines for adults has inspired relevant research interest in the combination of 24-hour movement behaviors among adult populations [37-42]. For instance, a national-level surveillance (2007-2013) found that Canadian adults (aged 18-79 years) who adhered to all 3 movement guidelines had more favorable BMI, waist circumference, aerobic fitness, and cardiometabolic biomarker levels (eg, C-reactive protein and insulin levels) [38]. A recent cross-sectional study indicated an inverse association between meeting 24-hour movement guidelines and mental health outcomes (eg, depression) among Chinese caregivers of preschoolers (mean age 35.5 years, SD 4.9 years) during the COVID-19 pandemic [41]. However, there is a lack of evidence among Chinese older adults.

Previous studies have found that meeting 24-hour movement guidelines was correlated with a range of demographic factors, such as sex, education level, marital status, health condition, and economic status among adult populations [37,38,42]. However, all the above studies focused on young individuals and middle-aged adults, while to the best of our knowledge, there is a scarcity of recent evidence on the correlates of meeting 24-hour movement guidelines among older adults (aged ≥60 years).

Therefore, this study aimed to (1) investigate the prevalence of meeting 24-hour movement guidelines among Chinese older adults; (2) investigate the correlates of meeting 24-hour movement guidelines among Chinese older adults; and (3) examine the association of adherence to 24-hour movement guidelines with physical (ie, BMI, waist circumference, WHR, PBF, SBP, DBP, and physical fitness) and mental (ie, depression and loneliness) health outcomes among Chinese older adults.
Methods

Participants and Procedure
Participants were recruited from the latest provincial health surveillance of Hubei China [43]. The HSHC is an ongoing consecutive cross-sectional surveillance that collects various health indicators on a representative sample of Chinese residents living in the Hubei province of China every 5 years [43]. A self-weighted stratified cluster random sampling approach was applied in the HSHC, where participants were randomly selected from 17 municipalities of Hubei province, stratified by communities, towns (villages), and districts (counties) for each municipality. In this study, eligible participants were required to meet the inclusion criteria, including (1) age of 60-79 years, (2) adequate language skills (ie, reading and writing capabilities in Chinese), and (3) no restriction of physical mobility (eg, passed the Physical Activity Readiness Questionnaire [PAR-Q]). Participants excluded from this study were those outside the specified age range, those unable to read and comprehend Chinese, those failing the PAR-Q, and those diagnosed with cognitive or sleep disorders. A total of 32,080 participants were contacted, and 27,826 agreed to participate in the surveillance (86.7% response rate). After eligibility checks, 4953 eligible participants were invited to complete data collection.

Eligible participants were further invited to complete the study measures at a multi-function stadium, lasting approximately 30 minutes per person. To ensure assessment quality, for each municipality, the data collection was conducted by a trained health surveillance team, which consisted of 15-20 qualified assessors who passed a competency examination, according to consistent standard operating procedures. Data were collected from July 25, 2020 (3 months after the lockdown was withdrawn) to November 19, 2020.

Ethical Considerations
This study followed the Declaration of the Helsinki World Medical Association [44] and the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement [45]. Ethics approval was obtained from the General Administration of Sport of China (CISS-2019-01-31) and Hubei Institute of Sport Science (HISS-2019-03-01). All participants who were interested in participating in the surveillance were asked to sign a written informed consent form before the study commencement.

Measures
Adherence to 24-Hour Movement Guidelines
The Chinese version of the International Physical Activity Questionnaire long form (IPAQ-LC) was used to measure PA, SB, and sleep (interconsistency coefficient [ICC]=0.79-0.87) [46,47]. For PA, participants were asked to report the frequency (days) and duration (minutes) of 3 intensities of PA (ie, light PA, moderate PA, and vigorous PA) during the past 7 days. Weekly time of PA was calculated using the following formula: frequency\times duration. For sedentary and sleep time, participants were asked to report the time (hours and minutes per day) they spent in these 2 behaviors on weekdays and weekends separately during the past week. Daily time of SB and sleep was calculated by dividing the weekly time of SB and sleep by 7 days. According to the Canadian 24-hour movement guidelines [31], participants were categorized as 0 (meeting none of the movement guidelines), 1 (meeting 1 of the movement guidelines), 2 (meeting 2 of the movement guidelines), and 3 (meeting all 3 movement guidelines).

Physical Health Outcomes
Physical health outcomes included objectively measured BMI, waist circumference, WHR, PBF, SBP, DBP, and physical fitness. Participants were informed to not participate in any vigorous PA 12 hours before the assessment. The assessments were conducted at indoor multi-function sport gymnasiums, with ambient temperature kept constant during the measurements for all participants.

Body weight and body height were measured using a portable stadiometer (GMCS-SGJ3; to the closest 0.05 kg) and a calibrated medical digital scale (GMCS-RCS3), which were further used to calculate BMI (kg/m²) [48]. Waist and hip circumference were measured using specific tape measures (GMCS-WD3; to the closest 0.1 cm). WHR was calculated as waist circumference (cm) divided by hip circumference (cm). PBF was measured using a portable bioelectrical impedance device (GMCS-TZL3). Participants were measured after either an overnight or 2-hour fast and were asked to remove their footwear and socks before stepping on to the measurement instrument. The whole assessment followed a standard procedure (eg, placing the feet on 4 pads, keeping the arms straight down, and not touching the inner thighs) guided by qualified assessors. SBP and DBP were measured after participants sat for 15 minutes, using digital instruments (GMCS-XY3).

Physical fitness was measured according to the standard protocol of the National Physical Fitness Surveillance of China [49,50]. The entire physical fitness assessment included 7 tests in this study: (1) vital capacity test using a spirometer (GMCS-FHL), (2) handgrip strength test using a mechanical dynamometer (GMCS-WCS3), (3) chair sit-and-reach test, (4) 30-second chair stand test, (5) 2-minute step test, (6) 1-leg standing with eyes closed balance test, and (7) choice reaction time test using a traditional test plate (GMCS-FYS). Each test was conducted twice, and the best performance of the 2 trials was recorded for analysis. The total physical fitness score was the sum of the weighted score of each test, ranging from 10 to 100, with a higher score indicating a greater physical fitness [49,50]. Prior to testing, all participants were fully familiarized with the measurement procedures.

Mental Health Outcomes
Depression was measured using the Chinese version of the Patient Health Questionnaire-9 (PHQ-9) [51,52]. Following the instruction question “how often were you bothered by the following problems over the past two weeks...”, participants were asked to give answers to 9 situations (eg, “little interest or pleasure in doing things”) on a 4-point Likert scale ranging from 0 (“not at all”) to 3 (“nearly every day”) (Cronbach α=.88). The total score of the 9 items was calculated (range 0-27), with
a higher score indicating a more severe level of depressive symptoms.

Loneliness was measured using the Chinese version of the 10-item Emotional and Social Loneliness Scale (ESLS-10) [53,54]. Participants were asked to answer how often the 10 designated feelings occurred over the past year (eg, “I feel as if nobody really understands me”). Responses were indicated on a 5-point Likert scale ranging from 1 (“not at all”) to 5 (“very often”) (Cronbach α=.85). The total score of the 10 items was calculated (range 10-50), with a higher score reflecting a higher level of loneliness.

Covariates
Covariates were chosen in accordance with prior research [30,31,37] and included age, gender, place of residence (urban/rural), educational attainment, marital status, chronic diseases (eg, hypertension, cardiovascular diseases, stroke, osteoporosis, cancer, and type 2 diabetes), current smoking and alcohol consumption behaviors, and the economic status of municipalities (as determined by provincial gross domestic product [GDP] statistics) [55].

Statistical Analysis
Data analyses were performed using SPSS 28.0 (IBM Corp). We excluded 248 cases owing to missing demographic details (73 cases lacking age and gender information), movement behaviors (73 cases), and physical and mental health outcomes (248 cases without data on BMI, PBF, physical fitness, and 2 mental health indicators) and 143 cases owing to invalid or abnormal values in movement behaviors (122 cases) and health-related outcomes (76 cases; eg, BMI and PBF), particularly those cases with skewness and kurtosis absolute values beyond ±1.5 and z-scores exceeding the ±3 SD threshold [56]. Thus, data from 4562 participants were retained for the final analysis. Based on a retrospective power estimate, the final sample size of 4562 was adequate to detect an effect size (Cohen $f^2$) of 0.01, with an $\alpha$ of .05 and a statistical power (1-β) of 0.8, in the regression model [35,41]. Odds ratios (ORs) of 1.68, 3.47, and 6.71 were considered to indicate a small, medium, and large effect size, respectively [57]. A statistical significance level of $P<.05$ (2-tailed) with 95% CI not covering 0 was used for all analyses.

Results
Sample Characteristics
Of 4953 participants, 4562 (retention rate of 92.1%) were included in the final data analysis (Figure 1). In addition, 69.2% (3157/4562) of participants were aged 65 years or older, and the mean age of the total sample was 67.68 (SD 5.03) years. Moreover, 55.8% (2544/4562) were female and 53.0% (2417/4562) were living in urban areas. Participants with an education level of secondary school accounted for the largest proportion (2115/4562, 46.4%), while only 12.2% (555/4562) of participants had an education level of college or above. Furthermore, 86.9% (3964/4562) of participants were married, and 46.8% (2134/4562) reported a history of chronic diseases. Details of the study sample are presented in Table 1.
Figure 1. STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) diagram of the study process.

- Participants invited (n=32,080)
  - Participants accepted (n=27,826)
    - Excluded for not meeting the eligibility criteria (n=22,873)
    - Participants accepted and eligible (n=4953)
      - Invalid, abnormal, and missing data (n=391):
        - Demographics (n=73)
        - 24-hour movement behaviors (n=195)
        - Health-related outcomes (n=324)
  - Participants for analyses (n=4562)
Table 1. Descriptive characteristics of the study sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=4562)</th>
<th>Age 60-64 years (n=1405)</th>
<th>Age ≥65 years (n=3157)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>67.68 (5.03)</td>
<td>62.12 (1.36)</td>
<td>70.15 (4.00)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2018 (44.2)</td>
<td>619 (44.1)</td>
<td>1399 (44.3)</td>
</tr>
<tr>
<td>Female</td>
<td>2544 (55.8)</td>
<td>786 (55.9)</td>
<td>1758 (55.7)</td>
</tr>
<tr>
<td><strong>Residence, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>2417 (53.0)</td>
<td>782 (55.7)</td>
<td>1635 (51.8)</td>
</tr>
<tr>
<td>Countryside</td>
<td>2145 (47.0)</td>
<td>623 (44.3)</td>
<td>1522 (48.2)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school or below</td>
<td>1892 (41.5)</td>
<td>456 (32.5)</td>
<td>1436 (45.5)</td>
</tr>
<tr>
<td>Secondary school</td>
<td>2115 (46.4)</td>
<td>753 (53.5)</td>
<td>1362 (43.1)</td>
</tr>
<tr>
<td>College or above</td>
<td>555 (12.2)</td>
<td>196 (14.0)</td>
<td>359 (11.4)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>3964 (86.9)</td>
<td>1277 (90.9)</td>
<td>2687 (85.1)</td>
</tr>
<tr>
<td>Single/divorced/widowed</td>
<td>598 (13.1)</td>
<td>128 (9.1)</td>
<td>470 (14.9)</td>
</tr>
<tr>
<td><strong>Chronic disease, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2428 (53.2)</td>
<td>836 (59.5)</td>
<td>1592 (50.4)</td>
</tr>
<tr>
<td>Yes</td>
<td>2134 (46.8)</td>
<td>569 (40.5)</td>
<td>1565 (49.6)</td>
</tr>
<tr>
<td><strong>Smoking, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not currently</td>
<td>648 (14.2)</td>
<td>235 (16.7)</td>
<td>413 (13.1)</td>
</tr>
<tr>
<td>Yes, but not everyday</td>
<td>184 (4.0)</td>
<td>65 (4.6)</td>
<td>119 (3.8)</td>
</tr>
<tr>
<td>Yes, almost everyday</td>
<td>3730 (81.8)</td>
<td>1105 (78.6)</td>
<td>2625 (83.1)</td>
</tr>
<tr>
<td><strong>Alcohol, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>3423 (75.0)</td>
<td>1030 (73.3)</td>
<td>2393 (75.8)</td>
</tr>
<tr>
<td>Seldom</td>
<td>627 (13.7)</td>
<td>220 (15.7)</td>
<td>407 (12.9)</td>
</tr>
<tr>
<td>Often</td>
<td>512 (11.2)</td>
<td>155 (11.0)</td>
<td>357 (11.3)</td>
</tr>
<tr>
<td><strong>Municipality economic status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥9th in terms of provincial GDP</td>
<td>2344 (51.4)</td>
<td>676 (48.1)</td>
<td>1668 (52.8)</td>
</tr>
<tr>
<td>&lt;9th in terms of provincial GDP</td>
<td>2218 (48.6)</td>
<td>729 (51.9)</td>
<td>1489 (47.2)</td>
</tr>
<tr>
<td><strong>Physical health outcomes, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
<td>24.36 (3.05)</td>
<td>23.91 (2.72)</td>
<td>24.55 (3.17)</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>86.46 (9.38)</td>
<td>84.72 (8.48)</td>
<td>87.24 (9.65)</td>
</tr>
<tr>
<td>WHR(^b)</td>
<td>0.92 (0.07)</td>
<td>0.91 (0.07)</td>
<td>0.93 (0.07)</td>
</tr>
<tr>
<td>PBF(^c) (% )</td>
<td>27.92 (6.87)</td>
<td>27.14 (6.55)</td>
<td>28.27 (6.98)</td>
</tr>
<tr>
<td>SBP(^d) (mmHg)</td>
<td>139.63 (18.52)</td>
<td>139.41 (18.49)</td>
<td>139.73 (18.55)</td>
</tr>
<tr>
<td>DBP(^e) (mmHg)</td>
<td>82.67 (10.98)</td>
<td>82.81 (10.82)</td>
<td>82.61 (11.06)</td>
</tr>
<tr>
<td>Physical fitness</td>
<td>59.50 (11.91)</td>
<td>60.10 (10.06)</td>
<td>59.24 (12.64)</td>
</tr>
<tr>
<td><strong>Mental health outcomes, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>2.59 (3.35)</td>
<td>1.99 (2.65)</td>
<td>2.86 (3.58)</td>
</tr>
<tr>
<td>Loneliness</td>
<td>21.59 (7.91)</td>
<td>20.01 (7.24)</td>
<td>22.03 (8.15)</td>
</tr>
</tbody>
</table>

\(^a\)GDP: gross domestic product.  
\(^b\)WHR: waist-hip ratio.
Prevalence of Adherence to 24-Hour Movement Guidelines

The rates of adhering to 0, 1, and 2 movement guidelines were 24.7% (1126/4562), 50.5% (2303/4562), and 23.0% (1050/4562), respectively. The proportion of participants meeting individual movement guidelines was 32.1% (1466/4562) for MVPA, 3.4% (155/4562) for SB, and 66.4% (3031/4562) for sleep. Prevalence rates of meeting a combination of 2 movement behaviors ranged from 1.8% (83/4562) to 23.3% (1063/4562), while only 1.8% (83/4562) of participants met all 3 movement guidelines (Figure 2).

Figure 2. Adherence to 24-hour movement guidelines among the study participants. The value within each circle is added to the percentage of participants meeting each individual guideline (ie, 32.1% for MVPA, 3.4% for SB, and 66.4% for sleep). The overall nonoverlapped area of each circle refers to the percentage of participants meeting 1 of the 3 guidelines (ie, 7.3%+0.1%+43.1%=50.5%). The overall overlapped area of 2 circles refers to the percentage of participants meeting 2 movement guidelines (ie, 21.5%+1.5%+0%=23.0%). The overlapped area of 3 circles refers to the percentage of participants meeting all 3 movement guidelines (ie, 1.8%). The outside area of the circle refers to the percentage of participants meeting none (0) of the guidelines (ie, 24.7%). MVPA: moderate-to-vigorous physical activity; SB: sedentary behavior.

Correlates of Adherence to 24-Hour Movement Guidelines

Table 2 presents the correlates of meeting 24-hour movement guidelines in the study sample. Participants who were older and who were female had a comparatively poorer adherence to either individual movement guidelines or the combinations of 2 or 3 behaviors (all P<.001). The municipality economic status was positively associated with meeting all 3 movement guidelines. Participants living in urban areas were more likely to adhere to MVPA and sleep guidelines (all P<.001). A higher education level and no chronic diseases were associated with a higher adherence to MVPA, sleep, and MVPA+SB guidelines, while marital status was not associated with guideline adherence.
Table 2. Correlates of adherence to 24-hour movement guidelines in the study sample (N=4562).

<table>
<thead>
<tr>
<th>Variable</th>
<th>MVPA&lt;sup&gt;a&lt;/sup&gt;, OR (95% CI)</th>
<th>SB&lt;sup&gt;c&lt;/sup&gt;, OR (95% CI)</th>
<th>Sleep, OR (95% CI)</th>
<th>MVPA+SB, OR (95% CI)</th>
<th>MVPA+sleep, OR (95% CI)</th>
<th>SB+sleep, OR (95% CI)</th>
<th>MVPA+SB+sleep, OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>0.97 (0.95-0.98)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.93 (0.90-0.97)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.97 (0.96-0.98)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.94 (0.90-0.97)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.97 (0.96-0.99)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.94 (0.89-0.98)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.94 (0.89-0.98)&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (reference)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Female</td>
<td>0.55 (0.47-0.64)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.40 (0.27-0.60)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.55 (0.47-0.64)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.39 (0.26-0.59)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.56 (0.47-0.66)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.36 (0.21-0.61)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.35 (0.21-0.61)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Residence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban (reference)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Countryside</td>
<td>0.75 (0.66-0.86)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.83 (0.59-1.16)</td>
<td>0.78 (0.69-0.89)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.80 (0.57-1.13)</td>
<td>0.77 (0.66-0.89)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.82 (0.52-1.32)</td>
<td>0.83 (0.52-1.33)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school or below (reference)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Secondary school</td>
<td>1.19 (1.03-1.37)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.93 (0.64-1.35)</td>
<td>1.06 (0.92-1.22)</td>
<td>0.96 (0.66-1.39)</td>
<td>1.21 (1.03-1.43)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.38 (0.80-2.39)</td>
<td>1.40 (0.81-2.41)</td>
</tr>
<tr>
<td>College or above</td>
<td>1.41 (1.13-1.75)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.81 (0.47-1.40)</td>
<td>1.86 (1.46-2.38)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.82 (0.47-1.42)</td>
<td>1.54 (1.22-1.95)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.46 (0.71-3.01)</td>
<td>1.50 (0.73-3.07)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married (reference)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Single/divorced/widowed</td>
<td>1.15 (0.95-1.39)</td>
<td>1.03 (0.61-1.74)</td>
<td>1.11 (0.92-1.33)</td>
<td>0.88 (0.50-1.55)</td>
<td>1.22 (0.99-1.51)</td>
<td>0.74 (0.31-1.72)</td>
<td>0.96 (0.45-2.03)</td>
</tr>
<tr>
<td><strong>Chronic diseases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Yes</td>
<td>0.87 (0.76-0.99)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.20 (0.86-1.67)</td>
<td>0.60 (0.53-0.69)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.20 (0.86-1.67)</td>
<td>0.85 (0.74-0.99)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.03 (0.66-1.62)</td>
<td>1.04 (0.66-1.63)</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not currently (reference)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Yes, but not everyday</td>
<td>0.63 (0.43-0.90)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.59 (0.19-1.30)</td>
<td>1.15 (0.79-1.67)</td>
<td>0.50 (0.19-1.29)</td>
<td>0.78 (0.52-1.16)</td>
<td>0.42 (0.10-1.83)</td>
<td>0.42 (0.10-1.85)</td>
</tr>
<tr>
<td>Yes, almost everyday</td>
<td>1.13 (0.92-1.38)</td>
<td>0.90 (0.58-1.40)</td>
<td>1.05 (0.85-1.30)</td>
<td>0.89 (0.57-1.38)</td>
<td>1.18 (0.95-1.47)</td>
<td>0.94 (0.51-1.72)</td>
<td>0.95 (0.52-1.74)</td>
</tr>
<tr>
<td><strong>Alcohol</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never (reference)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Seldom</td>
<td>1.33 (1.09-1.63)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.12 (0.69-1.81)</td>
<td>0.93 (0.76-1.15)</td>
<td>1.13 (0.70-1.82)</td>
<td>1.24 (1.01-1.54)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.94 (0.48-1.81)</td>
<td>0.94 (0.49-1.81)</td>
</tr>
<tr>
<td>Often</td>
<td>1.24 (0.99-1.54)</td>
<td>1.43 (0.89-2.28)</td>
<td>0.82 (0.66-1.04)</td>
<td>1.42 (0.89-2.28)</td>
<td>1.07 (0.84-1.35)</td>
<td>1.12 (0.89-2.13)</td>
<td>1.13 (0.59-2.15)</td>
</tr>
<tr>
<td><strong>Municipality economic status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;9th in terms of GDP&lt;sup&gt;b&lt;/sup&gt;  (reference)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>≥9th in terms of GDP</td>
<td>0.93 (0.82-1.07)</td>
<td>1.49 (1.06-2.09)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.83 (0.73-0.94)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.50 (1.06-2.10)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.94 (0.82-1.09)</td>
<td>1.67 (1.05-2.67)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.67 (1.04-2.66)&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>R&lt;sup&gt;2&lt;/sup&gt;</strong></td>
<td>0.06</td>
<td>0.05</td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>

https://publichealth.jmir.org/2024/1/e46072
Associations of Adherence to 24-Hour Movement Guidelines With Physical and Mental Health Outcomes

The associations between adherence to movement guidelines (individual and combination) and health-related outcomes are outlined in Tables 3 and 4. Adhering to either MVPA or sleep guidelines was associated with favorable physical and mental health outcomes (all $P<.001$), except 2 blood pressure indicators ($P=.08-.57$). Similar results were observed for adherence to SB guidelines, except a significant correlation with SBP ($P=.02$).

Adhering to both MVPA and SB guidelines was associated with greater health-related outcomes (all $P<.01$), except DBP ($P=.88$). Adhering to MVPA+sleep or SB+sleep guidelines was associated with better performance regarding physical health (all $P<.05$), except 2 blood pressure indicators ($P=.09-.78$). Adhering to MVPA+sleep guidelines was associated with lower levels of both depressive symptoms and loneliness (both $P<.001$), while adhering to SB+sleep guidelines was not significantly associated with loneliness ($P=.18$).

Relative to not meeting any movement guidelines, adherence to 1, 2, or 3 movement guidelines was associated with lower values of BMI, waist circumference, WHR, and PBF, alongside higher levels of physical fitness among participants (all $P<.001$). However, this association did not extend to blood pressure (both SBP and DBP; $P=.17-.89$). For mental health outcomes, adherence to 1, 2, or 3 movement guidelines was significantly and inversely associated with both depressive symptoms and loneliness (all $P<.001$) compared with not meeting any guidelines (Table 4). Dose-response associations were identified between the number of adhered guidelines and health-related outcomes (all $P<.01$), except blood pressure ($P=.18-.64$).
<table>
<thead>
<tr>
<th>Meeting movement guidelines</th>
<th>BMI ($kg/m^2$), $B$ (95% CI)</th>
<th>WC (cm), $B$ (95% CI)</th>
<th>WHR $^d$, $B$ (95% CI)</th>
<th>PBF $^e$, $B$ (95% CI)</th>
<th>SBP $^f$, (mmHg), $B$ (95% CI)</th>
<th>DBP $^g$, (mmHg), $B$ (95% CI)</th>
<th>Physical fitness, $B$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meeting individual guidelines</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least MVPA $^i$</td>
<td>$-1.22 (-1.41)$ to $-1.04$ $^j$</td>
<td>$-3.84 (-4.40)$ to $-3.27$ $^j$</td>
<td>$-0.02 (-0.02)$ to $-0.01$ $^j$</td>
<td>$-2.44 (-2.75)$ to $-2.12$ $^j$</td>
<td>$-1.05 (-2.22)$ to $0.12$</td>
<td>$-0.20 (-0.90)$ to $0.50$</td>
<td>$2.99 (2.64$ to $3.34)$ $^j$</td>
</tr>
<tr>
<td>At least SB $^k$</td>
<td>$-0.84 (-1.32)$ to $0.36$ $^j$</td>
<td>$-2.80 (-4.25)$ to $-1.35$ $^j$</td>
<td>$-0.01 (-0.02)$ to $-0.003$ $^{j,m}$</td>
<td>$-2.37 (-3.19)$ to $-1.56$ $^j$</td>
<td>$-3.61 (-6.57)$ to $-0.12$</td>
<td>$-0.01 (-1.77)$ to $1.76$</td>
<td>$2.80 (1.90$ to $3.71)$ $^j$</td>
</tr>
<tr>
<td>At least sleep</td>
<td>$-0.51 (-0.70)$ to $-0.32$ $^{j,m}$</td>
<td>$-1.40 (-1.97)$ to $-0.83$ $^{j,m}$</td>
<td>$-0.01 (-0.01)$ $^{j,m}$</td>
<td>$-0.66 (-0.98)$ to $-0.34$ $^{j,m}$</td>
<td>$0.36 (-0.80$ to $1.53)$</td>
<td>$-0.06 (-0.75)$ to $0.64$</td>
<td>$1.65 (1.30$ to $2.00)$ $^j$</td>
</tr>
<tr>
<td><strong>Meeting specific guideline combinations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least MVPA+SB</td>
<td>$-0.86 (-1.34)$ to $-0.38$ $^j$</td>
<td>$-2.95 (-4.41)$ to $-1.49$ $^j$</td>
<td>$-0.02 (-0.03)$ to $-0.004$ $^{j,m}$</td>
<td>$-2.42 (-3.25)$ to $-1.61$ $^j$</td>
<td>$-3.98 (-6.96)$ to $-0.99$ $^m$</td>
<td>$-0.13 (-1.91)$ to $1.65$</td>
<td>$2.80 (1.89$ to $3.72)$ $^j$</td>
</tr>
<tr>
<td>At least MVPA+sleep</td>
<td>$-0.97 (-1.17)$ to $-0.76$ $^j$</td>
<td>$-3.06 (-3.68)$ to $-2.43$ $^j$</td>
<td>$-0.01 (-0.02)$ to $-0.01$ $^j$</td>
<td>$-1.82 (-2.17)$ to $-1.47$ $^j$</td>
<td>$-1.12 (-2.40$ to $-0.17)$</td>
<td>$-0.22 (-0.99)$ to $0.54$</td>
<td>$2.76 (2.37$ to $3.14)$ $^j$</td>
</tr>
<tr>
<td>At least SB+sleep</td>
<td>$-0.89 (-1.54)$ to $-0.24$ $^o$</td>
<td>$-2.75 (-4.72)$ to $-0.79$ $^o$</td>
<td>$-0.02 (-0.03)$ to $-0.002$ $^j$</td>
<td>$-2.53 (-3.63)$ to $-1.43$ $^j$</td>
<td>$-3.98 (-8.00$ to $-0.03)$</td>
<td>$0.33 (-2.06$ to $2.73)$</td>
<td>$3.49 (2.27$ to $4.72)$ $^j$</td>
</tr>
<tr>
<td><strong>Number of guidelines met</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meeting 1</td>
<td>$-0.93 (-1.14)$ to $-0.71$ $^j$</td>
<td>$-5.48 (-7.47)$ to $-3.48$ $^j$</td>
<td>$-0.02 (-0.02)$ to $-0.01$ $^j$</td>
<td>$-1.50 (-1.86$ to $-1.13$ $^j$</td>
<td>$0.86 (-0.50$ to $2.21$</td>
<td>$0.02 (-0.79$ to $0.83$</td>
<td>$2.26 (1.86$ to $2.66)$ $^j$</td>
</tr>
<tr>
<td>Meeting 2</td>
<td>$-1.64 (-1.89)$ to $-1.38$ $^j$</td>
<td>$-4.97 (-5.75$ to $-4.20$ $^j$</td>
<td>$-0.02 (-0.03)$ to $-0.00$ $^j$</td>
<td>$-2.91 (-3.34$ to $-2.47$ $^j$</td>
<td>$-0.55 (-2.16$ to $1.06$</td>
<td>$-0.29 (-1.25$ to $0.67$</td>
<td>$4.30 (3.82$ to $4.77)$ $^j$</td>
</tr>
<tr>
<td>Meeting 3</td>
<td>$-1.84 (-2.50)$ to $-1.18$ $^j$</td>
<td>$-2.56 (-3.21$ to $-1.91$ $^j$</td>
<td>$-0.03 (-0.05$ to $-0.02$ $^j$</td>
<td>$-4.13 (-5.25$ to $-3.01$ $^j$</td>
<td>$-3.64 (-7.79$ to $-0.51)$</td>
<td>$0.28 (-2.20$ to $2.75$</td>
<td>$5.90 (4.67$ to $7.13)$ $^j$</td>
</tr>
<tr>
<td>Trend analysis</td>
<td>$-0.77 (-0.89$ to $-0.65$ $^j$</td>
<td>$-2.35 (-2.72$ to $-1.99$ $^j$</td>
<td>$-0.02 (-0.02$ to $-0.01$ $^j$</td>
<td>$-1.44 (-1.64$ to $-1.23$ $^j$</td>
<td>$-0.51 (-1.27$ to $0.24$</td>
<td>$-0.11 (-0.56$ to $0.34$</td>
<td>$2.11 (1.89$ to $2.33)$ $^j$</td>
</tr>
</tbody>
</table>

$^a$All models were adjusted for age, gender, residence, education, marital status, chronic diseases, smoking, alcohol consumption, and municipality economic status.

$^b$Independent variables were meeting movement guidelines and dependent variables were health outcomes.

$^c$WC: waist circumference.

$^d$WHR: waist-hip ratio.

$^e$PBF: percentage body fat.

$^f$SBP: systolic blood pressure.

$^g$DBP: diastolic blood pressure.

$^h$Not meeting individual guidelines as the reference group.

$^i$MVPA: moderate-to-vigorous physical activity.

$^j$P<.001.

$^k$SB: sedentary behavior.

$^l$P<.05.

$^m$The analysis was not robust in terms of the sensitivity analysis, with the exclusion of participants who were obese and those who had moderate or severe depressive symptoms.

$^n$Not meeting specific guideline combinations as the reference group.

$^o$P<.1.

$^p$Not meeting any guideline as the reference group.
Table 4. Associations of adherence to 24-hour movement guidelines with mental health outcomes (N=4562).

<table>
<thead>
<tr>
<th>Meeting movement guidelines</th>
<th>Depression, B (95% CI)</th>
<th>Loneliness, B (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meeting individual guidelines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least MVPA</td>
<td>−1.99 (−2.18 to −1.80)</td>
<td>−3.57 (−4.04 to −3.09)</td>
</tr>
<tr>
<td>At least SB</td>
<td>−2.05 (−2.55 to −1.54)</td>
<td>−1.95 (−3.19 to −0.71)</td>
</tr>
<tr>
<td>At least sleep</td>
<td>−2.60 (−2.79 to −2.42)</td>
<td>−4.00 (−4.47 to −3.53)</td>
</tr>
<tr>
<td>Meeting specific guideline combinations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At least MVPA+SB</td>
<td>−2.04 (−2.54 to −1.53)</td>
<td>−1.89 (−3.14 to −0.64)</td>
</tr>
<tr>
<td>At least MVPA+sleep</td>
<td>−2.01 (−2.23 to −1.80)</td>
<td>−3.53 (−4.05 to −3.00)</td>
</tr>
<tr>
<td>At least SB+sleep</td>
<td>−1.76 (−2.45 to −1.08)</td>
<td>−1.14 (−2.82 to 0.54)</td>
</tr>
<tr>
<td>Number of guidelines met</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meeting 1</td>
<td>−3.28 (−3.48 to −3.07)</td>
<td>−5.11 (−5.65 to −4.58)</td>
</tr>
<tr>
<td>Meeting 2</td>
<td>−4.43 (−4.67 to −4.19)</td>
<td>−7.35 (−7.99 to −6.72)</td>
</tr>
<tr>
<td>Meeting 3</td>
<td>−4.75 (−5.37 to −4.14)</td>
<td>−5.95 (−7.59 to −4.32)</td>
</tr>
<tr>
<td>Trend analysis</td>
<td>−2.05 (−2.16 to −1.93)</td>
<td>−3.28 (−3.58 to −2.98)</td>
</tr>
</tbody>
</table>

a All models were adjusted for age, gender, residence, education, marital status, chronic diseases, smoking, alcohol consumption, and municipality economic status.
b Independent variables were meeting movement guidelines and dependent variables were health outcomes.
c Not meeting individual guidelines as the reference group.
d MVPA: moderate-to-vigorous physical activity.
e P < .001.
f SB: sedentary behavior.
g P < .05.
h The analysis was not robust in terms of the sensitivity analysis, with the exclusion of participants who were obese and those who had moderate or severe depressive symptoms.
i Not meeting specific guideline combinations as the reference group.
j P < .01.
k Not meeting any guideline as the reference group.

Discussion

Principal Findings

This study provides timely evidence on the prevalence and correlates of adherence to 24-hour movement guidelines as well as its associations with physical and mental health outcomes among older adults. The results showed that only 1.8% (83/4562) of participants met all 3 movement guidelines, while 32.1% (1466/4562), 3.4% (155/4562), and 66.4% (3031/4562) met individual behavioral guidelines for MVPA, SB, and sleep, respectively (aim 1). Participants who were older, were female, and lived in municipalities with a lower economic status were less likely to comply with all 3 movement guidelines, while those who lived in urban areas, had higher education levels, and had no chronic diseases showed a higher adherence to specific individual movement guidelines or combinations of movement guidelines (aim 2). With regard to aim 3, adherence to either individual movement guidelines or combinations of movement guidelines was associated with a higher level of physical fitness and lower levels of BMI, waist circumference, WHR, PBF, and loneliness.
depressive symptoms, and loneliness. This did not extend to the relationship between adherence to SB+sleep guidelines and loneliness. Furthermore, adherence to movement guidelines did not correlate with 2 blood pressure indicators, except when adhering to SB or MVPA+SB guidelines.

Regarding guideline adherence (aim 1), the percentage of Chinese older adults meeting all 3 movement guidelines was lower than that of other age groups as reported in previous national surveys (eg, Chinese children and adolescents: 2.1%; Chinese caregivers of preschoolers: 15.1%; Canadian adults: 7.1%; and Thailand adults: 21.3%) [36,38,41,42]. This is consistent with previous studies, which indicated a poorer adherence to 24-hour movement guidelines among older adults compared with other age groups [30,37,38]. Age-related decreases in PA and sleep and increases in sedentary time have been demonstrated in previous studies [61,62]. This is not surprising as older adults generally experience physical and cognitive hypofunction with age, coupled with worries of life transition, diseases, and death [63], which, to some extent, may weaken the antecedents of behavioral initiation (eg, perceived capability and motivation), eventually leading to unhealthy patterns of movement behaviors (eg, physical inactivity, prolonged SB, and insufficient sleep) [64]. In addition to this explanation, the time frame of data collection should be considered. The data in our study were gathered during the COVID-19 period, and at that time, local preventive measures were still being undertaken (eg, mandatory quarantine, physical distancing, and emergent closure of some public areas). This might have contributed to the low adherence rate of meeting 24-hour movement guidelines among Chinese older adults [65,66]. Overall, the above findings underline the long-term requirement of effective behavioral promotion strategies and policy making. In addition, as there is a lack of evidence on the adherence of older adults to 24-hour movement guidelines, we could not make a comparison with previous studies using the same age group, which implies that more surveillance studies on this topic are warranted.

Regarding the correlates of movement guideline adherence among Chinese older adults (aim 2), we found that participants who were female and who were older were less likely to comply with movement guidelines, which is in accordance with the findings of previous studies involving children and adults [35-38]. Interestingly, we also found a lower adherence to all 3 movement guidelines among participants who lived in municipalities ranking <9th in terms of GDP. This might reflect the fact that compared with those who lived in cities with a lower economic status and lower modernization, Chinese older adults who lived in cities with a higher GDP had a healthy lifestyle. For other covariates, some were only associated with specific guideline adherence (eg, a higher adherence to MVPA and sleep was found among participants living in urban areas, yet residence was not associated with SB guideline adherence), where mixed results were also demonstrated in previous studies with children and adults [30,35,42,67]. Overall, the above findings imply that future health promotion programs and policy making should take age, gender, and municipality economic status differences into account. For example, more effective strategies for motivating female individuals and older age groups are needed (eg, designing programs based on participants’ preferences, tailoring the intervention content based on preidentified psychosocial determinants targeting these samples, and introducing a more supportive policy for these economically disadvantaged municipalities). In addition, more research on examining the role of other demographic factors is warranted.

Regarding the association between movement guideline adherence and health-related outcomes among Chinese older adults (aim 3), we found that adherence to movement guidelines, individually or in combination, was associated with greater physical health indicators, including BMI, PBF, waist circumference, WHR, and physical fitness. These findings are consistent with previous findings from children and adult populations [37-42], suggesting the broader applicability of these guidelines. Notably, our research showed that only adherence to SB or MVPA+SB guidelines was associated with a lower systolic pressure, contrasting with the lack of an association between adherence to other guidelines and blood pressure indicators. The limited evidence on the relationship between 24-hour movement guideline adherence and the health of older adults constrains a comprehensive comparison with existing studies. Notably, current research on movement behaviors and their relationships with blood pressure indicators (eg, using a compositional data analysis) has presented mixed findings among older adults [27,68,69]. This discrepancy may be attributed to different types of PAs (eg, muscle strength training and aerobic exercise) and dietary factors (eg, sodium intake) [70,71]. In relation to blood pressure among older adults, the quality of PA (eg, specific modality) and dietary patterns may serve as more sensitive and significant correlates than the quantity of PA (eg, minutes per week). Moreover, age-related physiological changes may reduce the sensitivity of blood pressure to movement behaviors [72], and the prevalent use of antihypertensive medications within the elderly population could also obscure the potential benefits of adhering to movement guidelines (eg, achieving the recommended levels of MVPA) [73]. Nevertheless, these assumptions were not explored in our study, highlighting the need for systematic investigations in future research.

For mental health outcomes, we found that adherence to movement guidelines was associated with lower levels of depressive symptoms and loneliness, except adherence to SB+sleep guidelines, which showed no significant relationship with loneliness. Engaging in a variety of physical activities, especially peer-based or group-oriented ones, significantly benefits the emotional and social well-being of elderly people [74]. The influence of SB and sleep duration on loneliness among older adults has not been convincingly demonstrated by prior research [27,75,76]. It is important to recognize that not all sedentary activities exert the same effect on mental well-being. Engaging in socially interactive sedentary activities (eg, internet-based social activities and playing chess with friends) may actually contribute to reducing depressive symptoms and feelings of loneliness [76,77]. The findings of our study support the connection between adherence to movement guidelines and mental health outcomes among older adults. However, evaluating activity solely by its duration does not provide a comprehensive understanding of its effect on
mental health. Future research is needed to investigate the specific mechanisms by which movement behaviors contribute to the improvement of mental health, thereby informing the development of more effective health promotion initiatives.

Finally, the results of the sensitivity analyses corroborate the robustness of our data analyses, with certain exceptions noted in the associations between adherence to SB guidelines and loneliness; adherence to sleep guidelines and BMI, waist circumference, and PBF; adherence to MVPA+SB guidelines and WHR; and adherence to MVPA+sleep guidelines and loneliness. These exceptions suggest that the weight status and the severity of depressive symptoms might affect the solidity of our findings, warranting further investigations in future studies. Additionally, our e-value analyses consistently yielded values exceeding 1, affirming the resilience of our findings against the influence of unmeasured confounders and therefore strengthening the stability of our results. These findings validate the integrity of the associations identified in our study. It is worth noting that our observations also reveal that lower adherence to movement guidelines corresponds with reduced e-values, suggesting a comparatively increased susceptibility to unobserved confounders. This discrepancy indicates the importance of cautious interpretation and necessitates further exploration of potential unmeasured confounders, such as social and environmental factors, to enhance our comprehension of the intricate relationship between adherence to 24-hour movement guidelines and health-related outcomes.

**Limitations**

Several limitations should be noted. First, although our study applied stratified random sampling with a large sample size, the study findings to some extent could only reflect the behavioral profiles of Chinese older adults living in the central region of China, and the generalizability to other regions (eg, the north and south of China) and different cultural contexts should be further examined. Second, the causal relationship between movement behaviors and health outcomes could not be well supported by the cross-sectional design. A further examination using longitudinal and experimental designs is warranted. Moreover, all the movement behaviors were evaluated by self-reported items. Although these kinds of measures have been well validated and have been shown to have advantages in several aspects (eg, could reach wider participants and be more feasible in a large sample surveillance), they might lead to measurement biases (eg, recall bias and social disabilities). Objective measures for movement behaviors are warranted in future research. In addition, for the correlates of movement guideline adherence, our findings could only explain a small percentage of the variance. Further examination of the potential correlates (eg, psychosocial and environmental factors) should be performed in future studies.

**Conclusions**

This study found that only 1.8% of Chinese older adults adhered to 24-hour movement guidelines. Older age, female sex, and lower municipality economic levels were associated with poor adherence to movement guidelines among Chinese older adults. Importantly, adherence to either individual or combined movement guidelines was associated with better physical and mental health outcomes. These findings suggest the potential benefits of promoting a holistic lifestyle encompassing adequate MVPA, reduced SB, and sufficient sleep for improving the physical and mental well-being of elderly populations.

**Acknowledgments**

This research was funded by the Humanities and Social Science Fund of the Ministry of Education of China (23YJCZH121) and the Humanities and Social Sciences Revitalization Grant of Shenzhen University (WKZX0312). The funding organization had no role in the study design, study implementation, manuscript preparation, or publication decision. This work is the responsibility of the authors.

**Authors' Contributions**

WL and YD contributed to design and methodology. YW, QH, BS, LZ, and NS contributed to investigation and data management. WL and YW contributed to data screening and analysis. WL wrote the first draft. WL, RER, NS, LZ, and YD revised the manuscript. RER and JSB performed language editing. All authors read and approved the final manuscript.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Results of sensitivity analyses.

[PDF File (Adobe PDF File), 269 KB - publichealth_v10i1e46072_app1.pdf ]

**References**


Abbreviations

- DBP: diastolic blood pressure
- GDP: gross domestic product
- HSHC: health surveillance of Hubei China
- MVPA: moderate-to-vigorous physical activity
- OR: odds ratio
- PA: physical activity
- PAR-Q: Physical Activity Readiness Questionnaire
- PBF: percentage body fat
- PHQ-9: Patient Health Questionnaire-9
- SB: sedentary behavior
- SBP: systolic blood pressure
- WHR: waist-hip ratio

Edited by A Mavragani, T Sanchez; submitted 29.01.23; peer-reviewed by MD Tarjo, S Jiang, J Feng; comments to author 17.03.24; revised version received 28.03.24; accepted 14.05.24; published 13.06.24.

Please cite as:

Adherence to 24-Hour Movement Guidelines Among Chinese Older Adults: Prevalence, Correlates, and Associations With Physical and Mental Health Outcomes
JMIR Public Health Surveill 2024;10:e46072
URL: https://publichealth.jmir.org/2024/1/e46072
doi:10.2196/46072
PMID:38869941
Malaria Vector Bionomics: Countrywide Surveillance Study on Implications for Malaria Elimination in India

Manju Rahi1,2,3, MD; AK Mishra4, PhD; Gyan Chand4, PhD; RK Baharia5, PhD; RK Hazara6, PhD; SP Singh5, PhD; Siraj Khan7, PhD; U Sreehari5, PhD; Divya Kamaraju1, PhD; Gaurav Kumar5, PhD; Sanjeev Kumar Gupta5, MSc; Amit Sharma2,5,8, PhD; K Raghavendra8, PhD; K Gunasekaran3, PhD; Om P Singh5, PhD; Sarala K Subbarao1, PhD

1Indian Council of Medical Research, Delhi, India
2Academy of Scientific and Innovation Research, Ghaziabad, Uttar Pradesh, India
3Vector Control Research Centre, Puducherry, India
4National Institute of Research in Tribal Health, Jabalpur, Madhya Pradesh, India
5National Institute of Malaria Research, New Delhi, India
6Regional Medical Research Centre, Bhubaneswar, Odisha, India
7Regional Medical Research Centre, Dibrugarh, Assam, India
8Molecular Medicine Group, International Centre of Genetic Engineering and Biotechnology, New Delhi, India

Corresponding Author:
Manju Rahi, MD
Indian Council of Medical Research
V. Ramalingaswami Bhawan, PO Box Number 4911
Ansari Nagar
Delhi, 110029
India
Phone: 91 9999002837
Email: drmanjurahi@gmail.com

Abstract

Background: The biological characteristics of mosquito vectors vary, impacting their response to control measures. Thus, having up-to-date information on vector bionomics is essential to maintain the effectiveness of existing control strategies and tools, particularly as India aims for malaria elimination by 2030.

Objective: This study aims to assess the proportions of vector species resting indoors and outdoors, determine their preference for host biting/feeding, identify transmission sites, and evaluate the susceptibility of vectors to insecticides used in public health programs.

Methods: Mosquito collections were conducted in 13 districts across 8 Indian states from 2017 to 2020 using various methods to estimate their densities. Following morphological identification in the field, sibling species of Anopheles mosquitoes were identified molecularly using polymerase chain reaction (PCR)–specific alleles. Plasmodium falciparum and Plasmodium vivax infections in the vectors were detected using enzyme-linked immunosorbent assay (ELISA) and PCR assays. In addition, we assessed the insecticide susceptibility status of primary malaria vectors following the World Health Organization (WHO) protocol.

Results: Anopheles culicifacies, a primary malaria vector, was collected (with a man-hour density ranging from 3.1 to 15.9) from all states of India except those in the northeastern region. Anopheles fluviatilis, another primary vector, was collected from the states of Madhya Pradesh, Maharashtra, Karnataka, and Odisha. In Haryana and Karnataka, An. culicifacies sibling species A predominated, whereas species C and E were predominant in Madhya Pradesh and Maharashtra. An. culicifacies displayed mainly endophilic behavior across all states, except in Madhya Pradesh, where the proportion of semigravid and gravid mosquitoes was nearly half of that of unfed mosquitoes. The human blood index of An. culicifacies ranged from 0.001 to 0.220 across all study sites. The sporozoite rate of An. culicifacies ranged from 0.06 to 4.24, except in Madhya Pradesh, where none of the vector mosquitoes were found to be infected with the Plasmodium parasite. In the study area, An. culicifacies exhibited resistance to DDT (dichlorodiphenyltrichloroethane; with <39% mortality). Moreover, it showed resistance to malathion (with mortality rates ranging from 49% to 78%) in all districts except Angul in Odisha and Palwal in Haryana. In addition, resistance to deltamethrin was observed in districts of Maharashtra, Gujarat, Haryana, and Karnataka.
Conclusions: Our study offers vital insights into the prevalence, resting behavior, and sibling species composition of malaria vectors in India. It is evident from our findings that resistance development in An. culicifacies, the primary vector, to synthetic pyrethroids is on the rise in the country. Furthermore, the results of our study suggest a potential change in the resting behavior of An. culicifacies in Madhya Pradesh, although further studies are required to confirm this shift definitively. These findings are essential for the development of effective vector control strategies in India, aligning with the goal of malaria elimination by 2030.

(JMIR Public Health Surveill 2024;10:e42050) doi:10.2196/42050

KEYWORDS
malaria; bionomics; sibling species; insecticide resistance; elimination; India

Introduction

Malaria continues to stand as a significant contributor to both morbidity and mortality within the realm of vector-borne diseases. As outlined in the World Malaria Report of 2021, there were 241 million reported cases and 627,000 deaths attributed to malaria, with the majority occurring in African nations (accounting for 95% of the total cases) [1]. The World Health Organization South-East Asia Region (WHO-SEARO) contributed approximately 2% to the global malaria burden. Over the past 2 decades, malaria incidence in this region has seen a notable decline, dropping from 23 million cases in 2000 (18 cases per 1000) to approximately 5 million cases in 2020 (3 cases per 1000). India has shown remarkable progress within the WHO-SEARO zone, achieving substantial reductions from approximately 20 million malaria cases in 2000 to around 4.1 million cases in 2020. However, despite this improvement, India still represented 83% of malaria cases and 82% of malaria fatalities in the South-East Asian region [1]. Data from the national program in 2020 further indicated that approximately 82% of malaria cases reported in India were concentrated in 6 states: Odisha, Chhattisgarh, Uttar Pradesh, Jharkhand, Maharashtra, and West Bengal [2].

The significant reduction in the malaria burden across India can be largely attributed to the expansion of control interventions. These include prompt diagnosis facilitated by the widespread availability of rapid diagnostics and microscopy, treatment using artemisinin-based combination therapy, and crucially, the effective implementation of vector control measures such as indoor residual spraying (IRS) and distribution of long-lasting insecticidal nets (LLINs). In India, 6 primary malaria vectors have been identified, namely, Anopheles culicifacies, Anopheles fluviatilis, Anopheles minimus, Anopheles stephensi, Anopheles baimaii, and Anopheles sundicus. The secondary vectors are Anopheles annularis, Anopheles philippinensis, Anopheles nivipes, and Anopheles varuna [3]. According to national guidelines, the distribution of LLINs aims for 80% coverage, with an average of 1 LLIN per 1.8 people, particularly targeting all subcenters reporting an annual parasite incidence of greater than 1. IRS, by contrast, is directed toward epidemic-prone areas and malaria-affected communities with limited access to health care services [4]. Vector control interventions have played a pivotal role and have been the primary contributors to the reduction in malaria transmission.

In the pursuit of malaria elimination, the effectiveness of vector control interventions hinges greatly on understanding the diverse biological characteristics of vectors [5]. These include physiological and behavioral traits that directly impact malaria transmission. Therefore, to maintain the efficacy of vector control measures, it is imperative to prioritize the entomological aspects of prevalent malaria vectors. This entails staying abreast of recent data on their bionomics and assessing the efficacy of chemical interventions. Such insights are crucial for evaluating the effectiveness of current vector control tools and for making informed decisions regarding the most suitable control strategies to be deployed in the field.

India’s diverse topographies, climatic conditions, and ecosystems give rise to varying levels of malaria endemcity across the country [6]. In addition to the presence of multiple vectors, India faces the challenge of vector incursion and migration. An. culicifacies, a primary malaria vector in India, has expanded its range into the northeastern states, where it has established itself as a malaria vector in this previously unaffected territory [7]. Likewise, An. stephensi, an urban malaria vector in India, has spread its presence into regions such as Sri Lanka, Djibouti, Ethiopia, and Sudan in Africa [8]. The emergence of these vectors in new areas highlights the critical need for continuous entomological surveillance to monitor and effectively address these evolving challenges. The migration of various malaria vectors across neighboring areas, ecological transitions, and colonization of new habitats through evolutionary adaptation to eco-climatic changes have been documented in numerous instances in India [9]. An entomological surveillance study [10] was conducted to monitor the Aedes invasion in the Guilan Province of Iran, which boasts several ports of entry. Although no specimens of Aedes aegypti or Aedes albopictus were collected, the importance of such regular surveillance was acknowledged [10]. Similarly, in India, the Indian Council of Medical Research’s (ICMR) Vector Control Research Centre is conducting a study to explore the relationship between human and mosquito mobility and its implications for the control of mosquito-borne infections. In this regard, this study aims to unravel the intricate phenomenon of dispersion and movement of short- and long-distance cargo, as well as human transportation, along with the surrounding travel infrastructure such as ports, stations, and markets. Recently, there has been an increasing demand for comprehensive nationwide vector surveillance studies [11]. The ICMR, headquartered in New Delhi, serves as the nodal research organization under the Ministry of Health and Family Welfare, tasked with conducting biomedical research across the country. The ICMR, through its permanent institutes located in several malaria-endemic states, has conducted extensive entomological research over the past several decades. It has made significant contributions to the understanding and dissemination of the science of vector biology
in the country, particularly in the context of malaria elimination goals [12]. The current multicenter study was conceptualized and formulated by the ICMR. In consultation with experts, common objectives were formulated, and a standardized methodology for studying vector bionomics was developed. The study was conducted by 4 research institutes of the ICMR and their respective field stations situated across different regions of the country from 2017 to 2020. The objectives of the study were to assess (1) the proportions of vector species resting indoors and outdoors over time; (2) host biting/feeding preferences, biting rhythms, and peak biting activity of vector species across different seasons; (3) the sites of transmission; and (4) the susceptibility status of vectors to various insecticides.

**Methods**

**Study Sites**

The study was conducted by 4 institutes of the ICMR: the National Institute of Malaria Research in Delhi, along with its field units in Bengaluru and Nadiad; the National Institute of Research in Tribal Health in Jabalpur, Madhya Pradesh; the Regional Medical Research Centre in Bhubaneswar, Odisha; and the Regional Medical Research Centre in Dibrugarh, Assam. A total of 13 districts from 8 Indian states, each characterized by distinct ecotypes, vector distributions, insecticide susceptibility statuses, among others, were included in the study. In each district, we identified 1 or more blocks (subdistrict level) with varying epidemiological situations. Within each block, we selected 2-3 villages, each representing a characteristic ecotype such as forest, foothill, riverine, or plains. The study encompassed all 4 prevailing seasons: premonsoon, monsoon, postmonsoon, and winter. Figure 1 illustrates the criteria, number of states, districts, blocks/community health centers (CHCs), and villages selected for the study. Figure 2 depicts the locations of the participating institutes and the districts covered by each institute.

**Figure 1.** Criteria for study site selection. GJ: Gujarat; HR: Haryana; KR: Karnataka; MG: Meghalaya; MH: Maharashtra; MP: Madhya Pradesh; NIMR: National Institute of Malaria Research; NIRTH: National Institute for Research on Tribal Health; OR: Odisha; RMRC: Regional Medical Research Center; TR: Tripura.
Ethical Approval

After community sensitization, verbal approval was sought from the chiefs and elders of the communities for mosquito collection in the selected villages. In addition, written consent was obtained from individuals who participated as human baits. Given that the study protocol entailed human landing collections, which could potentially result in mosquito bites on the human baits, efforts were made to collect mosquitoes immediately upon landing on the host, before engorgement, to prevent actual biting. The human baits received prophylaxis as part of the malaria prevention measures. Ethical approval for the study was obtained from the ethics committee of each respective participating institute.

Institutional Ethical Committee approval was obtained vide letter no. ECR/NIMR/EC/2017/142 dated 21 June 2017 by National Institute of Malaria Research.

Entomological Surveys

Overview

The various methods used for the collection of adult anopheline mosquitoes from the study sites are described in the following sections.

Indoor Resting Collections

Anopheles mosquitoes resting indoors were captured manually using a mouth aspirator and a flashlight. In each village, 4 human dwellings and 4 cattle sheds were searched for mosquitoes in the morning from 6 AM to 8 AM hours, allocating 15 minutes for each structure [13]. The standard morpho-taxonomic keys were used to identify the mosquito species collected from the field [14]. The density of each vector species, expressed as per man-hour density (MHD), was calculated from the number of female mosquitoes, as per the formula provided in the “Data Analysis” section.

Light Trap Collections (Center for Disease Control Light Trap)

Adult vector density was also monitored using light traps. These traps were suspended inside human dwellings near eaves, sleeping hosts, and doors for mosquito collection from 6 PM to 6 AM. Outdoor collections were conducted by hanging traps both near the houses and in open areas away from the houses. The collected mosquitoes were placed in cartons lined with wet towels at the bottom and kept outside during transportation to maintain a temperature of 26-28°C and a relative humidity of 70%-80%. Vector density, defined as the number of female mosquitoes collected per trap night indoors or outdoors, was calculated from the light trap collections.

Pyrethrum Spray Collection

Another method used in the study for collecting adult mosquitoes that were resting indoors was pyrethrum spray collection. Most, if not all, of the Anopheles mosquitoes resting indoors were collected using this method during the morning hours (8 AM to 10 AM). In this method, the entire floor of the room (human dwelling) was covered with a white cotton sheet. Using a flit sprayer, the complete room was sprayed with 0.1%-0.2% pyrethrum spray, causing all mosquitoes resting inside the room to be knocked down onto the sheet. The collected mosquitoes were then transferred into petri dishes lined with wet cotton or filter paper and transported to the laboratory. This method provided the total number of mosquitoes and species resting per house or structure.
Mosquito Landing Collections on Human Baits

A whole-night collection of mosquitoes landing on human bait was conducted from dusk to dawn (6 PM to 6 AM). Landing collections were made hourly, continuously for 12 hours. Mosquitoes were collected from the exposed legs of the human bait. Efforts were made to collect mosquitoes immediately upon landing on the host, before engorgement, to prevent actual biting. Mosquitoes collected from each hour were placed in separately labeled paper cups covered with a mosquito net. These collections were used for estimating the man-biting/landing density and determining the entomological inoculation rate of vector species.

Laboratory Processing

Mosquito Species Identification

Mosquitoes collected through various methods were morphologically identified to different Anopheles species using the taxonomic keys provided by Nagpal and Sharma [14].

Sibling Species Identification

The mosquito specimens preserved in isopropanol underwent molecular identification to distinguish sibling species of An. culicifacies and An. fluviatilis, following the methods outlined by Goswami et al [15] and Singh et al [16], respectively.

Blood Meal Preferences

Blood from the stomachs of fully-fed mosquitoes obtained from the field was collected onto Whatman No. 1 filter paper to identify the source of the blood meal against human and bovine antisera. The human blood index (HBI) of An. culicifacies was determined using the gel diffusion technique [17].

Vector Incrimination

Mosquitoes collected via various methods were examined for vector infection with human malaria parasites. The heads and thoraces of the mosquitoes were dissected and stored in isopropanol at –20°C until use. Both polymerase chain reaction (PCR) and enzyme-linked immunosorbent assay (ELISA) were used as the primary methods for detecting sporozoites in mosquitoes, and both methods were applied in our study. Vector incrimination was conducted using the ELISA-based method to detect species-specific circumsporozoite antigen of Plasmodium falciparum, Plasmodium vivax 210, and P. vivax 247, following the protocol outlined by Wirtz et al [18,19]. In addition, diagnostic PCR was used for the detection of malaria parasites. In our study, a total of 4067 An. culicifacies mosquitoes were tested by PCR, and 4164 An. culicifacies mosquitoes were tested by ELISA. In our study, specific antibodies against circumsporozoite protein were used in ELISA, enabling the identification of Plasmodium parasite species as well as the subtyping of P. vivax sporozoites. Furthermore, PCR was conducted using the nested PCR protocol outlined by Snounou et al [20].

Insecticide Susceptibility Tests

Vector susceptibility to the insecticides used in the national control program was assessed once during the peak abundance of vector species, adhering to the guidelines set by the WHO [21]. Field-collected, preferably from unsprayed villages or houses, mixed-age vector mosquitoes were exposed to WHO papers impregnated with insecticides at diagnostic concentrations (DDT [dichlorodiphenyltrichloroethane] 4%, malathion 5%, and WHO-recommended discriminating concentrations of various synthetic pyrethroids) using WHO test kits. For each test, 100 mosquitoes were exposed in 4 or 5 replicates, with 20-25 mosquitoes per replicate for the treatment group and 50 mosquitoes in 2 replicates, with 25 mosquitoes per replicate, for the control group. The tests were conducted in a room free from insecticide contamination and maintained at a temperature of 26-28°C and a relative humidity of 70%-80% both during exposure and the subsequent 24-hour holding period. Following the 24-hour holding period, the percent mortality was calculated based on the total number of alive and dead mosquitoes in the replicates. If the control mortality fell within the range of 5%-20%, the Abbott formula was applied to correct the treatment mortality. According to the criteria outlined by the WHO, mortality rates of 98% or higher were categorized as “susceptible,” mortality rates below 90% were classified as “resistant,” and mortality rates between 91% and 97% were labeled as “possible resistance.”

Data Analysis

All data generated during the study were inputted into a computer using Microsoft Excel, and the following parameters were analyzed:

- MHD, calculated as the number of mosquitoes collected by 1 person in 1 hour. It is determined by considering the total number of mosquitoes (n) collected, the time spent (t), and the number of persons involved in the collection (p). The formula for MHD is MHD = n × 60/t × p.
- The HBI, calculated based on the proportion of fed Anopheles mosquitoes found to contain human blood.
- The sporozoite rate, calculated based on the proportion of female Anopheles mosquitoes carrying sporozoites in their salivary glands.
- The human landing density, defined as the ratio of the total mosquitoes captured landing on a human bait for a given period to the total person-nights used for the same period.
- The entomological inoculation rate, calculated from human landing/biting catches as the product of the human landing density and the sporozoite rate of mosquitoes.

Results

Anopheles Species Collection Methods and Results: Insights from Resting, Pyrethrum Spray, Light Traps, and Human Landing Surveys

The results of Anopheles species collected through various methods including resting, pyrethrum spray, light traps, and human landing at different sites are described below and summarized in Table 1.
Table 1. Malaria vector density by various collection methods in districts of India.

<table>
<thead>
<tr>
<th>State and district</th>
<th>Anopheles species</th>
<th>Indoor resting collection (per man-hour)</th>
<th>Pyrethrum spray (number per room)</th>
<th>Light trap outdoor (per trap per night)</th>
<th>Light trap indoor (per trap per night)</th>
<th>Human landing indoors (per man per night)</th>
<th>Human landing outdoors (per man per night)</th>
<th>Animal bait collection (per bait per night)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kheda</td>
<td>An. culicifacies</td>
<td>3.15</td>
<td>1.92</td>
<td>3.9</td>
<td>3.1</td>
<td>1.85</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>Panchmahal</td>
<td>An. culicifacies</td>
<td>4</td>
<td>2.46</td>
<td>3.08</td>
<td>1.41</td>
<td>1.04</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Madhya Pradesh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sidhi</td>
<td>An. culicifacies</td>
<td>9.47</td>
<td>3.94</td>
<td>3.88</td>
<td>—</td>
<td>0.5</td>
<td>0.6</td>
<td>3.11</td>
</tr>
<tr>
<td>Haryana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuh</td>
<td>An. culicifacies</td>
<td>15.95</td>
<td>10.96</td>
<td>0.41</td>
<td>—</td>
<td>—</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>Palwal</td>
<td>An. culicifacies</td>
<td>15.46</td>
<td>36.03</td>
<td>1.02</td>
<td>—</td>
<td>—</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>Maharashtra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gadchiroli</td>
<td>An. culicifacies</td>
<td>12.1</td>
<td>3.7</td>
<td>0.06</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>Karnataka</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalaburagi</td>
<td>An. culicifacies</td>
<td>5.86</td>
<td>2.3</td>
<td>0.48</td>
<td>3.71</td>
<td>12</td>
<td>—</td>
<td>12</td>
</tr>
<tr>
<td>Odisha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalahandi</td>
<td>An. culicifacies</td>
<td>—</td>
<td>—</td>
<td>1.1</td>
<td>4.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Angul</td>
<td>An. culicifacies</td>
<td>—</td>
<td>—</td>
<td>1</td>
<td>2</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gujarat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kheda</td>
<td>An. stephensi</td>
<td>0.17</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Panchmahal</td>
<td>An. stephensi</td>
<td>0.025</td>
<td>0.016</td>
<td>0</td>
<td>0</td>
<td>0.125</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Haryana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuh</td>
<td>An. stephensi</td>
<td>8</td>
<td>20.44</td>
<td>2.92</td>
<td>—</td>
<td>—</td>
<td>4.5</td>
<td>—</td>
</tr>
<tr>
<td>Palwal</td>
<td>An. stephensi</td>
<td>21.39</td>
<td>45.03</td>
<td>8.5</td>
<td>—</td>
<td>—</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>Madhya Pradesh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sidhi</td>
<td>An. fluviatilis</td>
<td>0.74</td>
<td>0.33</td>
<td>0.49</td>
<td>—</td>
<td>0.01</td>
<td>0.01</td>
<td>0.24</td>
</tr>
<tr>
<td>Karnataka</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalaburagi</td>
<td>An. fluviatilis</td>
<td>0.06</td>
<td>0.06</td>
<td>2.96</td>
<td>18.6</td>
<td>22.4</td>
<td>—</td>
<td>22.7</td>
</tr>
<tr>
<td>Odisha</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalahandi</td>
<td>An. fluviatilis</td>
<td>—</td>
<td>—</td>
<td>2.3</td>
<td>3.8</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Angul</td>
<td>An. fluviatilis</td>
<td>—</td>
<td>—</td>
<td>0.5</td>
<td>1.2</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*aMan-hour density is defined as the number of mosquitoes collected by 1 person in 1 hour.

bNot done.
Density and Proportion of Indoor Resting Anopheline Vectors

**Madhya Pradesh (Sidhi District)**

An. culicifacies and An. fluviatilis vectors were prevalent, with an MHD of 9.47 (62%) and 0.47 (4%), respectively (Table 1). In addition, 13 other Anopheles species were collected, including An. annularis, a recognized secondary vector. The majority of An. culicifacies (92.3%) and An. fluviatilis (81%) were collected from cattle sheds. July and August recorded the highest density of anopheline mosquitoes, including the 2 vector species, whereas May and June had the lowest (Figure 3).

**Figure 3.** Monthly density of Anopheles culicifacies collected from selected districts of India. MHD: man-hour density.

**Gujarat (Kheda and Panchmahal Districts)**

Three vector species, namely, An. culicifacies, An. stephensi, and An. fluviatilis, were collected. An. culicifacies predominated in Kheda and Panchmahal, with an MHD of 3.15 and 4.0, respectively. The MHD of An. stephensi and An. fluviatilis was less than 0.2. Further, 3 other Anopheles species, namely, An. subpictus, An. vagus, and An. annularis, were collected. The density of An. culicifacies was higher in cattle sheds in Panchmahal compared with that in Kheda. However, overall, anopheline density was higher in human dwellings in Kheda than in Panchmahal. Two peaks of An. culicifacies densities were observed in Gujarat, specifically in February-April and July-August.

**Karnataka (Kalaburagi District)**

In this area, An. culicifacies, An. stephensi, and An. fluviatilis were the 3 recorded vector species, while other Anopheles species, including An. annularis, An. vagus, Anopheles nigerrimus, An. tessellatus, and An. barbirostris, were also prevalent. The MHD of An. culicifacies was higher in cattle sheds than in human dwellings. An. subpictus and An. culicifacies accounted for over 95% of all anophelines, and both were perennial species, with the highest density observed in August.

**Haryana (Nuh and Palwal Districts)**

An. culicifacies and An. stephensi were primarily collected from cattle sheds. The density of these vector species was generally low but increased considerably during the postmonsoon period and the onset of winter, with an MHD ranging between 8 and 21. Among other anophelines, An. subpictus was found in high densities, followed by An. annularis and Anopheles pulcherrimus. The monthly MHDs of An. culicifacies and An. stephensi exhibited high variations across different seasons. Two peaks of densities of An. culicifacies were observed in March-April and September-October. An. stephensi showed the first peak from January to April and a more prominent peak from September to December. The peak densities of An. culicifacies were in March (48.5) and October (29.5).

**Tripura (Dhalai and South Tripura Districts) and Meghalaya (West Garo Hills and South Garo Hills Districts)**

An. minimus and An. baimaii, the primary vectors in the states, were collected in the study districts of Tripura and Meghalaya. An. minimus was found in 9 out of the 16 selected villages, but their densities were low, especially in the state of Meghalaya. In addition to the primary vectors, An. annularis, An. vagus, An. varuna, and Anopheles jeyporiensis were among the other Anopheles species collected. In Tripura, the density of Anopheles species was higher in August, whereas in Meghalaya, it peaked in September.
**Odisha (Angul and Kalahandi Districts)**

The proportion of *An. culicifacies* and *An. fluviatilis* was higher in Kalahandi than in Angul, with an increased density of *An. culicifacies* observed in both districts. Throughout the study period, 15 and 18 different *Anopheles* species were collected in Kalahandi and Angul, respectively. The density of *An. fluviatilis* was significantly higher during winter than during the rainy and summer seasons in both districts. By contrast, the densities of *An. culicifacies* were considerably higher during the rainy season than during the postmonsoon and winter seasons.

**Anopheline Density From Spray Sheet Collections**

The districts of Palwal and Nuh in Haryana exhibited the highest density of *An. culicifacies*, with 36.0 and 10.9 mosquitoes captured per room, respectively. In the remaining districts (Sidhi, Kheda, Panchmahal, Gadchiroli, and Kalaburagi), the per-room density of *An. culicifacies* ranged from 1.9 to 3.9. Similarly, collections from Palwal (45.0) and Nuh (20.4) districts showed high densities of *An. stephensi*, while in Kheda (0.02) and Panchmahal (0.06) districts in Gujarat, its density was low. *An. fluviatilis* was collected from Sidhi (Madhya Pradesh) and Kalaburagi (Karnataka) districts with a per-room density of 0.33 and 0.06, respectively.

**Anopheline Density From Light Trap Catches**

The density of vectors recorded from light trap collections, that is, the number per trap night, both indoors and outdoors, is provided in Table 1. The density of *An. culicifacies* indoors and outdoors ranged from 0.4 to 3.9 mosquitoes per trap night, except in Gadchiroli, where the density was negligible. *An. stephensi* was only collected from light traps set outdoors in Palwal (8.5 mosquitoes per trap night) and Nuh (2.9 mosquitoes per trap night) districts. *An. fluviatilis* was collected from both indoor and outdoor traps in Kalaburagi, Kalahandi, and Angul, and only from indoor traps in Sidhi. In all these districts, the density of *An. fluviatilis* was below 3.8, with the exception of Kalaburagi, where it was 18.6 from indoor trap collections.

**Abdominal Condition of Vector Species in Different Collections**

**Madhya Pradesh (Sidhi District)**

In Kusumu CHC, the proportion of semigravid plus gravid *An. culicifacies* in indoor human dwelling collections (hand catch and pyrethrum spray collections) was 32.8%, while in Sumaria CHC, it was 31.16%. In both CHCs, the proportion of unfed plus fully fed mosquitoes was 65.75% and 69.07%, respectively.

**Gujarat (Kheda and Panchmahal Districts)**

In Panchmahal, the proportion of unfed plus fully fed mosquitoes was higher in March, May, and June, whereas in all the remaining months, the proportion of semigravid plus gravid mosquitoes was higher, ranging from 51% to 100%. Similarly, in Kheda district, the proportion of the semigravid plus gravid category was similar to Panchmahal, except in January, October, and December. With a few monthly variations, the data suggest a predominantly endophilic behavior of *An. culicifacies* in the 2 districts.

**Karnataka (Kalaburagi District)**

In Kalaburagi district in Karnataka, the proportion of semigravid plus gravid mosquitoes in indoor human dwelling collections in all 3 villages, Laxmipurwadi (60.4%), Shankerpurwadi (81.9%), and Muglegaon (65%), was higher than unfed plus fully fed mosquitoes, suggesting that *An. culicifacies* continues to exhibit endophilic behavior. Similarly, the proportion of semigravid plus gravid *An. fluviatilis* was significantly higher than unfed plus fully fed mosquitoes in Lashmipurwadi village, also indicating its endophilic behavior. From the collection of other villages, no conclusion can be drawn as only a few specimens of *An. fluviatilis* were collected from Muglegaon village, and in Shankerpurwadi, the collection was nil.

**Maharashtra (Gadchiroli District)**

In Gadchiroli, the proportion of semigravid plus gravid *An. culicifacies* in human dwellings (from resting and pyrethrum spray collections) was 57.6% in Ahiri CHC and 46.1% in Dhanora CHC, indicating its endophilic behavior in the district with little difference between the 2 CHCs.

**Blood Meal Preferences**

The HBI of *An. culicifacies* recorded in different study sites is shown in Figure 4. In Madhya Pradesh (Sidhi district), the HBI of *An. culicifacies* was 0.03 and of *An. fluviatilis* was 0.09. In Gujarat (Kheda and Panchmahal districts), the HBI of *An. culicifacies* in the canal-irrigated area of Kheda district was 0.02, whereas in the riverine area, none contained human blood. In Panchmahal district, the HBI was 0.016 in the canal-irrigated area and 0.01 in the riverine area. In Karnataka (Kalaburagi district), the HBI of *An. culicifacies* and *An. fluviatilis* was 0.22 and 0.06, respectively. In Maharashtra (Gadchiroli district), the HBI of *An. culicifacies* was 0.10 in the Gadchiroli district of Maharashtra. In Haryana (Nuh and Palwal districts), the HBI of *An. culicifacies* was low, 0.09 in Nuh and 0.07 in Palwal. Similarly, the HBI of *An. stephensi* was also low in the 2 districts, 0.01 and 0.004, respectively. In Tripura (Dhalai and South Tripura districts) and Meghalaya (West Garo Hills and South Garo Hills districts), the blood meal analysis revealed a higher HBI of 2 vector species: *An. minimus* (0.76-0.78) and *An. baimei* (0.85), indicating their higher anthropophagic behavior in northeast India. The HBI of *An. jeyporiensis* was also found to be higher (0.66-0.75) in the same region. In Odisha (Angul and Kalahandi districts), the HBI of *An. fluviatilis* was higher in both districts, indicating that the species was primarily anthropophagic. The HBI observed in Kalahandi was 0.52, and in Angul, it was 0.32. By contrast, *An. culicifacies* was found to be primarily a zoophagic mosquito, as the observed HBI was 0.086 and 0.062, respectively, in the 2 districts.
Sibling Species Composition
The proportion of An. culicifacies sibling species B, C, and E was 23%, 44.4%, and 33%, respectively, in the Sidhi district of Madhya Pradesh (Figure 5). In the Kheda district of Gujarat, sibling species A/D was 12%, and B/C/E was 82%, whereas in the Panchmahal district of the same state, the corresponding proportion was 8% and 91%, indicating the predominance of sibling species B/C/E over A/D in both districts. In the Kalaburagi district of Karnataka, An. culicifacies mostly comprised species A (57.2%) and species B (42.8%). The proportion of 3 sibling species S, T, and U of An. fluviatilis was found to be 4.6%, 94.2%, and 1.2%, respectively, in this state. A total of 1050 specimens of An. culicifacies were analyzed in Maharashtra (Gadchiroli district) for the sibling species composition, where sibling species C was predominant (43.9%), followed by species E (25%). In Haryana (Nuh and Palwal districts), An. culicifacies species A, which is an efficient vector, predominantly constituted 99.4% in Nuh and 99.4% in Palwal.

Vector Infection
Madhya Pradesh (Sidhi District)
A total of 1038 An. culicifacies and 27 An. fluviatilis specimens were screened for sporozoite positivity using PCR assay, and none were found positive (Figure 6).
Gujarat (Kheda and Panchmahal Districts)
In Kheda district, 916 An. culicifacies specimens were analyzed, and 12 were detected with malaria parasites (7 P. falciparum positive and 5. P. vivax positive), with a sporozoite positivity rate of 1.31%. In Panchmahal district, 4 An. culicifacies (3 P. falciparum positive and 1 P. vivax) specimens were found positive (n=900) with a sporozoite positivity rate of 0.44%.

Karnataka (Kalaburagi District)
In total, 1305 An. culicifacies and 691 An. fluviatilis specimens were screened, and the sporozoite rate was 0.08% and 0.14%, respectively.

Maharashtra (Gadchiroli District)
Out of 3029 An. culicifacies tested, only 2 were positive, showing a sporozoite positivity rate of 0.06%.

Haryana (Nuh and Palwal Districts)
In Nuh, 831 An. culicifacies specimens were analyzed for Plasmodium infection, and 18 were positive, with a sporozoite positivity rate of 2.16%. There was higher positivity for P. falciparum (n=11) than for P. vivax (n=7). By contrast, in Palwal district, 9 An. culicifacies specimens were detected positive (7 Pf positive and 2 Pv) out of 212 screened, and the sporozoite positivity rate was 4.24%.

Tripura (Dhalai and South Tripura Districts) and Meghalaya (West Garo Hills and South Garo Hills Districts)
None of the primary (An. minimus and An. baimaii) or secondary vectors (An. nivipes, An. philippinensis, and An. annularis) were found infected with malaria parasites.

Odisha (Angul and Kalahandi Districts)
The sporozoite rate of An. fluviatilis was 0.22% in Kalahandi. The sporozoite rate of An. culicifacies was 2.6% and 0.66% in Kalahandi and Angul districts, respectively.

Entomological Inoculation Rate
The entomological inoculation rate of An. culicifacies in the Kheda and Panchmahal districts of Gujarat was 3.09 and 0.47, respectively, followed by 0.25 in the Kalaburagi district of Karnataka. The entomological inoculation rate could not be calculated for other sites.

Vector Susceptibility to Insecticides
Overview
Susceptibility testing was conducted on An. culicifacies, An. stephensi, An. fluviatilis, An. minimus, and An. baimaii collected from various study sites against DDT, malathion, and synthetic pyrethroids, including deltamethrin, cyfluthrin, and alphacypermethrin, approved by the National Malaria Control Program. The susceptibility/resistance status is visually represented in Figures 7 and 8.
Figure 7. Insecticide susceptibility status of Anopheles culicifacies in India (≥98%-100% mortality: susceptible; 90-97% mortality: possible resistance; and <90% mortality: resistant). DDT: dichlorodiphenyltrichloroethane.
**Figure 8.** Insecticide susceptibility status of *Anopheles stephensi*, *Anopheles fluviatilis*, *Anopheles minimus*, and *Anopheles baimaii* in India (≥98-100% mortality: susceptible; 90-97% mortality: possible resistance; and <90% mortality: resistant).

**DDT (4%)**

*An. culicifacies* was tested against DDT in all districts except the Kheda and Panchmahal districts of Gujarat state, and this species exhibited resistance to DDT in all the tested districts (<39% mortality; Figure 7). *An. baimaii* and *An. minimus* were found to be susceptible to DDT in the study districts of Meghalaya and Tripura. *An. stephensi* displayed resistance to DDT in both the districts of Haryana, while *An. fluviatilis* showed possible resistance to DDT in the 2 districts of Odisha (Figure 7).

**Malathion (5%)**

*An. culicifacies* exhibited resistance to malathion (49%-78% mortality) in all districts except Angul in Odisha, where it was susceptible. *An. baimaii* was found to be susceptible to malathion in all study districts of the northeast except Dhalai, where the test could not be performed. *An. stephensi* demonstrated resistance to malathion in both districts of Haryana, whereas *An. fluviatilis* was susceptible to malathion in the 2 districts of Odisha (Figure 7).

**Deltamethrin (0.05%)**

*An. culicifacies* displayed resistance to the synthetic pyrethroid deltamethrin in Maharashtra, Gujarat, Haryana, and Karnataka study sites, but showed susceptibility in Sidhi (Madhya Pradesh) and Angul (Odisha) districts. Moreover, it exhibited possible resistance (92.5% mortality) in the Kalahandi district of Odisha. *An. stephensi* demonstrated resistance to deltamethrin in both districts of Haryana, whereas *An. fluviatilis* was found to be susceptible to this insecticide in the 2 districts of Odisha (Figure 8).

**Cyfluthrin (0.15%)**

Susceptibility tests against cyfluthrin, another synthetic pyrethroid, were conducted in Odisha and Maharashtra only. *An. culicifacies* exhibited resistance to cyfluthrin in the Gadchiroli district of Maharashtra and the Kalahandi district of Odisha, while “possible resistance” was observed in the Angul district of Odisha. *An. fluviatilis*, by contrast, was susceptible to cyfluthrin in both districts of Odisha.

**Alpha-cypermethrin (0.25%)**

Susceptibility to alpha-cypermethrin, another synthetic pyrethroid, was tested only for *An. culicifacies*. The mortality rate of *An. culicifacies* against this synthetic pyrethroid was 95%-96.7% in Panchmahal, Palwal, and Sidhi districts, indicating a possible resistance. However, in Gadchiroli (with 88.1% mortality) and Kalaburagi (with 16% mortality) districts, it exhibited resistance.
Discussion

Principal Findings

This study was conducted in 13 districts spanning 8 different states of India, encompassing various eco-epidemiological zones, including tribal, hilly, plain, and forested regions across the country. Notably, this is the first nationwide study covering nearly all parts of India, including the east, west, north, south, central, and northeast regions. Updated knowledge of bionomics is essential to bolster vector control activities and assess the effectiveness of such measures [22]. Therefore, the data generated in this study are crucial and highly pertinent to the country’s malaria elimination objective. Regular surveillance of malaria vectors is vital to identify any alterations in their distribution, behavior, and susceptibility to insecticides, which are essential for achieving the malaria elimination target. The findings from this study could contribute significantly to addressing these challenges.

Mosquito collections confirmed the presence of *An. culicifacies* in all study districts except those in Assam and Tripura states. The density of *An. culicifacies* obtained indoors through hand catch and pyrethrum spray sheet collection suggested that this species could be endophilic in most districts. However, the proportions of gravid and semigravid mosquitoes were lower (31%-33%) in Madhya Pradesh, indicating that *An. culicifacies* may not be entirely endophilic, and some proportions might also be resting outdoors. This could potentially indicate a shift in the resting behavior of *An. culicifacies* in Madhya Pradesh, although larger studies are necessary to confirm this possible change definitively. The alteration in resting behavior might be attributed to the use of LLINs inside houses. Similar changes in resting behavior have been reported for *An. fluviatilis* in Odisha, where it was predominantly found in mixed dwellings with a high anthropogenic nature [23]. Likewise, in Kenya and Tanzania, a shift in the resting behavior of *Anopheles gambiae* s.s. and *Anopheles funestus* was noted from endophily to exophily following the introduction of LLINs [24]. In the case of *An. stephensi*, the indoor resting density was higher than outdoors, suggesting a possible endophilic behavior as well.

*An. culicifacies*, a significant malaria vector in India, comprises a complex of 5 sibling species. It is estimated that up to about 70% of malaria transmission in India is facilitated by *An. culicifacies* [3]. Thus, the implementation of effective insecticide-based vector control interventions is crucial for malaria elimination efforts. *An. culicifacies* sibling species exhibit distinct distributions, with varying proportions of each species across different regions of the country [3], a pattern observed in this study as well. Understanding the distribution and proportions of *An. culicifacies* sibling species is crucial, as species A, C, D, and E are vectors, whereas species B is considered a poor or nonvector [25-28]. Moreover, these species vary in their susceptibility to different insecticides [28-30]. The study results indicated that species A predominated in Nuh, Palwal, and Kalaburagi districts, whereas species C was predominant in Sidhi and Gadchiroli districts. Species B was found in very low densities in Haryana, while in Gujarat and Karnataka states, its density was higher. This study offers a comprehensive overview of the contemporary distribution of sibling species in various eco-epidemiological settings across the country. Previously, the distribution pattern showed the ubiquitous presence of species B wherever *An. culicifacies* was encountered. Species A predominated in the northern part of the country, whereas species C was predominant in the western and eastern regions. In addition, species D was found in sympatric association with species A and B in the northwestern region [31].

Resistance to DDT in *An. culicifacies* was detected in all study sites, indicating widespread resistance throughout the country. Moreover, this species has developed resistance to malathion and deltamethrin in most study districts. Deltamethrin resistance in *An. culicifacies* was previously reported in the Gadchiroli district of Maharashtra [32,33], Kalahandi, and 4 other districts in Odisha [34]. Furthermore, in Nuh, Haryana, possible resistance to deltamethrin was observed [35]. A multidistrict study conducted in 2009 on the susceptibility of *An. culicifacies* to DDT, malathion, and deltamethrin in multiple districts of Madhya Pradesh reported its resistance to deltamethrin in Mandla and Dindori districts. Possible resistance was observed in Balaghat, Betul, Chhindwara, Jhabua, Sidhi, and Shadol districts, while the species remained susceptible in Guna district [36]. The national vector control program in India primarily depends on insecticide-based IRS and LLINs. Consequently, the development of resistance in malaria vectors to synthetic pyrethroids is of significant concern because of their widespread use in IRS and LLINs.

In the 1950s, DDT was used for IRS to control vectors. However, as a result of selection pressure, *An. culicifacies* was found to be resistant to this insecticide in 1959 [37]. Subsequently, malathion was introduced for vector control in 1969. However, resistance to malathion was reported in Gujarat as early as 1973 [38]. Further reports of resistance to malathion emerged from other states such as Andhra Pradesh, Odisha, Madhya Pradesh, Chhattisgarh, and Jharkhand [39]. In 1996, synthetic pyrethroids were introduced for IRS in India owing to their safety, excito-repellent, and knockdown properties. Just 6 years after their initial use, the first report of deltamethrin resistance in *An. culicifacies* emerged in Gujarat in 2002 [40]. Subsequently, triple resistance (to DDT, malathion, and deltamethrin) in *An. culicifacies* was reported from 31 districts in India [41]. The findings of this study align with these earlier reports.

*An. stephensi*, another malaria vector, was found to be resistant to deltamethrin, DDT, and malathion in the Nuh and Palwal districts of Haryana. Currently, IRS is not used as a vector control strategy for *An. stephensi* in India, except in Rajasthan, where it has been identified as the primary vector of malaria [41]. However, this species has been reported to be double resistant, specifically to DDT and malathion in 7 districts, and to malathion and deltamethrin in 1 district [41]. In our study, possibly for the first time in the country, triple resistance (to DDT, malathion, and deltamethrin) in *An. stephensi* was reported.

*An. fluviatilis* demonstrated susceptibility to malathion and synthetic pyrethroids, but exhibited possible resistance to DDT.
in the Kalahandi and Angul districts of Odisha. This vector species remained susceptible to deltamethrin and malathion in most parts of the country, except in Gadchiroli (Maharashtra), where possible resistance was reported [33]. In the Mayurbhanj district of Odisha, possible resistance to malathion was reported in *An. fluviatilis*. It was found to be susceptible to DDT in Odisha but resistant in Jharkhand and Chhattisgarh [41,42]. In this study, *An. minimus* and *An. baimaii*, the 2 primary malaria vectors in the northeast, were found to be susceptible to DDT and malathion. Another study by Raghavendra et al [41] reported that these 2 vector species were susceptible to DDT, malathion, and deltamethrin. However, in the eastern state of Odisha, there was only 1 report of resistance in *An. minimus* to DDT.

The WHO has proposed 3 main strategies for insecticide resistance management in malaria vectors: rotation, mosaics, and mixtures [43]. The primary objective of insecticide resistance management is to prevent the emergence of resistance in susceptible populations, delay the evolution of insecticide resistance, or reverse it to a level that allows for the effective use of insecticides for vector control [44]. Information regarding insecticide resistance among malaria vectors and the underlying mechanisms of resistance in various ecological settings is essential for formulating rational strategies for insecticide use and resistance management [45]. Therefore, further research studies on the frequency and intensity of insecticide resistance, as well as its underlying mechanisms among malaria vectors, are crucial to suggest an appropriate resistance management strategy. Using synergist piperonyl butoxide–based LLINs effectively reduced populations of insecticide-resistant (synthetic pyrethroid) vectors [46]. The ICMR-Vector Control Research Center assessed the efficacy of a piperonyl butoxide-αphacypermethrin–incorporated LLIN in experimental huts in Odisha and found it to be superior to the reference net in terms of mosquito mortality in the huts and cone bioassays [47]. Another potential method for managing insecticide-resistance populations could involve the use of attractive toxic sugar baits in combination with LLINs, although insecticide-resistant populations could involve the use of attractive toxic sugar baits in combination with LLINs, although this approach is still under development and evaluation [48-50]. These proposed measures hold significant epidemiological importance for India.

We screened over 8000 anopheline mosquitoes for *Plasmodium* parasite infection, but only 2 species, *An. culicifacies* and *An. fluviatilis*, tested positive. An earlier study conducted in Nuh in 2015-2016 reported a sporozoite positivity rate of 0.26% [35], which is lower than the rate reported in this study (2.1%). The observed variability in the sporozoite rate could be attributed to differences in sample size, usage of LLINs, and malaria endemicity of the study sites. Moreover, the high positivity rate of *An. culicifacies* in Nuh and Palwal districts might be due to the predominant proportion of sibling species A (>99%) in the area. This sibling species is recognized as an established malaria vector in the malaria-endemic regions of India [35]. In a previous study conducted in 8 southern districts of Odisha, sporozoite-positive *An. culicifacies* was reported only from Kandhamal district (with a 1.5% infectivity rate), and none of the mosquito pools tested positive in the rest of the districts, including Kalahandi [51]. Regarding *An. stephensi*, none were found positive for *Plasmodium* infection in Nuh and Palwal districts. This observation might be due to the abundance of *An. stephensi* mysorensis form, which, being primarily zoophagic in nature, is not considered a competent vector [3]. An observed HBI of 0-0.01 in this study confirms the zoophilic behavior of *An. stephensi* mysorensis in the area.

**Limitations**

This study conducted nationwide surveillance of malaria vectors, successfully collecting all primary vectors of malaria in India except *An. sundaicus*, which is the only vector in Andaman and Nicobar Islands, India. This represents a major limitation of our study. Another limitation is the unavailability of data on density bioassays for determining the level of insecticide resistance among the main malaria vectors. The results of intensity assays could inform decisions on whether to continue existing insecticide vector control measures or to change insecticides. The third limitation of the study was the lack of information on mosquito larval breeding habitats. This information could provide a more comprehensive understanding of the anopheline breeding habitats present in the selected study sites. The final limitation of the study was the lack of data on the parity status of vector mosquitoes, which is a robust indicator of mosquito age.

**Conclusions**

This study provides a comprehensive overview of essential aspects of vector bionomics, including seasonal prevalence, resting and biting behavior, insecticide resistance status, composition of sibling species, and malaria transmission potential. It clearly indicates a growing development of resistance in *An. culicifacies* against synthetic pyrethroids in the country. These findings underscore the importance of continuous monitoring of insecticide resistance for effective planning, implementation, and evaluation of malaria vector control strategies. Furthermore, because sibling species can vary in the rate of development of insecticide resistance, monitoring should be conducted at the sibling species level rather than at the sensu lato level. The observed shift from endophilic to exophilic behavior in parts of India among *An. culicifacies* highlights the necessity for continuous monitoring of such behavioral changes in vector species, especially in light of the extensive use of LLINs in the country. The localized and focal nature of malaria transmission is influenced by variations in the biological characteristics of vector species, cultural aspects of human populations, and environmental factors within a region. Therefore, further research on vector behavior is crucial to corroborate specific findings that have implications for malaria transmission and to strengthen control measures.

**Acknowledgments**

We thank the Indian Council of Medical Research, New Delhi for their financial support in carrying out the study.
Data Availability
The data sets generated during or analyzed in this study may be made available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References
2. Malaria Situation in India. National Center for Vector Borne Diseases Control. URL: https://nvd.cbdac.gov.in/WriteReadData/1892s/70838173921597401184.pdf [accessed 2021-10-03]


Abbreviations

DDT: dichlorodiphenyltrichloroethane
HBI: human blood index
ICMR: Indian Council of Medical Research
IRS: indoor residual spraying
LLIN: long-lasting insecticidal nets
MHD: man-hour density
PCR: polymerase chain reaction
WHO: World Health Organization

©Manju Rahi, AK Mishra, Gyan Chand, RK Baharia, RK Hazara, SP Singh, Siraj Khan, U Sreehari, Divya Kamaraju, Gaurav Kumar, Sanjeev Kumar Gupta, Amit Sharma, K Raghavendra, K Gunasekaran, Om P Singh, Sarala K Subbarao. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Spatiotemporal Characteristics and Risk Factors for All and Severity-Specific Preterm Births in Southern China, 2014-2021: Large Population-Based Study

Huazhang Miao1*, PhD; Hui He2*, MD; Chuan Nie3, PhD; Jianbing Ren3, PhD; Xianqiong Luo2, PhD

1Department of Healthcare, Guangdong Women and Children Hospital, Guangzhou, China
2Department of Pediatric Endocrinology and Inherited Metabolic Diseases, Guangdong Women and Children Hospital, Guangzhou, China
3Neonatology Department, Guangdong Women and Children Hospital, Guangzhou, China

*these authors contributed equally

Corresponding Author:
Xianqiong Luo, PhD
Department of Pediatric Endocrinology and Inherited Metabolic Diseases
Guangdong Women and Children Hospital
No. 521, Xingnan Road
Panyu District
Guangzhou, 510623
China
Phone: 86 39151822
Email: luoxqgz@126.com

Abstract

Background: The worldwide incidence of preterm births is increasing, and the risks of adverse outcomes for preterm infants significantly increase with shorter gestation, resulting in a substantial socioeconomic burden. Limited epidemiological studies have been conducted in China regarding the incidence and spatiotemporal trends of preterm births. Seasonal variations in risk indicate the presence of possible modifiable factors. Gender influences the risk of preterm birth.

Objective: This study aims to assess the incidence rates of preterm birth, very preterm birth, and extremely preterm birth; elucidate their spatiotemporal distribution; and investigate the risk factors associated with preterm birth.

Methods: We obtained data from the Guangdong Provincial Maternal and Child Health Information System, spanning from January 1, 2014, to December 31, 2021, pertaining to neonates with gestational ages ranging from 24 weeks to 42 weeks. The primary outcome measures assessed variations in the rates of different preterm birth subtypes over the course of the study, such as by year, region, and season. Furthermore, we examined the relationship between preterm birth incidence and per capita gross domestic product (GDP), simultaneously analyzing the contributing risk factors.

Results: The analysis incorporated data from 13,256,743 live births. We identified 754,268 preterm infants and 12,502,475 full-term infants. The incidences of preterm birth, very preterm birth, and extremely preterm birth were 5.69 per 100 births, 4.46 per 1000 births, and 4.83 per 10,000 births, respectively. The overall incidence of preterm birth increased from 5.12% in 2014 to 6.38% in 2021. The incidence of extremely preterm birth increased from 4.10 per 10,000 births in 2014 to 8.09 per 10,000 births in 2021. There was a positive correlation between the incidence of preterm infants and GDP per capita. In more developed economic regions, the incidence of preterm births was higher. Furthermore, adjusted odds ratios revealed that advanced maternal age, multiple pregnancies, and male infants were associated with an increased risk of preterm birth, whereas childbirth in the autumn season was associated with a protective effect against preterm birth.

Conclusions: The incidence of preterm birth in southern China exhibited an upward trend, closely linked to enhancements in the care capabilities for high-risk pregnant women and critically ill newborns. With the recent relaxation of China's 3-child policy, coupled with a temporary surge in advanced maternal age and multiple pregnancies, the risk of preterm birth has risen. Consequently, there is a pressing need to augment public health investments aimed at mitigating the risk factors associated with preterm birth, thereby alleviating the socioeconomic burden it imposes.
Neonatal diseases constitute the primary cause of mortality in children younger than 5 years, comprising approximately 37.7% of such deaths [1]. Globally, preterm birth (PTB) stands as the most significant risk factor for neonatal morbidity and mortality [2,3], contributing to around 18% of all fatalities in children younger than 5 years and as much as 35% of all fatalities among neonates (aged <28 days) [4]. The United Nations Sustainable Development Goal (SDG) 3.2 for 2030 aims to eliminate preventable child mortality worldwide, targeting a reduction of neonatal mortality to below 12 per 1000 live births and under-5 mortality to below 25 per 1000 live births. Data indicate that, although the global under-5 mortality rate declined by almost one-half, decreasing from 71.2 per 1000 to 37.1 per 1000 between 2000 and 2019, the neonatal mortality rate exhibited a slower reduction, dropping from 28.0 per 1000 to 17.9 per 1000. Notably, 32% of countries are projected to fall short of attaining any of the SDG 3.2 targets by 2030 [1]. Hence, directing attention to health care systems in regions experiencing a high prevalence of PTBs, augmenting investments in public health and medical interventions to mitigate PTB, and enhancing perinatal and neonatal care in focal areas may contribute to diminishing child mortality, progressively aligning with the 2030 SDGs.

The World Health Organization (WHO) defines PTB as the delivery of newborns before 37 completed weeks of gestation. Over the past 2 decades, there has been a global increase in the incidence of PTB, resulting in a global PTB rate of 11% [5]. In 2022, data from the US National Vital Statistics indicated that the rate of PTB for singletons was 10.38% [6]. Twins constitute 20% of all preterm deliveries, with 60% and 10.7% of twins delivered before 37 and 32 weeks of gestation, respectively [7]. Twin pregnancies are associated with a higher risk of PTB than singleton pregnancies. A meta-analysis conducted in China revealed that the pooled rate of preterm delivery in hospital-based studies was 7.2% (95% CI 6.9%-7.6%), whereas the population-based rates were 5.3% for eastern, 4.6% for central, and 3.8% for western regions [8]. Six countries—India, China, Nigeria, Pakistan, Indonesia, and the United States—collectively contribute to 45.9% (approximately 7.4 million) of the total global PTBs, and it is projected that China will have a PTB rate of around 6.9% [2]. In a study conducted in the United States concerning mortality, in-hospital morbidity, and a 2-year follow-up of extremely preterm infants, 78.3% survived until discharge. Among newborns delivered at gestational ages below 27 weeks, rehospitalization and neurodevelopmental disorders were more prevalent at 2 years of age [9]. This underscores that PTB constitutes a significant health care challenge and a substantial socioeconomic burden [4].

Our study specifically investigated the spatiotemporal aspects and risk factors associated with PTB in 21 cities in southern China from 2014 to 2021. This research aimed to guide the development of public health and health care policies that address regional health equity and balanced development.
Ethical Considerations

This study received approval from the Ethics Committee of Guangdong Women and Children Hospital (approval number: 202201332). The study relied on routine registration data and did not involve patients’ participation in questionnaires, outcome assessments, and other research methods. Additionally, privacy information such as personal identity and contact details have been concealed. The requirement for informed consent was waived by the hospital’s ethics review committee.

Statistical Analysis

Maternal and neonatal characteristics are presented as a mean (SD) for continuous variables and number (%) for categorical variables. Independent samples Satterthwaite t tests were conducted to examine differences in the means of continuous variables. Prevalences of total PTBs (<37 completed weeks of gestation, per 100 births), VPTBs (<32 completed weeks of gestation, per 1000 births), and EPTBs (<28 completed weeks of gestation, per 10,000 births) according to the year, area of residence, region of Guangdong, season of delivery, maternal age, neonatal gender, and number of fetuses were calculated using a binomial distribution exact method. Chi-square tests were used to compare prevalences and obtain the P value.

Crude or adjusted associations between PTB and season of delivery, maternal age, neonatal gender, and number of fetuses were examined separately using a logistic regression model. The univariable model included PTB as the dependent variable and included only season of delivery, maternal age, neonatal gender, and number of fetuses as separate independent variables. The multivariable model included PTB as the dependent variable and adjusted for season of delivery, maternal age, neonatal gender, and number of fetuses.

Analyses were done using SAS 9.4 (SAS Institute) and R version 4.2.1 (R Core Team) for the figures. A 2-sided P<.05 was considered statistically significant.

Results

We conducted an analysis of 13,501,312 infants delivered between January 1, 2014, and December 31, 2021, in 21 cities in southern China and spanning gestational ages from 24 weeks to 45 weeks (Multimedia Appendix 1). Exclusions from the data analysis were for the following reasons: 17,794 cases (0.13%) involved non-Chinese pregnant women, 148,351 cases (1.10%) had non-Chinese fathers, 9812 cases (0.07%) were not live births, 56 cases (<0.01%) had an unspecified gender, 6017 cases (0.04%) had a maternal age outside the range of 15 years to 60 years or a missing maternal age, 56,047 (<0.41%) had a paternal age beyond the range of 15 years to 60 years or a missing paternal age, 333 cases (<0.01%) had a birth weight outside the range of 350 g to 6000 g, and 7159 cases (0.05%) had a gestational age outside the range of 16 weeks to 45 weeks. The data set included a total of 13,256,743 live births (Multimedia Appendix 2).

Table 1 provides an overview of the characteristics of the live-born infants included in the analysis. The conception rate varied significantly by season, with winter showing a relatively higher proportion, at approximately 32.68% (4,332,432/13,256,743), while autumn had a higher rate of childbirth, at 27.96% (3,706,674/13,256,743). Over the period spanning from 2014 to 2021, there were 754,268 preterm infants and 12,502,475 full-term infants. Among the PTBs, 419,314 (55.59%) occurred among Aboriginal inhabitants, and a significantly larger proportion occurred in urban areas (89.86%) compared with rural areas. Notably, within the PRD region, there were 472,468 PTBs, constituting a substantially higher percentage (472,468/754,268, 62.64%) than in the eastern, western, and northern regions of Guangdong.
Table 1. Comparison of maternal and neonatal characteristics between preterm births (<37 completed weeks of gestation; n=754,268) and term births (≥37 completed weeks of gestation; n=12,502,475) in southern China, 2014-2021 (N=13,256,743).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Preterm births</th>
<th>Term births</th>
<th>Total</th>
<th>Statistical test results (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Area of residence, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>677,772 (89.86)</td>
<td>10,724,873 (85.78)</td>
<td>11,402,645 (86.01)</td>
<td>9825.20 (1)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Rural</td>
<td>76,496 (10.14)</td>
<td>1,777,602 (14.22)</td>
<td>1,854,098 (13.99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Household registration, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local</td>
<td>419,314 (55.59)</td>
<td>7,406,819 (59.24)</td>
<td>7,826,133 (59.04)</td>
<td>3920.14 (1)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Nonlocal</td>
<td>334,954 (44.41)</td>
<td>5,095,656 (40.76)</td>
<td>5,430,610 (40.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Region of GD&lt;sup&gt;c&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRD&lt;sup&gt;d&lt;/sup&gt;</td>
<td>472,468 (62.64)</td>
<td>6,834,971 (54.67)</td>
<td>7,307,439 (55.12)</td>
<td>21,462.59 (3)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Eastern GD</td>
<td>85,806 (11.38)</td>
<td>1,982,067 (15.85)</td>
<td>2,067,873 (15.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western GD</td>
<td>97,305 (12.90)</td>
<td>1,983,107 (15.86)</td>
<td>2,080,412 (15.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern GD</td>
<td>98,689 (13.08)</td>
<td>1,702,330 (13.62)</td>
<td>1,801,019 (13.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Season of conception, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>189,427 (25.11)</td>
<td>2,848,584 (22.78)</td>
<td>3,038,011 (22.92)</td>
<td>2225.09 (3)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Summer</td>
<td>155,520 (20.62)</td>
<td>2,631,794 (21.05)</td>
<td>2,787,314 (21.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autumn</td>
<td>172,078 (22.81)</td>
<td>2,926,908 (23.41)</td>
<td>3,098,986 (23.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>237,243 (31.45)</td>
<td>4,095,189 (32.76)</td>
<td>4,332,432 (32.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Season of delivery, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>172,829 (22.91)</td>
<td>2,856,963 (22.85)</td>
<td>3,029,792 (22.85)</td>
<td>1492.84 (3)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Summer</td>
<td>192,966 (25.58)</td>
<td>3,144,972 (25.15)</td>
<td>3,337,938 (25.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autumn</td>
<td>197,651 (26.20)</td>
<td>3,509,023 (28.07)</td>
<td>3,706,674 (27.96)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter</td>
<td>190,822 (25.30)</td>
<td>2,991,517 (23.93)</td>
<td>3,182,339 (24.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal age (years), mean (SD)</td>
<td>29.18 (5.59)</td>
<td>28.05 (5.03)</td>
<td>28.12 (5.07)</td>
<td>170.44 (829,409)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Maternal age (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>622,431 (82.52)</td>
<td>11,128,746 (89.01)</td>
<td>11,751,177 (88.64)</td>
<td>29,772.81 (1)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≥35</td>
<td>131,837 (17.48)</td>
<td>1,373,729 (10.99)</td>
<td>1,505,566 (11.36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paternal age (years), mean (SD)</td>
<td>31.56 (6.14)</td>
<td>30.43 (5.58)</td>
<td>30.49 (5.62)</td>
<td>155.98 (831,145)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Paternal age (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characteristics</td>
<td>Preterm births</td>
<td>Term births</td>
<td>Total</td>
<td>Statistical test results (df)</td>
<td>P value</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>----------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>&lt;35</td>
<td>538,114 (71.34)</td>
<td>9,869,950 (78.94)</td>
<td>10,408,064 (78.51)</td>
<td>112,969.00 (2)⁷</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≥35</td>
<td>216,154 (28.66)</td>
<td>2,632,525 (21.06)</td>
<td>2,848,679 (21.49)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Method of delivery, n (%)**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaginal</td>
<td>348,215 (46.17)</td>
<td>7,822,312 (62.57)</td>
<td>8,170,527 (61.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cesarean section</td>
<td>321,930 (42.68)</td>
<td>3,166,736 (25.33)</td>
<td>3,488,666 (26.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>84,123 (11.15)</td>
<td>1,513,427 (12.11)</td>
<td>1,597,550 (12.05)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Neonatal gender, n (%)**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>436,768 (57.91)</td>
<td>6,641,625 (53.12)</td>
<td>7,078,393 (53.39)</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female</td>
<td>317,500 (42.09)</td>
<td>5,860,850 (46.88)</td>
<td>6,178,350 (46.61)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Number of previous childbirths, n (%)**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>373,655 (49.54)</td>
<td>6,410,764 (51.28)</td>
<td>6,784,419 (51.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>271,527 (36.00)</td>
<td>4,491,143 (35.92)</td>
<td>4,762,670 (35.93)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>109,086 (14.46)</td>
<td>1,600,568 (12.80)</td>
<td>1,709,654 (12.90)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Number of fetuses, n (%)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>&lt;.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>599,814 (79.52)</td>
<td>12,342,371 (98.72)</td>
<td>12,942,185 (97.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>152,342 (20.20)</td>
<td>159,039 (1.27)</td>
<td>311,381 (2.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>2112 (0.28)</td>
<td>1065 (0.01)</td>
<td>3117 (0.02)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Gestational age (weeks), mean (SD)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>&lt;.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;28</td>
<td>6399 (0.85)</td>
<td>0 (0.00)</td>
<td>6399 (0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28-32</td>
<td>84,797 (11.24)</td>
<td>0 (0.00)</td>
<td>84,797 (0.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33-36</td>
<td>663,072 (87.91)</td>
<td>0 (0.00)</td>
<td>663,072 (5.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37-41</td>
<td>12,424,849 (99.38)</td>
<td>12,424,849 (93.72)</td>
<td>12,424,849 (93.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥42</td>
<td>77,626 (0.62)</td>
<td>77,626 (0.59)</td>
<td>77,626 (0.59)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Birth weight (g), mean (SD)**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>&lt;.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1000</td>
<td>6170 (0.82)</td>
<td>108 (0.00)</td>
<td>6278 (0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000-1499</td>
<td>38,543 (5.11)</td>
<td>972 (0.01)</td>
<td>39,515 (0.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500-2499</td>
<td>362,945 (48.12)</td>
<td>316,810 (2.53)</td>
<td>679,755 (5.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2500-3999</td>
<td>345,227 (45.77)</td>
<td>11,775,742 (94.19)</td>
<td>12,120,969 (91.45)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Additionally, Table 1 presents data on the average maternal age, which was 28.12 (SD 5.07) years, with an 11.36% (1,505,566/13,256,743) rate of advanced maternal age (≥35 years old), and the average paternal age, which was 30.49 (SD 5.62) years, with a 21.49% (2,848,679/13,256,743) rate of advanced paternal age (≥35 years old). In their first birth, 373,655 women delivered a preterm baby. Among PTB, 79.52% (599,814/754,268) were singleton pregnancies. The average gestational age of infants was 38.76 (SD 1.53) weeks, with an average birth weight of 3153.76 (SD 448.06) g. In contrast, preterm infants had an average gestational age of 34.66 (SD 1.92) weeks and an average birth weight of 2381.40 (SD 519.20) g.

Figure 1 illustrates the temporal trend in the 3 distinct categories of preterm infants from 2014 to 2021. PTB, VPTB, and EPTB increased year on year. PTB increased from 5.12 (per 100 births) in 2014 to 6.38 (per 100 births) in 2021, VPTB rose from 3.93 (per 1000 births) in 2014 to 4.46 (per 1000 births) in 2021, and EPTB surged from 4.10 (per 10,000 births) in 2014 to 8.09 (per 10,000 births) in 2021. Notably, a marked reduction in the incidence of preterm infants was evident at the start of the subsequent year, with the rates of PTB and VPTB experiencing a sharp increase in 2016 followed by a decline in 2017.

Figure 2 and Multimedia Appendix 3 spatially delineate the 3 distinct categories of preterm infants. Generally, PTB rates across all types are higher in the PRD region, with Guangzhou, as the provincial capital, having the highest prevalence. It is noteworthy that, in the case of EPTB, the cities with higher quantities of PTBs are not predominantly concentrated in the PRD region. Additionally, there are 1 to 2 cities in eastern, western, and northern Guangdong with high EPTB rates.
Figure 2. In southern China, 2014-2021: (A) spatial distributions of the prevalence of preterm births (<37 completed weeks of gestation, per 100 births); (B) spatial distributions of the prevalence of very preterm births (<32 completed weeks of gestation, per 1000 births); (C) spatial distributions of the prevalence of extremely preterm births (<28 completed weeks of gestation, per 10,000 births).

Figure 3 provides a detailed description of the incidence and temporal trends of preterm infants in 21 cities within southern China. Irrespective of the type of PTB, the rate was lower in eastern Guangdong than in the other regions. For both VPTB and EPTB, the rates in Guangzhou and Shenzhen were notably high, with the incidence of EPTB in Shenzhen continuously increasing relative to Guangzhou since 2015.

Figure 3. Across 21 cities in southern China, 2014-2021: heat maps of the yearly prevalences of (A) preterm births (<37 completed weeks of gestation, per 100 births); (B) very preterm births (<32 completed weeks of gestation, per 1000 births); (C) extremely preterm births (<28 completed weeks of gestation, per 10,000 births).

Table 2 provides a comparative analysis of the incidences of various types of preterm infants. Over the period of 2014 to 2021, the incidence rates for PTB, VPTB, and EPTB were 5.69 (per 100 births), 4.46 (per 1000 births), and 4.83 (per 10,000 births), respectively. All 3 types of preterm infants occurred more frequently in urban areas than in rural regions. It is noteworthy that the 3 types of preterm infants have the lowest birth rates in the autumn season. Additionally, advanced maternal age ($\geq 35$ years) and multiple pregnancies posed a higher risk for PTB, regardless of PTB, VPTB, or EPTB. Furthermore, PTB is significantly more common in boys than in girls.
### Table 2. Prevalence of total preterm births (<37 completed weeks of gestation, per 100 births), very preterm births (<32 completed weeks of gestation, per 1000 births), and extremely preterm births (<28 completed weeks of gestation, per 10,000 births) in southern China, 2014-2021.

<table>
<thead>
<tr>
<th>Subgroups</th>
<th>Births, n</th>
<th>Total preterm births</th>
<th>Very preterm births</th>
<th>Extremely preterm births</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Results, n</td>
<td>Prevalence (95% CI)</td>
<td>Results, n</td>
<td>Prevalence (95% CI)</td>
<td>Results, n</td>
<td>Prevalence (95% CI)</td>
</tr>
<tr>
<td><strong>Years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>1,816,688</td>
<td>92,985</td>
<td>5.12 (5.09-5.15)</td>
<td>7143</td>
<td>3.93 (3.84-4.02)</td>
<td>745</td>
<td>4.10 (3.81-4.41)</td>
</tr>
<tr>
<td>2015</td>
<td>1,662,742</td>
<td>86,386</td>
<td>5.20 (5.16-5.23)</td>
<td>6441</td>
<td>3.87 (3.78-3.97)</td>
<td>497</td>
<td>2.99 (2.73-3.26)</td>
</tr>
<tr>
<td>2016</td>
<td>1,876,740</td>
<td>107,738</td>
<td>5.74 (5.71-5.77)</td>
<td>8098</td>
<td>4.31 (4.22-4.41)</td>
<td>695</td>
<td>3.70 (3.43-3.99)</td>
</tr>
<tr>
<td>2017</td>
<td>1,906,491</td>
<td>99,665</td>
<td>5.23 (5.20-5.26)</td>
<td>7230</td>
<td>3.79 (3.71-3.88)</td>
<td>713</td>
<td>3.74 (3.47-4.02)</td>
</tr>
<tr>
<td>2018</td>
<td>1,699,744</td>
<td>100,617</td>
<td>5.92 (5.88-5.96)</td>
<td>7939</td>
<td>4.67 (4.57-4.77)</td>
<td>901</td>
<td>5.30 (4.96-5.66)</td>
</tr>
<tr>
<td>2019</td>
<td>1,637,556</td>
<td>98,089</td>
<td>5.99 (5.95-6.03)</td>
<td>8028</td>
<td>4.90 (4.80-5.01)</td>
<td>934</td>
<td>5.70 (5.34-6.08)</td>
</tr>
<tr>
<td>2020</td>
<td>1,409,671</td>
<td>89,182</td>
<td>6.33 (6.29-6.37)</td>
<td>7238</td>
<td>5.13 (5.02-5.25)</td>
<td>901</td>
<td>6.42 (6.01-6.85)</td>
</tr>
<tr>
<td>2021</td>
<td>1,247,111</td>
<td>79,606</td>
<td>6.38 (6.34-6.43)</td>
<td>6948</td>
<td>5.57 (5.44-5.70)</td>
<td>1009</td>
<td>8.09 (7.6-8.61)</td>
</tr>
<tr>
<td>2014-2021</td>
<td>13,256,743</td>
<td>754,268</td>
<td>5.69 (5.68-5.70)</td>
<td>59,065</td>
<td>4.46 (4.42-4.49)</td>
<td>6399</td>
<td>4.83 (4.71-4.95)</td>
</tr>
<tr>
<td><strong>Area of residence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>11,402,645</td>
<td>677,772</td>
<td>5.94 (5.93-5.96)</td>
<td>55,539</td>
<td>4.87 (4.83-4.91)</td>
<td>6112</td>
<td>5.36 (5.23-5.50)</td>
</tr>
<tr>
<td>Rural</td>
<td>1,854,098</td>
<td>76,496</td>
<td>4.13 (4.10-4.15)</td>
<td>3526</td>
<td>1.90 (1.84-1.97)</td>
<td>287</td>
<td>1.55 (1.37-1.74)</td>
</tr>
<tr>
<td><strong>Region of GD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern GD</td>
<td>2,067,873</td>
<td>85,806</td>
<td>4.15 (4.12-4.18)</td>
<td>6476</td>
<td>3.13 (3.06-3.21)</td>
<td>527</td>
<td>2.55 (2.34-2.78)</td>
</tr>
<tr>
<td>Western GD</td>
<td>2,080,412</td>
<td>97,305</td>
<td>4.68 (4.65-4.71)</td>
<td>6702</td>
<td>3.22 (3.14-3.30)</td>
<td>612</td>
<td>2.94 (2.71-3.18)</td>
</tr>
<tr>
<td>Northern GD</td>
<td>1,801,019</td>
<td>98,689</td>
<td>5.48 (5.45-5.51)</td>
<td>6181</td>
<td>3.43 (3.35-3.52)</td>
<td>555</td>
<td>3.08 (2.83-3.35)</td>
</tr>
<tr>
<td><strong>Season of delivery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>3,029,792</td>
<td>172,829</td>
<td>5.70 (5.68-5.73)</td>
<td>14,163</td>
<td>4.67 (4.60-4.75)</td>
<td>1668</td>
<td>5.51 (5.24-5.78)</td>
</tr>
<tr>
<td>Summer</td>
<td>3,337,938</td>
<td>192,966</td>
<td>5.78 (5.76-5.81)</td>
<td>16,090</td>
<td>4.82 (4.75-4.90)</td>
<td>1644</td>
<td>4.93 (4.69-5.17)</td>
</tr>
<tr>
<td>Autumn</td>
<td>3,706,674</td>
<td>197,651</td>
<td>5.33 (5.31-5.36)</td>
<td>15,208</td>
<td>4.10 (4.04-4.17)</td>
<td>1559</td>
<td>4.21 (4.00-4.42)</td>
</tr>
<tr>
<td>Winter</td>
<td>3,182,339</td>
<td>190,822</td>
<td>6.00 (5.97-6.02)</td>
<td>13,604</td>
<td>4.27 (4.20-4.35)</td>
<td>1528</td>
<td>4.80 (4.56-5.05)</td>
</tr>
<tr>
<td><strong>Maternal age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>11,751,177</td>
<td>622,431</td>
<td>5.30 (5.28-5.31)</td>
<td>47,353</td>
<td>4.03 (3.99-4.07)</td>
<td>4891</td>
<td>4.16 (4.05-4.28)</td>
</tr>
<tr>
<td>≥35</td>
<td>1,505,566</td>
<td>131,837</td>
<td>8.76 (8.71-8.80)</td>
<td>11,712</td>
<td>7.78 (7.64-7.92)</td>
<td>1508</td>
<td>10.02 (9.52-10.53)</td>
</tr>
<tr>
<td><strong>Neonatal gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7,078,393</td>
<td>436,768</td>
<td>6.17 (6.15-6.19)</td>
<td>34,985</td>
<td>4.94 (4.89-4.99)</td>
<td>3959</td>
<td>5.59 (5.42-5.77)</td>
</tr>
<tr>
<td>Female</td>
<td>6,178,350</td>
<td>317,500</td>
<td>5.14 (5.12-5.16)</td>
<td>24,080</td>
<td>3.90 (3.85-3.95)</td>
<td>2440</td>
<td>3.95 (3.79-4.11)</td>
</tr>
<tr>
<td><strong>Number of fetuses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>12,942,185</td>
<td>599,814</td>
<td>4.63 (4.62-4.65)</td>
<td>45,393</td>
<td>3.51 (3.48-3.54)</td>
<td>4611</td>
<td>3.56 (3.46-3.67)</td>
</tr>
<tr>
<td>2</td>
<td>311,381</td>
<td>152,342</td>
<td>48.92 (48.75-49.10)</td>
<td>13,228</td>
<td>42.48 (41.78-43.20)</td>
<td>1723</td>
<td>55.33 (52.76-58.00)</td>
</tr>
<tr>
<td>≥3</td>
<td>3177</td>
<td>2112</td>
<td>66.48 (64.81-68.12)</td>
<td>444</td>
<td>139.75 (127.88-152.30)</td>
<td>65</td>
<td>204.6 (158.25-260.04)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Significant difference compared to the previous year.
Chi-square tests were used to compare the prevalences for all of these subgroups.

GD: Guangdong.

PRD: Pearl River Delta.

Includes March, April, and May.

Includes June, July, and August.

Includes September, October, and November.

Includes December, January, and February.

Figure 4 illustrates the association between the incidence of the 3 distinct types of preterm infants and per capita gross domestic product (GDP). As the economy developed, there was an increase in the incidences of PTB, VPTB, and EPTB. Notably, the positive correlation between VPTB and GDP per capita was the most pronounced (r=0.64).

Figure 4. Across 21 cities in southern China, 2014-2021: correlations between per capita gross domestic product (GDP) and (A) preterm births (PTB; <37 completed weeks of gestation, per 100 births); (B) very preterm births (VPTB; <32 completed weeks of gestation, per 1000 births); (C) extremely preterm births (EPTB; <28 completed weeks of gestation, per 10,000 births).

Table 3 examines the risk factors associated with the 3 categories of preterm infants. The adjusted odds ratios (ORs) revealed that advanced maternal age (adjusted OR 1.59, 95% CI 1.58-1.60), multiple pregnancies (adjusted OR 40.52, 95% CI 37.63-43.64), and male infants (adjusted OR 1.24, 95% CI 1.23-1.25) were risk factors for PTB. In contrast, giving birth in the autumn season was found to be a protective factor against PTB.
enhancing the equity of maternal and child health services and disparities in health care but also points toward directions for by a more developed economy. This not only highlights regional higher rate of PTB in the PRD region, which is characterized of premature infants. The findings indicated a classification of premature infants. The findings indicated a in China. Additionally, our research presents a comprehensive data set available for southern China. These cities, which are 8 years, using the largest, most recent, and most comprehensive reviewed the epidemiological trends and risk factors associated the survival rate among PTBs. In this study, we analyzed and systems have played significant roles in the gradual increase of maternal and child health service network and 2-way referral As medical technology continues to advance, the Chinese 4-level Principal Findings

Discussion

Table 3. Risk factors for total preterm births (<37 completed weeks of gestation), very preterm births (<32 completed weeks of gestation), and extremely preterm births (<28 completed weeks of gestation) in southern China, 2014-2021.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Preterm births</th>
<th>Very preterm births</th>
<th>Extremely preterm births</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Crude OR (95% CI)</td>
<td>P value</td>
<td>Adjusted OR (95% CI)</td>
</tr>
<tr>
<td>Season of delivery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>1.07 (1.07-1.08)</td>
<td>&lt;.001</td>
<td>1.04 (1.03-1.05)</td>
</tr>
<tr>
<td>Summer</td>
<td>1.09 (1.08-1.10)</td>
<td>&lt;.001</td>
<td>1.06 (1.05-1.07)</td>
</tr>
<tr>
<td>Autumn</td>
<td>Reference</td>
<td>—</td>
<td>Reference</td>
</tr>
<tr>
<td>Winter</td>
<td>1.13 (1.13-1.14)</td>
<td>&lt;.001</td>
<td>1.11 (1.10-1.12)</td>
</tr>
<tr>
<td>Maternal age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>Reference</td>
<td>—</td>
<td>Reference</td>
</tr>
<tr>
<td>≥35</td>
<td>1.72 (1.71-1.73)</td>
<td>&lt;.001</td>
<td>1.59 (1.58-1.60)</td>
</tr>
<tr>
<td>Neonatal gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.21 (1.21-1.22)</td>
<td>&lt;.001</td>
<td>1.24 (1.23-1.25)</td>
</tr>
<tr>
<td>Female</td>
<td>Reference</td>
<td>—</td>
<td>Reference</td>
</tr>
<tr>
<td>Number of fetuses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Reference</td>
<td>—</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>19.71 (19.56-18.86)</td>
<td>&lt;.001</td>
<td>19.39 (19.25-19.54)</td>
</tr>
<tr>
<td>≥3</td>
<td>40.61 (37.72-43.71)</td>
<td>&lt;.001</td>
<td>40.52 (37.63-43.64)</td>
</tr>
</tbody>
</table>

aOR: odds ratio.
bMultivariate adjustment for season of delivery, maternal age, neonatal sex, and number of fetuses.
cNot applicable.

Discussion

Principal Findings

As medical technology continues to advance, the Chinese 4-level maternal and child health service network and 2-way referral systems have played significant roles in the gradual increase of the survival rate among PTBs. In this study, we analyzed and reviewed the epidemiological trends and risk factors associated with preterm infants in 21 cities in southern China over the past 8 years, using the largest, most recent, and most comprehensive data set available for southern China. These cities, which are densely populated with a significant migrant population, offer a representative sample for assessing premature birth incidence in China. Additionally, our research presents a comprehensive classification of premature infants. The findings indicated a higher rate of PTB in the PRD region, which is characterized by a more developed economy. This not only highlights regional disparities in health care but also points toward directions for enhancing the equity of maternal and child health services and proposing prepregnancy health care guidance. These efforts aim to promote a healthy development model for women and children in the new era in China.

In general, the PTB rate shows a positive correlation with GDP levels, a relationship deeply intertwined with the specific conditions and the maternal and child health system in China. Roughly 15 million preterm infants, representing an adverse pregnancy outcome, are born globally each year. It is worth noting that the rate is as low as 5% even among healthy pregnant women at low risk during pregnancy [13]. Hence, the vast majority of PTBs are linked to multiple pregnancies, pregnancy complications, placental abnormalities, and other contributing factors [14]. Our results support that advanced maternal age and multiple pregnancies are risk factors for the incidence of PTB. Using the maternal and child health service network and referral system in China, the majority of high-risk pregnant women are referred to tertiary hospitals. This results in a scenario in which a significant proportion of preterm infants are born in urban or economically developed regions. Additionally, it is evident that the population of permanent
residents in urban areas or more economically developed areas is substantially higher than in less economically developed areas, contributing to the elevated PTB rates in these regions. An investigation spanning from 1990 to 2019 that focused on the global, regional, and national incidences of and mortality associated with neonatal PTB revealed a noteworthy positive correlation between the prevalence of neonatal PTB and the sociodemographic index in 2019, as well as the universal health coverage index in 2019. This correlation was particularly pronounced in high-income countries, such as Greece, Bahrain, Japan, the United Kingdom, and the United States, which corroborates our findings [15].

In October 2015, the Chinese government initiated the universal 2-child policy, which officially took effect in July 2016. The alteration in fertility policy had the potential to result in shifts in the sociodemographic characteristics of pregnant women, including a transient increase in advanced maternal age (≥35 years) [16]. Consequently, the rates of PTBs in 2016 and 2017 notably increased compared with previous years, particularly in the case of EPTB. Despite intensified public health and medical intervention efforts aimed at reducing PTB, the rate has increased in the majority of countries. For instance, PTB rates in the United States increased from 9.98% in 2010 to 10.09% in 2020 [17], while the rate of EPTB decreased from 0.71% to 0.64%. Since 1990, worldwide neonatal mortality has decreased by 37%, and the reduction in the incidence of EPTB has played a role in further lowering neonatal mortality rates [18]. According to survey data from the National Maternal and Child Health Center in China, the incidence of preterm infants in China increased from 5.9% in 2012 to 6.4% in 2018, with an overall rate of approximately 6.1% [16]. The PTB rate in South China also exhibited an upward trend, with the rate of EPTB significantly lower than that in the United States.

The province housing 21 cities in southern China ranks as China’s most densely populated region, boasting a preterm rate significantly below the global average. This achievement can be attributed to the establishment of a robust maternal and child health system and the successful implementation of the national health care reform plan. Nonetheless, there exists a glaring issue of regional health care inequality, exemplified by the concentration of PTBs in the PRD region. Apart from the PRD region, 1 to 2 prominent cities in other regions also had high PTB rates. This suggests that the standard of care for newborns facing severe complication is consistently improving in these leading cities, while health care in other cities remains underdeveloped. Hence, regional management, augmented investment in public health initiatives within these leading cities, and the gradual establishment of a core medical hub centered around these cities represent an appropriate strategy. Ultimately, this approach aims to catalyze the advancement of health care services in the surrounding cities, aims to alleviate the health care burden on provincial capital cities and the PRD areas, and holds significant implications for the promotion of child health in China.

In our study, we observed that season had a notable impact on PTB. Our observation revealed that, although the number of births was higher during autumn than in other seasons, the incidence of PTBs was at its lowest during that time of the year. Simultaneously, our analysis of risk factors suggested that giving birth in autumn (September to November) was associated with a protective effect against PTB. In general, there were 2 peaks in the incidence of PTBs during summer (June 1 to August 31) and winter (December 1 to February 30). These findings align with a Norwegian survey that explored the seasonal impact on PTBs by analyzing data from 2,321,652 birth registrations. When analyzing the underlying reasons for this pattern, it is possible that an increase in unplanned pregnancies during longer summer and winter holidays is a contributing factor. Moreover, individuals with unplanned pregnancies may be less likely to engage in preconception preparation [19].

Specifically, the incidence of PTBs exhibited substantial annual fluctuations, with a notable decrease near the onset of the subsequent year. This phenomenon may be attributed to the subtropical monsoon climate characteristic of southern China, which implies that December has relatively mild temperatures, and such variations in temperature may influence the occurrence of PTBs. Liang et al [20] conducted an analysis of the impact of the 2008 cold spell on PTBs in 2 cities in South China. Regarding the total number of vaginal PTBs, their study revealed that there was a 22.44% increase in Dongguan and a 21.25% increase in Shenzhen during the 2008 cold spell, thereby supporting the influence of a cold spell on PTB. Additionally, a global analysis encompassing 14 lower- to middle-income countries investigated the connection between extreme heat and the occurrence of PTB and stillbirth. The results indicated that higher maximum temperatures and a narrower diurnal temperature range experienced during the last week before birth were associated with an increased risk of PTB and stillbirth [21]. This finding reinforces our findings.

Furthermore, data from 13 studies conducted in the United States and Europe revealed that 54.6% of PTBs were male infants, with a risk ratio of 1.14 (95% CI 1.11-1.17) for PTBs of male infants compared with female infants, aligning with our findings [22]. If the risk ratio for PTB between genders remains consistent within each region and is applied to sex-specific live birth data to estimate sex-specific PTB prevalence, it can offer guidance for future public health investments.

In this study, we investigated the prevalence and risk factors associated with preterm infants in southern China, approaching the analysis both from a spatiotemporal perspective and using a substantial data set. Our findings suggest that, despite a year-on-year increase in the incidence of PTBs, it remains below the global average, which bodes well for the advancement of public health initiatives in China. However, this study has certain limitations. Although the sample size was large, we lacked information on maternal pregnancy complications and general health status, preventing a more in-depth analysis of risk factors and the elimination of potential confounding variables. Additionally, the causal mechanism behind the seasonal influence on PTBs, potentially associated with temperature variations, remains unexplored. In the future, further investigations involving extensive sampling and questionnaire surveys are essential to delve deeply into the impact of climatic factors on PTBs, ultimately contributing to the development of theories for prepregnancy health care.
Conclusions
The incidence of PTB in South China exhibited an upward trend, closely linked to enhancements in the care capabilities for high-risk pregnant women and critically ill newborns. With the recent relaxation of China’s 3-child policy, coupled with a temporary surge in advanced maternal age and multiple pregnancies, the risk of PTB has risen. Consequently, there is a pressing need to augment public health investments aimed at mitigating the risk factors associated with PTB, thereby alleviating the socioeconomic burden it imposes.

Data Availability
All data generated or analyzed during this study are included in this published article and its supplementary information.

Authors' Contributions
XL obtained funding for the study. HM, HH, and XL conceptualized and designed the study. XL, CN, and JR provided administrative, technical, or material support and supervised the study. All authors contributed to the acquisition, analysis, or interpretation of the data. HM, HH, and XL had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. HH and HM performed the statistical analysis and drafted the manuscript, and all authors critically revised the manuscript for important intellectual content.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Geographical locations of the study area. Guangdong province is located in South China. This study included all 21 cities in Guangdong province, which are Guangzhou, Shenzhen, Foshan, Dongguan, Zhuhai, Zhongshan, Zhaoqing, Huizhou, Jiangmen, Chaozhou, Shantou, Jieyang, Meizhou, Shanwei, Shaoguan, Heyuan, Qingyuan, Yunfu, Yangjiang, Maoming, and Zhanjiang.

Multimedia Appendix 2
Study flowchart for selection of newborns from birth records in South China, 2014-2021.

Multimedia Appendix 3
Disease distribution maps at sub-county levels: (A) spatial distributions of the prevalence of preterm births (PTB; <37 completed weeks of gestation, per 100 births) in southern China, 2014–2021; (B) spatial distributions of the prevalence of very preterm births (VPTB; <32 completed weeks of gestation, per 1000 births) in southern China, 2014–2021; (C) spatial distributions of the prevalence of extremely preterm births (EPTB; <28 completed weeks of gestation, per 10,000 births) in southern China, 2014-2021.

References


18. Lawn JE, Blencowe H, Oza S, You D, Lee ACC, Waiswa P. Lancet Newborn Study Group. Every Newborn: progress, priorities, and potential beyond survival. Lancet 2014 Jul 12;384(9938):189-205. [doi: 10.1016/S0140-6736(14)60496-7] [Medline: 24853593]


Abbreviations

- EPTB: extremely preterm birth
- GDP: gross domestic product
- OR: odds ratio
- PRD: Pearl River Delta
- PTB: preterm birth
- SDG: Sustainable Development Goal
- VPTB: very preterm birth
- WHO: World Health Organization
Miao et al

©Huazhang Miao, Hui He, Chuan Nie, Jianbing Ren, Xianqiong Luo. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 18.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Prevalence, Mortality, and Access to Care for Chronic Kidney Disease in Medicaid-Enrolled Adults With Sickle Cell Disease in California: Retrospective Cohort Study

Jhaqueline Valle¹, MPH; Jeffrey D Lebensburger², DO; Pranav S Garimella³, MPH, MBBS; Srila Gopal⁴, MD

¹Tracking California, Public Health Institute, Oakland, CA, United States
²Division of Pediatric Hematology-Oncology, Department of Pediatrics, University of Alabama at Birmingham, Birmingham, AL, United States
³Division of Nephrology-Hypertension, Department of Medicine, University of California San Diego, San Diego, CA, United States
⁴Division of Hematology/Oncology, Department of Medicine, University of California San Diego, La Jolla, CA, United States

Corresponding Author:
Jhaqueline Valle, MPH
Tracking California
Public Health Institute
555 12th Street
Oakland, CA, 94607
United States
Phone: 1 4085942436
Email: jhaqueline.valle@trackingcalifornia.org

Abstract

Background: Chronic kidney disease (CKD) is a significant complication in patients with sickle cell disease (SCD), leading to increased mortality.

Objective: This study aims to investigate the burden of CKD in Medicaid-enrolled adults with SCD in California, examine differences in disease burden between male and female individuals, and assess mortality rates and access to specialized care.

Methods: This retrospective cohort study used the California Sickle Cell Data Collection program to identify and monitor individuals with SCD. Medicaid claims, vital records, emergency department, and hospitalization data from 2011 to 2020 were analyzed. CKD prevalence was assessed based on ICD (International Classification of Diseases) codes, and mortality rates were calculated. Access to specialist care was examined through outpatient encounter rates with hematologists and nephrologists.

Results: Among the 2345 adults with SCD, 24.4% (n=572) met the case definition for CKD. The SCD-CKD group was older at the beginning of this study (average age 44, SD 14 vs 34, SD 12.6 years) than the group without CKD. CKD prevalence increased with age, revealing significant disparities by sex. While the youngest (18-29 years) and oldest (>65 years) groups showed similar CKD prevalences between sexes (female: 12/111, 10.8% and male: 12/101, 11.9%; female: 74/147, 50.3% and male: 34/66, 51.5%, respectively), male individuals in the aged 30-59 years bracket exhibited significantly higher rates than female individuals (30-39 years: 49/294, 16.7%, \(P=0.01\); 40-49 years: 52/182, 28.6%, \(P=0.02\); and 50-59 years: 76/157, 48.4%, \(P<0.001\)). During this study, of the 2345 adults, 435 (18.5%) deaths occurred, predominantly within the SCD-CKD cohort (226/435, 39.5%). The median age at death was 53 (IQR 61-44) years for the SCD-CKD group compared to 43 (IQR 33-56) years for the SCD group, with male individuals in the SCD-CKD group showing significantly higher mortality rates (111/242, 45.9%; \(P=0.009\)) than female individuals (115/330, 34.9%). Access to specialist care was notably limited: approximately half (281/572, 49.1%) of the SCD-CKD cohort had no hematologist visits, and 61.9% (354/572) did not see a nephrologist during this study’s period.

Conclusions: This study provides robust estimates of CKD prevalence and mortality among Medicaid-enrolled adults with SCD in California. The findings highlight the need for improved access to specialized care for this population and increased awareness of the high mortality risk and progression associated with CKD.

(JMIR Public Health Surveill 2024;10:e57290) doi:10.2196/57290
KEYWORDS
sickle cell disease; chronic kidney disease; prevalence; mortality; access to care; Medicaid; California; United States; retrospective; cohort study; investigate; emergency department; hospitalization; specialized care; adult; adults; hematologist; hematologists; nephrologist; nephrologists; t-test

Introduction
Chronic kidney disease (CKD) presents a significant complication in sickle cell disease (SCD), marked by its high prevalence and a notable increase in mortality. Nearly 70% of adults with SCD develop albuminuria [1], and roughly 20% progress to overt CKD, characterized by a glomerular filtration rate (GFR) of less than 60 ml/min/1.73 m² [2]. Each year, over 100 patients with SCD advance to end-stage renal disease (ESRD) [3]. The mortality risk in patients with SCD-related ESRD is 3-fold higher compared to those without CKD [4], particularly with proteinuria. Key predictors of mortality in SCD are a reduced GFR and rapidly deteriorating kidney function [5-7], contributing to 16% to 18% of overall mortality in SCD [8]. Increasing age is a risk factor for CKD [9], and male individuals are more likely to have a faster and more rapid kidney function decline [10]. As life expectancy in patients with SCD increases [11], the burden of SCD-associated CKD is expected to rise.

While the impact of CKD on SCD outcomes is well recognized, contemporary data on its prevalence and burden, especially in the aging population with SCD, are limited. This is particularly true for patients outside of specialized sickle cell clinics. Most current prevalence estimates are derived from pediatric sickle cell centers [2] or older patient cohorts [12]. However, the situation for adults with SCD is starkly different. Comprehensive care for adults is notably lacking [13], and most do not have access to specialized hematological care [14]. Thus, patients who are followed at established sickle cell centers may not represent the entire population of patients with SCD, and CKD in the population with SCD may be underrecognized. Furthermore, the majority of patients with SCD, often reliant on Medicaid, face significant barriers to accessing specialized care [15]. Access to timely and early nephrology care can impact the treatment of CKD and subsequent complications. Thus, more robust and contemporary population-wide approaches may help us in improving our understanding of the burden and prevalence of CKD in adults with SCD.

To address the current limitations in the field, our study aims to investigate the burden of CKD among Medicaid-enrolled adults with SCD in California using the Sickle Cell Data Collection (SCDC) program’s large administrative database over a 10-year time frame (2011-2020). The SCDC program offers a comprehensive view as a state-based, population-wide public health surveillance system for SCD [16]. Our report estimates the prevalence of CKD in the population with SCD by age groups, providing insights into mortality rates and access to specialized care, both hematology and nephrology, for adults with and those without CKD.

Methods
Study Design
This is a retrospective cohort study of Medicaid-enrolled people with SCD in California. All study data were obtained from the California SCDC program. The SCDC program leverages a variety of data sources to identify and longitudinally monitor individuals with SCD within the state. These data sources include newborn screening, nonfederal hospital discharge records, emergency room visits, ambulatory surgery encounters, vital records death files, Medicaid claims, and enrollment data, as well as clinical case reports from SCD care centers within the state. The data are linked and deduplicated across various sources and multiple years. Detailed information about the SCDC program and case definitions have been previously published [16-19].

Ethical Considerations
The SCDC program and this study were reviewed and approved by the California Committee for the Protection of Human Subjects (protocol 15-10-2249) and the Public Health Institute Institutional Review Board (protocol I16-001). California SCDC received a waiver of consent.

Data Sources
Data sources used in this analysis include Medicaid claims, hospital discharge data, and vital records data from 2011 to 2020. The source data included all outpatient provider claims, inpatient hospital, emergency department encounters, and all-cause mortality.

Study Population
Individuals aged 18 years and older, who met the case definition for SCD, were enrolled in Medicaid in 2011, and who maintained their Medicaid coverage for at least 75% (75/100) of their time in this study were included. For example, if someone’s enrollment spanned 6 years, they were required to be enrolled for at least 4.5 years cumulatively, regardless of whether the enrollment was continuous.

Study Measures
The demographic characteristics of individuals are reported by the total study population. Age was calculated on the first date of this study, January 1, 2011. SCD subtype was recorded from State Newborn Screening records or clinical case reports from SCD centers. Subtypes were categorized as hemoglobin (Hb)–SS or Hb-Sβ0 thalassemia (sickle cell anemia), Hb-Sβ+ thalassemia, Hb-SC, other compound homozygous forms of SCD, or unknown if laboratory confirmation was not available within the SCDC database. An individual was considered dual eligible, meaning they were enrolled in both Medicare and Medicaid, if any Medicaid claims data indicated dual eligibility. For the prevalence estimation, age is calculated at the last date of this study, December 31, 2020, or date of death, whichever came first.
first. Disposition codes within the emergency department and hospitalization data as well as linked vital death records were used to identify all-cause deaths occurring during this study’s period. The proportion of deaths was calculated as the number of people with CKD who died, regardless of cause, by age group divided by the total number of people who developed CKD by age group, calculated at age on the last date of this study, December 31, 2020, or date of death, whichever came first.

We ascertained individuals’ CKD status (none versus any, the latter encompassing stages 1 through 5, unspecified stage, and ESRD), stage of CKD is based on ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) and ICD-10-CM (International Classification of Diseases, Tenth Revision, Clinical Modification) diagnosis codes. Individuals met the case definition for CKD if, within 5 years, they had 3 or more outpatient or emergency department encounters with CKD-coded diagnoses (ICD-9-CM: 585, 5851-5856, or 5859; ICD-10-CM: N18, N181-N185, or N188-N189) recorded in any diagnostic position, or at least 1 hospitalization with such codes in any position [20]. We identified a cohort of 572 individuals who by the end of this study’s period met the case definition for CKD. Subsequently, we will reference this subset as the SCD-CKD cohort. The remaining that did not meet the case definition will be referred to as the SCD cohort.

Hematologist and nephrologist encounters were identified using the National Provider Identifier of the rendering provider, as listed in the Medicaid claims records [21]. Providers with any health care provider taxonomy code listed as hematologist (207RH0000X, 207RH0003X, or 2080P0207X) were categorized as a hematologist; nephrologists were categorized as such if they had any health care provider taxonomy code listed as a nephrologist (207RN0300X). To calculate an individual’s total number of encounters with a hematologist (both groups) or nephrologist (CKD only), Medicaid claims were deduplicated by stipulating that an individual could only be recorded as having 1 encounter with a specific provider per day. The outpatient encounter rate was calculated by dividing the total number of unique encounters by the total person-years attributed to each group. To calculate the proportion of individuals who had no visits with a nephrologist or hematologist, the entire study duration was examined, and those with no encounters with either specialist were identified.

Statistical Analysis
Categorical variables were summarized using frequencies and percentages and compared for statistical significance using chi-square tests. Continuous variables were summarized by means and rates, and the Wilcoxon-Mann-Whitney U test was used to test for differences between SCD-CKD and the SCD groups. All analyses were performed using SAS (version 9.4; SAS Institute Inc).

Results
Characteristics of Study Cohorts
Table 1 presents the demographic characteristics of the 2345 individuals in this study’s cohort from 2011 to 2020.
### Table 1. Demographic characteristics of individuals included in this study\(^a\).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=2345)</th>
<th>SCD(^b)-CKD(^c) (n=572)</th>
<th>SCD (n=1773)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>916 (39.1)</td>
<td>105 (18.4)</td>
<td>811 (45.7)</td>
</tr>
<tr>
<td>30-39</td>
<td>530 (22.6)</td>
<td>107 (18.7)</td>
<td>423 (23.9)</td>
</tr>
<tr>
<td>40-49</td>
<td>453 (19.3)</td>
<td>148 (25.8)</td>
<td>305 (17.2)</td>
</tr>
<tr>
<td>50-59</td>
<td>308 (13.1)</td>
<td>136 (23.8)</td>
<td>172 (9.7)</td>
</tr>
<tr>
<td>60-64</td>
<td>65 (2.7%)</td>
<td>32 (5.6)</td>
<td>33 (1.9)</td>
</tr>
<tr>
<td>65+</td>
<td>73 (3.1)</td>
<td>44 (7.7)</td>
<td>29 (1.6)</td>
</tr>
<tr>
<td><strong>Age(^d) (years), mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36 (17.7)</td>
<td>44 (14)</td>
<td>34 (12.6)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1495 (63.8)</td>
<td>330 (57.7)</td>
<td>1165 (65.7)</td>
</tr>
<tr>
<td>Male</td>
<td>850 (36.3)</td>
<td>242 (42.3)</td>
<td>608 (34.3)</td>
</tr>
<tr>
<td><strong>SCD subtype , n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCA(^e) (Hb(^f)(^{-})SS or Hb-S(^g)(^{-})(\beta) Thal(^g))</td>
<td>332 (14.2)</td>
<td>96 (16.8)</td>
<td>236 (13)</td>
</tr>
<tr>
<td>Hb-S(^g)(^{-}) Thal or Hb-SC</td>
<td>144 (6.14)</td>
<td>28 (4.9)</td>
<td>116 (6.5)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1869 (79.7)</td>
<td>448 (78.3)</td>
<td>1421 (80.2)</td>
</tr>
<tr>
<td>Dual eligible, n (%)</td>
<td>941 (40.1)</td>
<td>333 (58.2)</td>
<td>608 (34.3)</td>
</tr>
</tbody>
</table>

\(^a\)Met case definition by the end of this study period.

\(^b\)SCD: sickle cell disease.

\(^c\)CKD: chronic kidney disease.

\(^d\)Age at the start of this study from January 1, 2011.

\(^e\)SCA: sickle cell anemia.

\(^f\)Hb: hemoglobin.

\(^g\)Thal: thalassemia.

At the start of this study’s period, the SCD-CKD group (n=572) were older in age, with nearly half (n=284, 49.7%) being in the aged 40-59 years categories (n=148, 25.8%, 40-49 years; n=136, 23.8%, 50-59 years). The mean age at the start of this study for the SCD-CKD cohort was 44 (SD 14) years. In contrast, the SCD cohort (n=1773) had an average age of 34 (SD 12.6) years, with 45.7% (n=811) of them between the ages of 18 and 29 years. In both the SCD-CKD and SCD groups, female individuals were more predominant at 57.7% (330/572) and 65.7% (1165/1773) respectively. Among the people with a known genotype across both groups, sickle cell anemia was more prevalent than Hb-S\(^g\)\(^{-}\) thalassemia or Hb-SC. However, nearly 79.7% (1869/2345) had unknown genotypes. More than half (333/572, 58.2%) of the SCD-CKD cohort were considered dual eligible for Medicaid and Medicare.

### CKD Prevalence by Age Group and Sex

The prevalence of CKD by age group and sex, as presented in Figure 1, shows a notable pattern of increasing prevalence with advancing age, alongside a distinct variation between female and male individuals. In the youngest age group (18-29 years) and oldest age groups (>65 years), the prevalence is similar between female and male individuals, at 10.8% (12/111) female and 11.9% (12/101) male and 50.3% (74/147) female and 51.5% (34/66) male, respectively. Male individuals have a significantly higher CKD prevalence in the 30 (\(P=0.01\)), 40 (\(P=0.02\)), and 50 (\(P<0.001\)) years age groups. The highest prevalence is observed in the age group of 65 years and older, with 50.3% (74/147) in female individuals and 51.5% (34/66) in male individuals, resulting in a combined prevalence of 50.7% (108/213). These estimates are from the end of this study’s period.
Of the 2345 adults, there were 435 (18.5%) deaths that occurred during this study’s period, with the majority of deaths occurring in the SCD-CKD cohort (226/572, 39.5% of the SCD-CKD group). The median age at death was 53 (IQR 61-44) years for the SCD-CKD cohort and 43 (IQR 33-56) years for the SCD cohort. When compared to the SCD cohort, the SCD-CKD cohort had higher death rates for all age groups (Table 2). The lowest total number of deaths was identified in the youngest age group with SCD-CKD (>18-29 years); however, we identified a 63% mortality rate among individuals with SCD-CKD in this youngest age group. The remaining age groups also had high rates of mortality that ranged from 34% (33/96; 30-39 years) to 45% (73/163) among the 50-59 years age group (Figure 2). Male individuals with SCD-CKD had a significantly higher rate of mortality (45.9%;111/242, \(P=.009\)) compared to female individuals (34.9%, 115/330) with SCD-CKD; this trend, although not statistically significant, persisted across most age groups (Table 2).

**Figure 1.** Sex-specific prevalence of chronic kidney disease among individuals with sickle cell disease by age group, 2011-2020. Fisher exact \(\chi^2\) test: *\(P<.05\); **\(P<.001\).
Table 2. All-cause mortality rates by age group, sex, and chronic kidney disease status among individuals with sickle cell disease, 2011-2020.

<table>
<thead>
<tr>
<th>Age group</th>
<th>SCD\textsuperscript{a}-CKD\textsuperscript{b} (n=226/572)</th>
<th>SCD (n=209/1,773)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female, n (%)</td>
<td>Male, n (%)</td>
</tr>
<tr>
<td>Total</td>
<td>115/330 (34.9)</td>
<td>111/242 (45.9)</td>
</tr>
<tr>
<td>&gt;18-29 years</td>
<td>CS\textsuperscript{d}</td>
<td>CS</td>
</tr>
<tr>
<td>30-39 years</td>
<td>14/47 (29.8)</td>
<td>19/49 (38.8)</td>
</tr>
<tr>
<td>40-49 years</td>
<td>21/67 (31.3)</td>
<td>19/52 (36.5)</td>
</tr>
<tr>
<td>50-59 years</td>
<td>36/87 (41.3)</td>
<td>37/76 (48.7)</td>
</tr>
<tr>
<td>60-64 years</td>
<td>CS</td>
<td>CS</td>
</tr>
<tr>
<td>&gt;65 years</td>
<td>25/74 (33.8)</td>
<td>15/34 (44.1)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}SCD: sickle cell disease.  
\textsuperscript{b}CKD: chronic kidney disease.  
\textsuperscript{c}Fisher exact chi-square test.  
\textsuperscript{d}CS: cell suppressed.

Access to Care

Figure 3 and Table 3 show the rate of outpatient encounters with a hematologist or a nephrologist per person-year and the proportion of people with SCD-CKD who had zero visits with either specialist. Among those with SCD-CKD, visit rates with a hematologist were approximately 2 visits per person-year; however, almost half (281/572, 49.1%) of people had no encounters with a hematologist over the entire study period. Access to nephrology care was also limited; in the group with SCD-CKD, individuals had only approximately 1 visit to a nephrologist per person-year, and 61.9% (354/572) of people did not have any encounters with nephrologists throughout the entire study period.

Figure 3. Proportion of individuals with sickle cell disease, with and without chronic kidney disease, who had 0 hematologist encounters for over 10 years, 2011-2020. CKD: chronic kidney disease; SCD: sickle cell disease.
Table 3. Specialist outpatient visit rates in individuals with sickle cell disease, with and those without chronic kidney disease, 2011-2020.

<table>
<thead>
<tr>
<th>Rate of outpatient hematologist visits per person-year</th>
<th>SCD(^a)-CKD(^b) (n=572)</th>
<th>SCD (n=1773)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>People with 0 visits with a hematologist(^d), n (%)</td>
<td>281 (49.1)</td>
<td>925 (52.1)</td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>Rate of outpatient nephrologist visits per person-year</td>
<td>0.9</td>
<td>—</td>
<td>.21(^e)</td>
</tr>
<tr>
<td>People with 0 visits with a nephrologist(^d), n (%)</td>
<td>354 (61.9)</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

\(^a\)SCD: sickle cell disease.
\(^b\)CKD: chronic kidney disease.
\(^c\)Mann-Whitney \(U\) test.
\(^d\)During the entire study period.
\(^e\)Fisher exact chi-square test.
\(^f\)Not applicable.

In our analysis of the SCD-CKD group, comparing patients who are deceased to those who are alive at this study’s end, we found a significantly higher proportion of patients who are deceased had never consulted a hematologist (152/226, 67.2% vs 129/346, 37.3%; \(P<.001\)) or seen a nephrologist (154/226, 68.1% vs 200/346, 57.8%; \(P=.007\)), as shown in Table 4.

Table 4. Rate of outpatient visits with a specialist by mortality among SCD\(^a\)-CKD\(^b\).

<table>
<thead>
<tr>
<th>Rate of outpatient hematologist visits per person-year</th>
<th>Deceased (N=226)</th>
<th>Alive (N=346)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>People with 0 visits with a hematologist(^d), n (%)</td>
<td>152 (67.2)</td>
<td>129 (37.3)</td>
<td>&lt;.001(^e)</td>
</tr>
<tr>
<td>Rate of outpatient nephrologist visits per person-year</td>
<td>1.3</td>
<td>0.8</td>
<td>&lt;.001(^c)</td>
</tr>
<tr>
<td>People with 0 visits with a nephrologist(^d), n (%)</td>
<td>154 (68.1)</td>
<td>200 (57.8)</td>
<td>.007(^e)</td>
</tr>
</tbody>
</table>

\(^a\)SCD: sickle cell disease.
\(^b\)CKD: chronic kidney disease.
\(^c\)Mann-Whitney \(U\) test.
\(^d\)During the entire study period.
\(^e\)Fisher exact chi-square test.

Discussion

Principal Findings

Our study estimates the prevalence and mortality associated with CKD in Californian adults with SCD using a large administrative database. Similar to other single and multi-institutional cohorts, we found a high prevalence of CKD (572/2345, 24.4%) among adults with SCD. Prevalence of CKD increased with age, with significant differences between male and female individuals. Mortality rates were higher in the SCD-CKD cohort. Access to specialized care was limited, with a significant proportion of individuals having no encounters with hematologists or nephrologists.

Prevalence of CKD by Age Group and Sex

We identified 2345 adults (aged >18 years) during 10 years within the SCDC cohort. Of these, 572 fulfilled case definitions for CKD, with an estimated CKD prevalence of 24.4% (572/2345). Prior studies show a high prevalence of CKD in the population with SCD, and our study confirms these findings. Further, 1 observational study tracked 427 patients over 4 years, finding a baseline CKD prevalence of 21.4% in sickle cell anemia and 17.2% in Hb-S\(^B\) thalassemia and Hb-SC SCD [2].

Studies from countries such as Nigeria and Ghana report even higher prevalence rates, around 38% and 39%, respectively [22,23]. Another study, a 25-year cohort study [24], identified the prevalence of renal failure at 4.2% in patients with sickle cell anemia (median age of onset was 23.1, range 13-57 years) and 2.4% in patients with Hb-SC SCD (median age of onset was 49.9, range 29-65 years). On additional follow-up, the prevalence increased to 12% [24]. The prevalence of CKD, calculated using the age at the end of this study or age at death, demonstrated an expected increase with advancing age. Our study continues to highlight the high prevalence of SCD-CKD in patients aged 18-29 years. By age >50 years, around 50% of patients will have developed SCD-CKD. Interestingly, our cohort predominantly comprised female individuals (1495/2345, 63.75%); this overrepresentation may be attributed to higher Medicaid coverage in female individuals, especially during their reproductive years [25]. Despite this female predominance, male individuals appear to have a higher prevalence of CKD compared to female individuals for each age category: a higher percentage of male individuals in our study had CKD (242/850, 28.5% vs 330/1495, 22.1%), deviating from the general population’s CKD distribution where female individuals slightly outnumber male individuals [26]. This finding further confirms
sex differences in the prevalence of SCD-CKD at a population level [27]. This may be explained by the fact that male individuals with SCD have a higher rate of decline in GFR when compared to female individuals [28,29], leading to a greater burden of CKD.

While male individuals appear to have a higher prevalence of CKD for each age category, CKD prevalences for male and female individuals are the same in the aged 60-64 years group. This phenomenon may be partially explained by higher mortality rates for male individuals in the SCD-CKD cohort, and hence the comparable CKD rates in the oldest age groups may be because more male individuals with SCD-CKD are dying at a younger age. Taken together, these findings further emphasize the importance of the early CKD screening and intervention in SCD, especially in younger and male populations.

**CKD and Mortality**

During this study’s period, 435 deaths occurred, translating to an overall mortality rate of 18.6% (435/2345). Notably, the mortality rate in the SCD-CKD cohort was high at 39.5% (226/572). Male individuals in the SCD-CKD cohort exhibited a higher death rate (45.9%, 111/242) compared to female individuals (34.9%, 115/330), deviating from previous studies that did not find a sex difference in mortality among patients with CKD and with SCD [1]. Notably, 50.9% (115/226) of the deaths in the SCD-CKD cohort occurred in individuals with ESRD, reinforcing the association of kidney disease with heightened mortality risk in SCD. The high mortality rate seen in our study is notably higher than the 18% found in the study by Platt et al [8]. In total, 1 key difference is the level of health care access between the 2 populations; the study by Platt et al [8] followed patients enrolled and regularly seen at established SCD centers, whereas nearly half or more of the patients included in this analysis never had an encounter with a hematologist (281/572, 49.1%) or nephrologist (354/572, 61.9%). In addition, the definition for renal failure was different in Platt et al [8], where they defined renal failure as a 20% increase in baseline creatinine concentration and a creatinine clearance rate below 100 mL/min [8], while we used ICD (International Classification of Diseases) codes to identify individuals with CKD. In our study, the median age at death was 10 years higher in the SCD-CKD cohort compared to the non-CKD SCD cohort, which was an unexpected finding given that CKD is a known risk factor for mortality in SCD [7]. This could be attributed to the older age distribution of the SCD-CKD cohort. This suggests that CKD may be more prevalent in the aging population with SCD, potentially characterizing it as a complication of the aging population with SCD and perhaps can be considered a disease of the “survivors.”

The significantly higher mortality rate in the SCD-CKD group, especially in younger age groups, is another important finding. The stark contrast in mortality rates between male and female individuals with SCD-CKD in the youngest age group indicates a need for focused research on gender-specific factors contributing to these outcomes. This elevated mortality risk associated with CKD in patients with SCD underscores the importance of the early detection and management of kidney disease.

**Access to Care**

Severe lack of access to specialty services remains a critical issue for adults with SCD [14]. Notably, while the SCD-CKD group had more hematologist visits than the SCD group, nearly half had no specialist visits over the 10 years, highlighting a critical gap in specialty care access for adults with SCD. This lack of access is even more considerable given that 67.2% (152/226) of individuals with SCD and CKD who died had not consulted a hematologist, compared to 37.3% (129/346) among survivors. Furthermore, only 38.1% (n=218) of those with CKD had consulted a nephrologist, and likely only in the advanced stages of their disease. This is not a surprising observation, since it is likely that patients who had more access to care also had more access to preventative care, early detection of CKD, and likely disease-modifying strategies. Despite ESRD qualifying for Medicare, only 58.2% (333/572) were dual eligible, indicating that Medicaid coverage alone may not provide sufficient access to nephrology care [3]. This issue is not unique to the population with SCD and reflects broader systemic disparities in health care access, particularly for younger, predominantly Black and Hispanic individuals on Medicaid [3]. These findings emphasize the need for improved access to specialized care to better manage CKD and potentially reduce mortality in the population with SCD.

**Limitations**

While there are several strengths to our study approach as outlined above, there are also several limitations. The major limitation of our study is that it is a retrospective administrative data-based study. The use of ICD codes to capture CKD diagnoses has been validated previously [20], however, we are likely underestimating the burden of CKD for multiple reasons. We did not include ICD codes for proteinuria since we felt this would be less reliable when relying on administrative data alone for data collection. While we included all stages of CKD, we likely underestimated CKD 1 and 2 where the GFR is >60 mL/min because not all of these may have been recognized or coded correctly. In assessing the limitations of our methodology for calculating the prevalence, it is important to consider the following aspects. First, individuals were categorized as either having CKD or not based on meeting the case definition at any point within a 10-year timeframe. This approach simplifies the classification process but may overlook cases of CKD who did not use health care or were not identified or documented with the relevant codes during the specified period. Additionally, because we used ICD codes within a set period of time, we could only identify when the first ICD CKD code appeared and not when a person was first diagnosed with CKD. Lastly, only Medicaid recipients were included in our study and about a third of patients with SCD may have commercial insurance or Medicare only and may not have been included in our analysis. In our assessment of care access, we focused exclusively on visits to hematologists, without distinguishing providers with specific training in SCD. Previous research indicates that adult patients with SCD in California face significant access barriers, and including primary care providers with SCD expertise—defined as having 20 or more patients with SCD—did not substantially improve access [14]. In addition, we could not ascertain whether individuals who disenrolled
from Medi-Cal then enrolled in other health plans or became uninsured. While our data does not enable us to confirm whether individuals accessed specialty care during periods they were not enrolled in Medicaid, the literature suggests that accessing specialist services without insurance is highly unlikely [30]. Since this was an administrative database-based study, we had limited access to detailed clinical information, and could not adjust for covariates or perform robust statistical analysis, hence our results are presented descriptively.

**Conclusions**

We used administrative data to estimate the prevalence of CKD in SCD among California state Medicaid recipients. Our prevalence estimates for CKD are higher than what has been reported previously. Those with SCD and concurrent CKD experienced significantly higher mortality rates compared to individuals with SCD who did not meet the case definition for CKD. In California, individuals with SCD face substantial barriers to accessing specialty care. Our findings highlight the critical need for this data to guide state health care stakeholders and government policy makers in addressing these access issues. As the SCDC program expands to include additional states, similar studies may inform our understanding of the true burden of CKD in the population with SCD, including its effects on mortality, morbidity, health care use, and quality of life.

**Acknowledgments**

This work is funded by the Centers for Disease Control and Prevention’s DD20-2003 Sickle Cell Data Collection Program (NU58DD000016).

**Data Availability**

These data are sourced from third parties, and data use agreements that prohibit sharing even deidentified person-level data. Should readers wish to explore gaining access to the data, they may contact the original data stewards. Requests for minimal data set data can be directed to the corresponding author.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

CKD: chronic kidney disease
ESRD: end-stage renal disease
GFR: glomerular filtration rate
Hb: hemoglobin
ICD: International Classification of Diseases
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
SCD: sickle cell disease
SCDC: sickle cell data collection

Edited by A Mavragani; submitted 11.02.24; peer-reviewed by B Adesina, N Koizumi; comments to author 05.03.24; revised version received 25.03.24; accepted 17.05.24; published 15.07.24.

Please cite as:
Valle J, Lebensburger JD, Garimella PS, Gopal S
Prevalence, Mortality, and Access to Care for Chronic Kidney Disease in Medicaid-Enrolled Adults With Sickle Cell Disease in California: Retrospective Cohort Study
JMIR Public Health Surveill 2024;10:e57290
URL: https://publichealth.jmir.org/2024/1/e57290
doi: 10.2196/57290
PMID: 39008353

© Jhaqueline Valle, Jeffrey D Lebensburger, Pranav S Garimella, Srila Gopal. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 15.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Syndromic Surveillance Tracks COVID-19 Cases in University and County Settings: Retrospective Observational Study

Lily Minh Wass¹, BS; Derek O'Keeffe Hoare², MS; Georgia Elena Smits², MS; Marwan Osman¹,³, MPH, PhD; Ning Zhang⁴, MPH; William Klepack⁵, MD; Lara Parrilla¹, MPH; Jefferson M Busche⁶, MS; Marin E Clarkberg⁶, PhD; Sumanta Basu², PhD; Casey L Cazer¹,⁷, DVM, PhD

Abstract

Background: Syndromic surveillance represents a potentially inexpensive supplement to test-based COVID-19 surveillance. By strengthening surveillance of COVID-19–like illness (CLI), targeted and rapid interventions can be facilitated that prevent COVID-19 outbreaks without primary reliance on testing.

Objective: This study aims to assess the temporal relationship between confirmed SARS-CoV-2 infections and self-reported and health care provider–reported CLI in university and county settings, respectively.

Methods: We collected aggregated COVID-19 testing and symptom reporting surveillance data from Cornell University (2020 - 2021) and Tompkins County Health Department (2020 - 2022). We used negative binomial and linear regression models to correlate confirmed COVID-19 case counts and positive test rates with CLI rate time series, lagged COVID-19 cases or rates, and day of the week as independent variables. Optimal lag periods were identified using Granger causality and likelihood ratio tests.

Results: In modeling undergraduate student cases, the CLI rate ($P=.003$) and rate of exposure to CLI ($P<.001$) were significantly correlated with the COVID-19 test positivity rate with no lag in the linear models. At the county level, the health care provider–reported CLI rate was significantly correlated with SARS-CoV-2 test positivity with a 3-day lag in both the linear ($P<.001$) and negative binomial model ($P=.005$).

Conclusions: The real-time correlation between syndromic surveillance and COVID-19 cases on a university campus suggests symptom reporting is a viable alternative or supplement to COVID-19 surveillance testing. At the county level, syndromic surveillance is also a leading indicator of COVID-19 cases, enabling quick action to reduce transmission. Further research should investigate COVID-19 risk using syndromic surveillance in other settings, such as low-resource settings like low- and middle-income countries.

(JMIR Public Health Surveill 2024;10:e54551) doi:10.2196/54551

KEYWORDS

COVID-19; epidemiology; epidemiological; SARS-CoV-2; syndromic surveillance; surveillance system; syndromic; surveillance; coronavirus; pandemic; epidemic; respiratory; infectious; predict; predictive; prediction; predictions

Introduction

SARS-CoV-2 continues to be one of the most significant causes of morbidity and mortality, with over 697,000,000 total COVID-19 cases and 6,900,000 deaths recorded worldwide as of November 1, 2023 [1]. Although rates of COVID-19 infection, hospitalizations, and mortality have significantly declined due to global vaccination efforts [2], the emergence of the highly mutated Omicron variant and following subvariants continue to fuel spikes in COVID-19 cases [3]. National public health programs have primarily relied on diagnostic testing to gauge COVID-19 case trends. While useful for tracking the incidence of COVID-19 cases, large-scale polymerase chain reaction–based surveillance testing programs
have high supply and labor costs. Coupled with the shift to at-home testing—a practice that mainly grew in the first Omicron wave beginning in December 2021—which is largely unreported [4], this has led to a substantial decline in reported daily tests in the United States. Shifting the responsibility of case reporting from health care facilities to patients has led to the underestimation of true COVID-19 trends and has challenged public health measures such as isolation and quarantine of individuals who are infected, as well as the measurement of negative effects or financial costs of COVID-19. Furthermore, May 11, 2023, marked the expiration of the US federal COVID-19 Public Health Emergency declaration, along with authorizations to collect certain public health data [5]. This triggered a pivot from COVID-19 case-based surveillance to COVID-19–associated hospital admissions as the leading indicator of COVID-19 trends, with many national COVID-19–reporting platforms like the Centers for Disease Control and Prevention (CDC) removing case-based metrics from their websites. Looking at hospitalization data as an indication of the current state of COVID-19, over 16,000 hospitalizations were reported in the week of October 15, 2023 [6].

As endemic COVID-19 seems to be the likely future of the pandemic [7], sustainable disease monitoring systems are warranted. Syndromic surveillance, or the detection and recording of symptoms before a diagnosis is confirmed, could serve as a less resource-intensive method to monitor trends in COVID-19–like illness (CLI) for public health departments and health care facilities. The National Syndromic Surveillance Project of the CDC defines CLI as “fever and cough or shortness of breath or difficulty breathing with or without the presence of a coronavirus diagnosis code. Visits meeting the CLI definition that also have mention of flu or influenza are excluded” [8].

Syndromic surveillance has already been used in a few ways throughout the pandemic. The National Syndromic Surveillance Project pivoted to reporting on CLI using local, academic, and private partnerships [9], while county health departments like Seattle and King County of Washington State have reported CLI trends from hospitalizations and emergency department (ED) visits [10]. COVID Control, a Johns Hopkins University study, piloted a symptom reporting app in 1019 counties and identified loss of taste or smell to be a leading predictor of SARS-CoV-2 rates, appearing 5 days before confirmatory diagnostic tests [11]. Limitations of these prior syndromic surveillance studies include reliance on ED data [10], which may exclude patients with COVID-19 who are asymptomatic or have mild symptoms. Although COVID-19 hospitalizations are ongoing, they have significantly declined since the onset of the pandemic and subsequent Omicron spikes, and are largely composed of populations older than 70 years [6], making surveillance that relies exclusively on ED admissions and hospitalizations biased. Additionally, the COVID Control study selected participants experiencing CLI, which does not indicate how rates of symptoms relative to the total community or population track true SARS-CoV-2 positivity rates [11]. To date, there is still limited information on how CLI surveillance can be used to accurately monitor community transmission, including asymptomatic and mild infections.

Cornell University and Tompkins County Health Department (TCHD) each developed a CLI surveillance system and a robust tracking system for confirmed COVID-19 cases. In the university setting, surveillance testing and daily symptom reporting were enforced among all individuals on campus, resulting in a daily record of testing, symptom reporting, and exposure data. In contrast, the county recorded CLI data from patients who voluntarily sought out practitioner-based care, and they had a robust free COVID-19 testing program. In this paper, we assess whether levels of reported COVID-19 symptoms are correlated with test positivity levels during the Delta and B.1 variant predominant period in a county and university setting. We identify the optimal temporal lag between changes in symptom reporting and increases in COVID-19 cases, and which symptom or health screening questions had the strongest association with test positivity rates. We provide recommendations that can be used by institutions of higher education and local county health departments to improve SARS-CoV-2 surveillance in the endemic phase of the pandemic.

Methods

Study Design and Population

During the 2020-2021 academic year, Cornell University students (n=12,988 undergraduates, n=6549 graduate and professional students), faculty (n=1332), and staff (n=6739) participated in a robust COVID-19 testing program and daily syndromic surveillance, called the Daily Check. Undergraduate and professional students were tested twice a week, and graduate students, faculty, and staff were tested weekly for SARS-CoV-2 via anterior nares swab and polymerase chain reaction [12], as described elsewhere [13,14]. Supplemental SARS-CoV-2 testing occurred at the request of individuals and when increased case counts were noted in specific subpopulations. For the Daily Check, all on-campus students were asked to report daily whether they had COVID-19–like symptoms or had potential exposures to a confirmed SARS-CoV-2 case or a person with CLI. Each day, students who answered “yes” to any of these questions were labeled with a “red” status until cleared by Cornell Health with either a SARS-CoV-2 test or follow-up questions. A flowchart with the labeling logic is provided in Multimedia Appendix 1. Cornell University’s Office of Institutional Research and Planning provided aggregated data on the daily number of students, faculty, and staff who were flagged as “red,” experiencing CLI themselves, had been exposed to a confirmed COVID-19 case, had been exposed to someone experiencing CLI, and tested positive for SARS-CoV-2. The daily total number of SARS-CoV-2 tests conducted at Cornell and Daily Checks was also provided [15]. For this study, we analyzed data collected between August 17, 2020 (the first day of data collection for both the syndromic surveillance and testing programs) and February 3, 2021 (the last date before questions on the Daily Check survey changed substantially) from undergraduate students. Days with fewer
than 100 SARS-CoV-2 tests conducted in the undergraduate population were excluded from the analysis.

Starting in June 2020, local health practitioners in Tompkins County were encouraged by the health department to record how many patients they saw each day and, of those patients, how many presented with CLI. TCHD defines CLI as “cough and/or shortness of breath OR at least two of the following: fever; chills; repeated shaking with chills; myalgias; headache; sore throat; new loss of sense of taste or smell.” Practices were excluded from this study if they reported fewer than 50 times between June 2020 and March 2022. A pediatric practice was excluded because the prevalence of CLI varies greatly in children relative to adults, and children also experience differing clinical presentations of COVID-19 [16,17]. Data from the remaining practices were aggregated by the date of data collection. CLI data reported on weekends were excluded because most practices were not open on the weekends. This resulted in a data set of total patient encounters, CLI encounters, and the number of practitioners who contributed to reporting each day.

Free SARS-CoV-2 testing has been available to Tompkins County residents since March 2020. A record of new positive results, cumulative total positive results, and cumulative total tests were kept between March 14, 2020, and December 20, 2022, and is publicly available on TCHD’s website [18]. As Cornell University is within Tompkins County, surveillance and diagnostic testing data from the university were included in the data set provided by Tompkins County. Starting December 1, 2020, the number of daily laboratory tests was also reported. Our analysis focused on the period from December 1, 2020 (after the number of tests was available to calculate the proportion of positive SARS-CoV-2 tests) to March 11, 2022, the last week of surveillance before the BA 2.2.12.1 variant (Omicron subvariant) exceeded more than 50% of new cases in New York State [19]. The predominant variants within our study were Delta and BA.1 [19], which minimizes possible variation due to increased transmissibility or different predominant symptoms caused by BA.2. Dates when fewer than 300 SARS-CoV-2 tests were conducted were excluded.

Ethical Considerations

This study was determined to not meet the definition of human subject research by the Cornell University Institutional Review Board because the data were deidentified and aggregated. The original data were collected as part of public health surveillance activities by the TCHD and Cornell University, and therefore did not meet the definition of human subject research.

Data Analysis

Within both the Cornell and TCHD data sets, the rate of SARS-CoV-2 positivity was calculated by dividing the count of new daily positive test results by the number of daily tests conducted within the undergraduate or county setting. For Cornell undergraduates, rates of CLI, exposure to CLI, and exposure to COVID-19 were calculated as the sum of each Daily Check variable over the total number of Daily Checks completed that day. The CLI rate in Tompkins County was the sum of patients experiencing CLI divided by the total number of patients seen that day. The rate of county CLI encounters on weekends was imputed as the average of Friday and Monday rates. The county CLI rate on US holidays was imputed in the same manner, by averaging rates from the two nearest dates before and after the holiday. In addition, a variable for the day of the week was encoded based on the date of data collection for both data sets.

All variables defined for the statistical analysis are detailed in Supplemental Tables 1 and 2 in Multimedia Appendix 1. For the initial descriptive analysis, the moving 7-day averages of the CLI and SARS-CoV-2 positivity rate time series were plotted. The moving average was symmetrical; the values of each variable from ±3 days and at present were summed and divided by 7.

The autocorrelation of SARS-CoV-2 positivity rates and the cross-correlation between the CLI and SARS-CoV-2 positivity time series were plotted at up to 10-day lags to identify the optimal lag period between independent variables and the outcome of SARS-CoV-2 positivity. The Granger causality test assesses whether previous values of variable x are useful for forecasting current values of variable y [20]. The grangertest function of the lmtest R package compares an autoregressive model of y that only uses previous values of y as independent variables with a model that includes previous values of x and y as independent variables [21]. We used the Granger causality test to determine if the addition of CLI rates (x) forecasted SARS-CoV-2 positivity rates (y) better than autocorrelation of y alone. The temporal lag of the CLI rate to be included in subsequent models was selected based on which previous days of x had the smallest P values in the Granger causality test.

Based on the findings of the autocorrelation and cross-correlation plots, two types of models were built to assess the association between CLI rates at present and from the prior 1-6 days, and the current SARS-CoV-2 positivity measure. The linear model used the SARS-CoV-2 positivity rate as the outcome. Independent variables included CLI rates at present (day 0) and from the prior 1-6 days, day of the week, and the SARS-CoV-2 positivity rate from the prior 1-6 days. In the university setting, two linear models were constructed with independent variables including either the rate of CLI or rate of exposure to CLI from up to 6 days prior. Significant variables were identified if their associated P value was less than .05.

\[
\text{SARS-CoV-2rate (day i)} = \text{CLIrate (day i)} + \text{SARS-CoV-2rate (day i− weekday)}
\]

where i=0, ..., I, and I is the selected maximum number of days prior.

Next, the county data was tested for overdispersion, and a negative binomial model was fitted to the data. In the negative binomial model, counts of daily SARS-CoV-2 positive tests were the outcome. CLI rates (day 0 through 6 days prior), day of the week, and counts of SARS-CoV-2 infection from the prior 1-6 days were independent variables. The CLI variable was transformed to a (0, 100) scale for the model estimates to be interpreted as percentage changes in the outcome variable of SARS-CoV-2 counts. To account for variation in the number of positive tests caused by the daily number of tests conducted.
administered, an offset of the log of daily tests conducted was included in the model.

$$\log (\text{SARS}_\text{CoV}_2 \text{ count}) = \phi_0 + \sum_{i=1}^{7} \phi_i \log (\text{CLI rate})_{i+1} + \sum_{i=1}^{16} \phi_i \log (\text{SARS}_\text{CoV}_2 \text{ count})_{i+\text{weekday} + \log (\text{daily tests})}$$

where \(i=0, \ldots, I\), and \(I\) is the selected maximum number of days prior.

The significance of independent variables was assessed by comparing the \(P\) value associated with each input with an \(\alpha\) of 0.05. As the negative binomial model applies a log transformation to both independent and outcome variables, the incident rate ratio estimates of the model were exponentiated to undo the log transformation.

Models within the same class were compared using the likelihood ratio test to determine whether the addition of CLI information and the day of the week yielded a significantly better fit than a model based on previous measurements of SARS-CoV-2 positivity alone.

Data were analyzed using R version 4.1.0 (R Foundation for Statistical Computing) in the RStudio environment [22].

## Results

### Analysis of CLI and Cases in University Setting

We collected 124 days of Daily Check and surveillance testing data between August 17, 2020, and February 3, 2021, from Cornell University. The symmetrical 7-day moving average of SARS-CoV-2 positivity rates and the rate of “red” flagged students are plotted in Figure 1.
The SARS-CoV-2 test positivity rate ranged between 0 and 0.021 with an average of 0.002. The maximum number of SARS-CoV-2 positive cases on a single day was 20. On average, 1893 tests were conducted daily. The CLI rate ranged from 0 to 0.005 with a maximum of 58 students reporting CLI in one day. The rate of exposure to CLI ranged from 0 to 0.006, with a daily maximum of 69 students reporting exposure to CLI. Finally, the rate of exposure to a confirmed COVID-19 case ranged from 0 to 0.008 with a maximum count of 87 exposures reported in one day.

The rate of SARS-CoV-2 test positivity from 1 to 4 days prior and 7 days prior was significantly autocorrelated with the SARS-CoV-2 test positivity rate at present (Supplemental Figure 1 in Multimedia Appendix 1). We observed significant cross-correlation coefficients between the present-day SARS-CoV-2 positivity rate and lagged rates of students experiencing COVID-19 symptoms, contact with someone experiencing COVID-19 symptoms, and contact with a confirmed COVID-19 case (Supplemental Figure 2 in Multimedia Appendix 1).

The results of Granger causality test, comparing autoregression of the rate of positive tests with and without previous daily rates of CLI or exposure among undergraduates, are shown in Figure 2. The rate of COVID-19 symptoms from up to 6 days prior was found to improve forecasting of the test positivity rate relative to prior values of the positivity rate alone ($P<.001$).
Similarly, including the rate of newly exposed students to CLI from up to 6 days prior significantly improved the forecasting of test positivity ($P=.02$). However, the addition of data regarding exposure to a confirmed COVID-19 case from up to 6 days prior did not significantly improve forecasting relative to an autoregression of the test positivity variable alone ($P=.51$). The results of the regressions used in the Granger test can be seen in Supplemental Tables 3-5 in Multimedia Appendix 1.

Based on the linear models, a 1 percentage point (pp) increase in the rate of undergraduates experiencing CLI was significantly associated with a 1.36 (95% CI 0.46-2.26) pp increase in the rate of SARS-CoV-2 test positivity on the same day ($P=.003$). A 1 pp increase in the rate of students reporting new contacts with people experiencing CLI was significantly associated with a 1.66 (95% CI 0.83-2.50) pp increase in the SARS-CoV-2 positivity rate on the same day as well ($P<.001$; Table 1). The full model outputs can be found in Supplemental Tables 6 and 7 in Multimedia Appendix 1.

The rate of CLI and exposure to CLI had the largest statistically significant model coefficients in their models, even compared to previous rates of SARS-CoV-2 positivity (Supplemental Tables 6 and 7 in Multimedia Appendix 1). Using the likelihood ratio test, models were found to be significantly different from the nested model that removed the time series of CLI or exposure to CLI ($P<.001$). The likelihood ratio test also showed that the addition of the day of the week did not significantly improve the rate of the CLI model ($P=.15$) but did significantly improve the exposure to the CLI model ($P=.02$) and was therefore included as an independent variable for both models.
Figure 2. Granger causality test $P$ values for models including the proportion of undergraduate students with new COVID-19 symptoms (blue), students newly in contact with people who have COVID-19 symptoms (red), and students newly in contact with a confirmed COVID-19 case (yellow) from up to 6 days prior as independent variables, with the SARS-CoV-2 positivity rate as the outcome variable, relative to the autoregressive model of the positivity rate alone. Values below the green dotted line are significant ($\alpha=.05$). pos: positive test results.
Table. Optimal lag time between previous rates of surveillance indicators (x) and current SARS-CoV-2 positivity (y), and regression coefficient for the optimal lag. A lag time of 1 indicates COVID-19–like illness (CLI) data are taken from 1 day prior to the COVID-19 testing data, a lag of 2 is from 2 days prior, and so on.

<table>
<thead>
<tr>
<th>Indicator (model type)</th>
<th>Optimal lag time(^a) (days)</th>
<th>Largest model coefficient (95% CI)</th>
<th>P value for model</th>
<th>P value for Granger causality</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLI rate among undergraduates (linear)</td>
<td>0</td>
<td>1.36 (0.46-2.26)</td>
<td>.003</td>
<td>b</td>
</tr>
<tr>
<td>CLI exposure rate among undergraduates (linear)</td>
<td>0</td>
<td>1.66 (0.83-2.50)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>CLI rate in Tompkins County (linear)</td>
<td>3</td>
<td>0.20 (0.10-0.30)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>CLI rate in Tompkins County (negative binomial)</td>
<td>3</td>
<td>1.04 (1.01-1.07)</td>
<td>.005</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\)Optimal lag time is selected based on the largest model coefficient for linear models and largest incidence rate ratio for negative binomial models that is significant at \(\alpha=0.05\).

\(^b\)Granger causality assesses correlation of previous days of CLI with current measures of positivity; therefore day 0 is not included.

Analysis of CLI and Cases in the County Setting

In the Tompkins County setting, 403 days of CLI and SARS-CoV-2 testing data between December 1, 2020, and March 11, 2022, were included after data cleaning. The 7-day moving averages of the CLI rate and SARS-CoV-2 positivity rate are plotted in Figure 3.

The SARS-CoV-2 positivity rate ranged from 0.00 to 0.64, with the maximum value observed around the peak of the first Omicron (BA.1.1) wave [23] on December 19, 2021. The average positivity rate was 0.02. The minimum daily SARS-CoV-2 positivity count was 0, while the maximum was 523 cases. On average, 3235 SARS-CoV-2 tests were conducted daily. The CLI encounter (CLI encounters per total patient encounters) rate ranged from 0 to 0.71; the most patients with CLI seen on a single day was 60 with an average of 15 patients with CLI seen daily.

There was substantial autocorrelation of SARS-CoV-2 positivity rates; 1- through 5-day lags of the SARS-CoV-2 positivity rate had autocorrelation coefficients greater than 0.20, after which values declined (Supplemental Figure 3 in Multimedia Appendix 1). The cross-correlation plots revealed nonzero correlation coefficients between CLI rates up to 10 days prior and current rates of SARS-CoV-2 positivity (Supplemental Figure 4 in Multimedia Appendix 1). However, the highest correlation coefficients were observed in the 4-day lag (\(\beta=0.295\)) and 6-day lag (\(\beta=0.346\)) of the CLI rate. This result was confirmed by Granger causality, which revealed that including CLI rates from 3 to 6 days prior yielded the smallest P values (Figure 4). For CLI rates from 1 week prior (lags 1-6), minimum Granger causality P values were observed at a lag of 4 and 6 days (\(P<.001\); Supplemental Table 8 in Multimedia Appendix 1). Based on these results, rates of CLI from 1 to 6 days prior were selected to be included in subsequent models. Previous rates of SARS-CoV-2 positivity from up to 6 days were included based on the significant autocorrelation observed (Supplemental Figure 3 in Multimedia Appendix 1). Significant autocorrelation and Granger causality of rates from 7 days prior and at present were observed, but 7-day lagged variables were not included to account for the variation that the day of the week may cause in surveillance testing and CLI reporting capabilities. Instead, the weekday variable was included to account for possible variation due to the day of the week that data was collected.
Figure 3. Seven-day symmetrical mean of the CLI encounter rate, shown by the solid line, and SARS-CoV-2 positivity rate, shown by the dotted line, from December 1, 2020, to March 11, 2022, in Tompkins County. CLI: COVID-19–like illness.
Figure 4. Granger causality test $P$ values for a model using the COVID-19–like illness rate from 1 to 6 days prior as independent variables, with present-day SARS-CoV-2 positivity rates as the outcome variable. All values are significant at $\alpha=.05$; however, $P$ values decline sharply upon including lags from up to 3 or more days prior. Values below the dashed green line are significant at $\alpha=.05$.

In the linear regression model (Supplemental Table 9 in Multimedia Appendix 1), a 1 pp increase in the rate of CLI 3 days prior was significantly associated with a 0.20 (95% CI 0.10-0.30) pp increase in the current SARS-CoV-2 positivity rate ($P<.001$). Similarly, a 1 pp increase in the rate of CLI 4 and 6 days prior was significantly associated with a 0.15 (95% CI 0.04-0.26) pp and 0.19 (95% CI 0.08-0.30) pp increase in current SARS-CoV-2 positivity rates ($P=.01$ and $P<.001$), respectively. The rate of CLI from up to 3 days prior had the largest regression coefficient across all linear model independent variables (Supplemental Table 8 in Multimedia Appendix 1).

In the negative binomial model (Supplemental Table 10 in Multimedia Appendix 1), a 1 pp increase in the rate of CLI encounters at a 3-day lag was significantly associated with a 4% (95% CI 1.01-1.07) increase in the number of SARS-CoV-2 positive tests at present ($P=.005$). The rate of CLI from up to 3 days prior had the highest incidence rate ratio across all numeric independent variables, including previous counts of SARS-CoV-2 positive results (Supplemental Table 10 in Multimedia Appendix 1).

Both full models were found to be significantly different from the nested model, which excluded all CLI independent variables, using the likelihood ratio test (linear model: $P<.001$; negative
binomial model: \( P = .02 \). In addition, the likelihood ratio test also found that the inclusion of the day of the week variable significantly improves the model (linear model: \( P < .001 \); negative binomial model: \( P < .001 \)). Therefore, this variable was included in the final model output (Supplemental Tables 9 and 10 in Multimedia Appendix 1).

**Discussion**

**Overview**

Tompkins County and the Cornell University community presented a unique opportunity to investigate the utility of CLI syndromic surveillance as an indicator of COVID-19 cases due to required testing and CLI reporting at Cornell University, in addition to widespread free testing and robust reporting of CLI by health practitioners encouraged by TCHD. We looked for congruence in rates of CLI and SARS-CoV-2 positivity to understand whether CLI trends were temporally correlated with measures of SARS-CoV-2–positive tests.

**Principal Findings**

Significant autocorrelation of the SARS-CoV-2 positivity variable was observed in the university and county setting, suggesting current test positivity is influenced by levels of test positivity from earlier in the week. Cross-correlation analysis also revealed CLI to lead and lag SARS-CoV-2 rates in both settings, indicating that the two measures can offer valuable insights into each other. This reflects the expected nature of infectious disease transmission in a population without complete immunity, in which exponential growth of infections can lead to more exposures, symptoms, and positive SARS-CoV-2 tests. Given our interest in the potential use of CLI to monitor SARS-CoV-2 positivity, we demonstrate that the addition of CLI and CLI exposure data can improve efforts to assess SARS-CoV-2 positivity in low-resource settings (ie, when large-scale testing is not possible). However, the rate of exposure to confirmed COVID-19 cases was not significantly associated with the SARS-CoV-2 test positivity rate and may not be a useful measure of COVID-19 risk in universities. A possible explanation for COVID-19 exposure not being a significant variable in our analysis is that confirmed exposure to a COVID-19 case is dependent on testing capacity to deliver a quick test result. In the university’s syndromic surveillance, students could have already been flagged as “red” based on reporting exposure to CLI before a positive test result has confirmed the case. Depending on testing capacity and availability, possible exposure to COVID-19 could be a stronger variable than confirmed exposure to COVID-19 for tracking SARS-CoV-2 in real time.

In the university setting, linear models demonstrated that reporting of CLI and exposure to CLI among students track changes in the SARS-CoV-2 positivity rate in real time. Within Tompkins County, both the linear and negative binomial model indicated that the strongest variable associated with present-day SARS-CoV-2 positivity is the rate of CLI from up to 3 days prior. This discrepancy in the lag time of the CLI indicator could reflect the difference in adherence to the two programs. The university surveillance program was strictly enforced for the entire on-campus population and may have been able to identify changes in SARS-CoV-2 cases immediately. By comparison, the testing and CLI reporting of Tompkins County was voluntary. This may have caused lags between when symptoms appeared and diagnosis by a positive test result due to delays in testing based on appointment availability, patients’ work schedules, access to transportation, and other conflicting factors. Cornell University demonstrates how mandatory symptom and exposure reporting could track SARS-CoV-2 positivity in real time. However, with mandatory surveillance programs being unrealistic to implement, Tompkins County illustrates how the utilization of existing infrastructure like local practitioners is also able to effectively measure CLI, which is associated with cases. Practitioners were asked to count CLI cases each day, a process that could be applied to larger practitioner and hospital networks. We demonstrated that CLI rates from practitioner-based surveillance track SARS-CoV-2 positivity rates with a small lag. Trends such as spikes in COVID-19 identified from syndromic surveillance programs could then be validated by brief surveillance testing or distribution of at-home tests.

**Limitations**

The study findings should be carefully interpreted within the context of the study population. Cornell University is located in a small county in upstate New York, with a population of 105,162 in 2021 [24]. Tompkins County is somewhat isolated geographically, so these results may not generalize to denser urban areas or more rural areas. While influenza rates substantially decreased during the 2020–2021 flu season [25], some presentations of CLI in the study population are inevitably due to flu and other respiratory illnesses. This may explain a spike in university cases of CLI during November and December 2021 that was not accompanied by an increase in SARS-CoV-2 positivity, as flu rates did increase in New York State during this period [26]. On the other hand, given that both flu seasons included in our study period were less severe than usual, it is possible that the associations we observed may not be as strong when other common respiratory illnesses are in wider circulation. In addition, participating practitioners of our study voluntarily partook in syndromic surveillance without reimbursement. To apply this model of surveillance on a larger scale, financial incentives should be considered to ensure robust reporting—a cost that would need to be considered before implementation. Finally, as this was an observational study, we can only infer relationships and not their underlying cause.

**Significance**

Our findings are contextualized by the shift away from robust SARS-CoV-2 testing and reporting in the United States. At-home tests are now commonly used and typically are not included in COVID-19 case count reporting [4]. Federal funding for SARS-CoV-2 testing is diminishing [27], and COVID-19 response teams of health departments are being scaled back. On May 11, 2023, the US federal government ended the COVID-19 Public Health Emergency declaration [5]. As a result, the CDC and other reporting platforms no longer publish SARS-CoV-2 testing and reporting in the United States.

Our findings are contextualized by the shift away from robust SARS-CoV-2 testing and reporting in the United States. At-home tests are now commonly used and typically are not included in COVID-19 case count reporting [4]. Federal funding for SARS-CoV-2 testing is diminishing [27], and COVID-19 response teams of health departments are being scaled back. On May 11, 2023, the US federal government ended the COVID-19 Public Health Emergency declaration [5]. As a result, the CDC and other reporting platforms no longer publish SARS-CoV-2 testing and reporting in the United States.

Our findings are contextualized by the shift away from robust SARS-CoV-2 testing and reporting in the United States. At-home tests are now commonly used and typically are not included in COVID-19 case count reporting [4]. Federal funding for SARS-CoV-2 testing is diminishing [27], and COVID-19 response teams of health departments are being scaled back. On May 11, 2023, the US federal government ended the COVID-19 Public Health Emergency declaration [5]. As a result, the CDC and other reporting platforms no longer publish SARS-CoV-2 testing and reporting in the United States.
Low-resource alternatives like syndromic surveillance could help to fill this gap in our knowledge of SARS-CoV-2 local transmission and build on the currently favored methods of hospital-based surveillance. Alternatives to resource-intensive testing are especially important for detecting outbreaks in low-resource settings where diagnostic testing infrastructure is poor [28,29]. This includes low- and middle-income countries as well as impoverished communities in the United States.

Conclusions

Finally, we emphasize that the estimates of our regressions, both linear and negative binomial, should not necessarily be interpreted as definite predicted changes in the rate of SARS-CoV-2. Our goal was not to build a predictive model but to instead investigate associations between the outcome, current levels of SARS-CoV-2 cases and test positivity, and recent measures of CLI. We focused on the statistically significant lagged CLI rate variables that demonstrate changes in CLI rates can lead SARS-CoV-2 rates by a few days, as confirmed by both the negative binomial and linear models in the county setting. The university setting models the advantage of an enforced daily health questionnaire whose measures of CLI and exposure to CLI appear to track on-campus SARS-CoV-2 positivity in real time. The strongest takeaway from our results is that the two variables, CLI and SARS-CoV-2 positivity, are correlated, and CLI should be further explored as a low-resource way to monitor the risk of SARS-CoV-2 in the absence of robust testing. Our results support the push to integrate CLI symptom reporting into the routine services of health practitioners as a potentially easy and cost-effective approach to monitoring SARS-CoV-2 transmission. This forms a foundation for future research that should further characterize the relationship between COVID-19 symptoms and SARS-CoV-2 positivity rates.

Acknowledgments

This project was supported by the David and Lucile Packard Foundation (#2021-72608). MO was supported by the Cornell University Office of the Provost. We would like to thank all of the health care practices in Tompkins County who contributed to the COVID-19–like illness syndromic surveillance program.

Data Availability

Researchers who wish to request access to the Cornell University data may contact Cornell's Office of the Vice President for Research and Innovation (vp_research@cornell.edu), which oversees research access to Cornell COVID-19 surveillance data. Data from the Tompkins County COVID-19–like illness syndromic surveillance program can be requested from Tompkins County Whole Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary materials.

[DOCX File, 589 KB - publichealth_v10i1e54551_app1.docx ]

References

10. Public health - Seattle & King County syndromic surveillance summary. King County Health Department. URL: https://tableaupub.kingcounty.gov/t/Public/views/KingCountySyndromicSurveillanceDashboard/SyndromicSurveillanceData?%3Aembed=y&%3Atoolbar=n&%3Adisplay_count=n&%3AshowShareOptions=n&%3AshowVizHome=n [accessed 2023-03-28]


Abbreviations

CDC: Centers for Disease Control and Prevention
CLI: COVID-19–like illness
ED: emergency department
pp: percentage point
TCHD: Tompkins County Health Department
Using EpiCore to Enable Rapid Verification of Potential Health Threats: Illustrated Use Cases and Summary Statistics

Nomita Divi¹, MSPH; Jaś Mantero¹, MD, PhD; Marlo Libel¹, MD, MPH; Onicio Leal Neto¹², PhD; Marinanicole Schultheiss³, MPH; Kara Sewalk³, MPH; John Brownstein³⁴, PhD; Mark Smolinski¹, MD, MPH

¹Ending Pandemics, San Francisco, CA, United States
²Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States
³Computational Epidemiology Lab, Boston Children’s Hospital, Boston, MA, United States
⁴Harvard Medical School, Harvard University, Boston, MA, United States

Corresponding Author:
Nomita Divi, MSPH
Ending Pandemics
870 Market Street
San Francisco, CA, 94201
United States
Phone: 1 6173591733
Email: nomita@endingpandemics.org

Abstract

Background: The proliferation of digital disease-detection systems has led to an increase in earlier warning signals, which subsequently have resulted in swifter responses to emerging threats. Such highly sensitive systems can also produce weak signals needing additional information for action. The delays in the response to a genuine health threat are often due to the time it takes to verify a health event. It was the delay in outbreak verification that was the main impetus for creating EpiCore.

Objective: This paper describes the potential of crowdsourcing information through EpiCore, a network of voluntary human, animal, and environmental health professionals supporting the verification of early warning signals of potential outbreaks and informing risk assessments by monitoring ongoing threats.

Methods: This paper uses summary statistics to assess whether EpiCore is meeting its goal to accelerate the time to verification of identified potential health events for epidemic and pandemic intelligence purposes from around the world. Data from the EpiCore platform from January 2018 to December 2022 were analyzed to capture request for information response rates and verification rates. Illustrated use cases are provided to describe how EpiCore members provide information to facilitate the verification of early warning signals of potential outbreaks and for the monitoring and risk assessment of ongoing threats through EpiCore and its utilities.

Results: Since its launch in 2016, EpiCore network membership grew to over 3300 individuals during the first 2 years, consisting of professionals in human, animal, and environmental health, spanning 161 countries. The overall EpiCore response rate to requests for information increased by year between 2018 and 2022 from 65.4% to 68.8% with an initial response typically received within 24 hours (in 2022, 94% of responded requests received a first contribution within 24 h). Five illustrated use cases highlight the various uses of EpiCore.

Conclusions: As the global demand for data to facilitate disease prevention and control continues to grow, it will be crucial for traditional and nontraditional methods of disease surveillance to work together to ensure health threats are captured earlier. EpiCore is an innovative approach that can support health authorities in decision-making when used complementarily with official early detection and verification systems. EpiCore can shorten the time to verification by confirming early detection signals, informing risk-assessment activities, and monitoring ongoing events.

(JMIR Public Health Surveill 2024;10:e52093) doi:10.2196/52093

KEYWORDS
disease surveillance; surveillance; verification; early detection; epidemic intelligence, risk assessment; threat; threats; crisis; crises; outbreak; outbreaks; warning; warnings; crowdsource; crowdsourcing; surveillance; digital health; detect; detection; risk; risks
Introduction

Background

World Health Organization (WHO) leadership has recently suggested the need for collaborative surveillance within and beyond the health sector for more robust event detection, risk assessment, and response monitoring [1-3]. Furthermore, the recent quadripartite partnership among the Food and Agriculture Organization of the United Nations, the World Organization for Animal Health, the UN Environment Programme, and the WHO has noted the need for intelligence systems to use an integrated “One Health” approach to reduce the risk of ongoing and emerging threats [2].

Although early warning “signals” from such systems are leading to earlier detection and swifter responses to emerging threats, the proliferation of these systems can also generate a large volume of data that must be processed before contributing toward the early warning or risk assessment of ongoing threats [3]. This large volume of data may also result in false alarms that might propagate rumors and quickly overwhelm the surveillance infrastructure [2,4]. This risk of false alarms underscores the crucial role of human input for data curation and signal verification.

Delays in the effective response to a genuine health threat are often the result of the time to verification of a health event—determining if an unofficial reported threat is real so an appropriate response can ensue. An analysis by Chan et al [5] based on WHO-verified outbreaks reported in Disease Outbreak News noted median time from “outbreak start” to “outbreak discovery” dropped from 29.5 days (95% CI 13-59 days) to 13.5 days (95% CI 3.5-44.5 days) during that time. More recently, the time to detection has been reduced due in part to the advancement of digital epidemiology, but the task of timely verification remains a challenge [5,6]. WHO has presented data on additional delays in verifying that outbreaks are indeed real events, noting up to a 1-week delay [7]. It was the extra several days for outbreak verification that was the main impetus for creating EpiCore, a virtual multi-sectoral community of human, animal, and environmental health professionals that supports the rapid verification of potential health threats. EpiCore aims to contribute to the reduction in time to verify human, animal, or environmental health events to obtain event details within 24 hours of initiating a request for information (RFI) within a digital secure information system. The objective of this paper is to describe the potential of the crowdsourcing of information from qualified professionals through EpiCore to facilitate the verification of early warning signals of potential outbreaks and support the monitoring and risk assessment of ongoing threats.

About EpiCore

Conceptualized in 2012 by Ending Pandemics as an initial partnership between the TEPHINET (Training Programs in Epidemiology and Public Health Interventions Network) [8,9], Program for Monitoring Emerging Diseases (ProMED)–mail [10], and HealthMap of Boston Children’s Hospital, EpiCore became operational in 2016 [11]. EpiCore does not aim to replace any official verification system and is meant to be a complementary verification tool. An epidemic intelligence expert in the role of the focal point oversees the system’s operations. This individual is responsible for ensuring the system is accessible and running 24/7, providing support and training to the users of the system and ensuring privacy protocols are being followed. The focal point also oversees communications between the various users as well as the distribution of a quarterly newsletter.

Several epidemic intelligence and disease prevention and control organizations use EpiCore to confirm signals detected from primarily digital sources. These groups are referred to as “requesters” in the system. Requesters have included members from ProMED, GeoSentinel, HealthMap, the Hungarian National Association of Radio Distress-Signaling and Info-communication that operates the Emergency and Disaster Information Services, and Médecins Sans Frontières (MSF) Operational Centre Barcelona [10,12-15]. Requesters send out RFIs, a specific question sent through EpiCore about known or potential “events” to EpiCore responders (see below) for verification.

The virtual community of health professionals in EpiCore are referred to as “responders” in the system. To become an EpiCore responder, a health professional applies and must have at least two of the following qualifications: (1) a degree in public health (eg, MS, MPH, and PhD) or a related field (eg, MD, DVM, and RN); (2) health profession certification or licensure (eg, livestock officer and food inspector); (3) at least 3 years of experience in human or animal health or environmental health; (4) affiliation with a medical center, ministries or departments of health, or other health-related organizations (eg, nongovernmental organizations and private sector); and (5) successful completion of a field epidemiology training program. In the first 5 years after it became operational in 2016, EpiCore membership grew to over 3300 individuals with members spanning 161 countries. Over one-third of members identify themselves as experts spanning multiple sectors (n=1134, 34.3%). Most members (n=2807, 84.9%) list themselves as human health experts; 926 (28%) listed themselves as animal health experts, and 962 (29.1%) as environmental health experts. Responders join the network and contribute as individuals (ie, they do not represent any organization and their contributions can be provided anonymously if so desired). As an incentive to increase participation, responders have access to select resources related to infectious disease prevention and control, including web-based courses and publications.

When evidence of a health event is detected, local information is requested from EpiCore responders to rapidly provide details on the event (Figure 1). Responders near (in the district, country, or region) where events are reported are selected, in essence, providing the “eyes and ears” on the ground. The radius around the event where responders are selected depends on the density of the responders in the location of the potential event. In areas with fewer responders, the radius is increased by requesters to the country or regional level to include more responders. Through EpiCore’s secure digital platform, responders with knowledge of the event can easily and quickly provide details and expert opinions to inform event verification or ongoing event monitoring. Information provided by responders can only be seen by requesters. Other EpiCore responders are unable to
see individual responses to ensure the privacy and anonymity of the responders. However, all EpiCore requesters can view the RFIs from other requesters as well as the responses to those requests. This ensures that information silos between requesters and their organizations are minimized.

**Figure 1.** Illustration of key steps in the EpiCore verification process. RFI: request for information.

An RFI is considered verified if an EpiCore requester deems sufficient reliable additional details have been collected from the responders to confirm a reported event is true, or enough details have been collected to confirm the reported event or early warning signal is a false alarm. RFIs are considered verified if the report is confirmed to be valid or a false alarm. EpiCore requesters give the highest priority to reliable sources of information, including official statements and complementary reports or documents obtained at the locality of the reported event. Requesters create an RFI summary of verified health events, which constitutes a synopsis of the information received via the platform as well as other information publicly available that contributes to the understanding of the situation. These RFI summaries are made available on EpiCore’s public dashboard as well as via the requesters’ own dissemination channels, so anyone can have access to details on events verified by EpiCore [11,16]. Requesters are urged to close RFIs as soon as helpful information is available to ensure timely dissemination of the information to the public. If no information is received from responders, RFIs are automatically closed after 7 days.

In addition to disseminating information via the public dashboard, EpiCore is also contributing information to the Epidemic Intelligence from Open Sources platform and supports the goal of the WHO Hub for Pandemic and Epidemic Intelligence to combine information from traditional surveillance, event-based surveillance, participatory or community surveillance, and on-the-ground investigations with contextual information to generate an assessment of public health risk [17]. EpiCore is also facilitating organizations to improve their information-dissemination activities. For example, ProMED noted an increased number of responses to their RFIs about health events on the EpiCore platform compared to requests sent through their network [10,11].

**Methods**

**Study Design**

Data on the EpiCore platform from January 2018 to December 2022 were analyzed to evaluate key descriptive statistics, as well as to identify use cases illustrative of the various utilities of the system. All RFIs created over the 5 years were analyzed to calculate the annual mean and median time to reply to an RFI. Annual RFI response rates and verification rates were also calculated. RFIs that received a response, regardless of whether the information was considered useful or not, were categorized as “RFIs with a response.” The response rate was calculated as the number of RFIs with a response as a percentage of the total number of RFIs to indicate members’ engagement on the platform. All RFIs that requesters deem to have sufficient and useful information through responder contributions with additional information to confirm an event or to debunk misinformation or disinformation were categorized as “verified.” The verification rate was calculated as the number of RFIs that the requesters have classified as verified as a percentage of the total number of RFIs with a response. This rate informs the proportion of RFIs obtaining information from EpiCore members that were deemed useful in the verification of a reported event.

**Ethical Considerations**

This study does not constitute human subject research as it is a descriptive analysis of a system. The data collected were limited to the time of information flow and summarization of information about health events to describe EpiCore. The data set has no personal identifiers. This rationale is consistent with the Harvard University policies on human subjects research [18].
Results
EpiCore Platform and Statistics
During the observed period, 622 RFIs were sent out globally. Of these RFIs, 485 (78%) were related to health events regarding human exposure, 121 (19.4%) to health events regarding animal exposure, 12 (1.9%) to health events with human and animal exposure, and 4 (0.6%) to environmental health events. A total of 398 RFIs received a response during the observed period. The mean and median response times are indicated in Table 1.

Table 1. EpiCore key summary statistics on all RFIs issued by year between 2018 and 2022.

<table>
<thead>
<tr>
<th>Time period</th>
<th>Total RFIs sent, n</th>
<th>RFIs with a response, n</th>
<th>Response time (h), mean (SD)</th>
<th>Response time (h), median (IQR)</th>
<th>Response rate (%)</th>
<th>Reacted within 24 h, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>205</td>
<td>134</td>
<td>19.5 (40.6)</td>
<td>4.2 (16.5)</td>
<td>65.4</td>
<td>108 (80.6)</td>
</tr>
<tr>
<td>2019</td>
<td>167</td>
<td>100</td>
<td>17.1 (35.1)</td>
<td>5.2 (15.8)</td>
<td>59.9</td>
<td>81 (81)</td>
</tr>
<tr>
<td>2020</td>
<td>96</td>
<td>61</td>
<td>12.1 (26.24)</td>
<td>2.2 (6.9)</td>
<td>63.5</td>
<td>54 (88.5)</td>
</tr>
<tr>
<td>2021</td>
<td>106</td>
<td>70</td>
<td>17.4 (31.6)</td>
<td>4.2 (14.3)</td>
<td>66</td>
<td>57 (82.4)</td>
</tr>
<tr>
<td>2022</td>
<td>48</td>
<td>33</td>
<td>10.1 (26.9)</td>
<td>1.6 (6.2)</td>
<td>68.8</td>
<td>31 (94.0)</td>
</tr>
</tbody>
</table>

*RFI: request for information.*

Figure 2 illustrates the locations of the RFIs with a response, demonstrating the geographic distribution of the RFIs over the 5 years. A majority of the RFIs with a response received responses within 24 hours (n=334, 84%). Furthermore, 247 (62.1%) RFIs of those that received a response (n=398) were deemed to have sufficient and useful information through responder contributions to verify the reported event and were summarized on the public dashboard. The annual verification rate has increased each year from 2018 to 2022 ranging from 50.7% to 84.8% (Table 2). The response rate increased from 65.4% in 2018 to 68.8% in 2022.

Figure 2. Map of frequency of EpiCore RFIs by location of potential event between 2018 and 2022. RFIs represented in this map are generated by EpiCore requesters for additional input for verification by EpiCore responders. The map is not a representation of all potential health events during the observed period, but is limited to the subset of events identified by the requesters on EpiCore. RFI: request for information.
EpiCore plays a crucial role in reducing the time to verification by confirming early detection signals, informing risk assessment activities, and monitoring ongoing events. Gathering timely information is pertinent both during the early days of a potential outbreak and for continued situational awareness. EpiCore provides access to a network of health experts who can share information about a specific event in response to an RFI. Information provided by members of this trusted network is valuable, including knowing how a disease is spreading within a country, if an outbreak has crossed borders, the epidemiological characteristics, severity and transmissibility, and laboratory results. The following case studies highlight the various uses of EpiCore.

Case Study 1: Ruling Out False Alarms

Unnecessary alarms can result in wasted valuable resources in responding to essentially a nonevent. EpiCore members provide information to help rule out false alarms that may circulate through various media outlets, especially during the early days of a potential infectious disease outbreak. For example, EpiCore members provided timely information during an uptick in rash illness observed in November 2021 among fishermen in Senegal as reported in the local media. EpiCore responders provided information within 24 hours of the RFI being sent to those within the geographical area of the reported “event.” Their responses clarified that the health conditions were likely caused by environmental factors and not an infectious disease [19].

Case Study 2: Real Time Details on an Evolving Event

In another situation in October 2020, local media in South Sudan reported 3 sudden deaths in Western Bahr el Ghazal State in patients presenting with hemorrhagic symptoms, raising fears of Ebola. Responders confirmed the occurrence of the event and shared additional specifics that also quoted national health authorities. They noted that the 3 reported cases were buried before the arrival of the rapid response team and samples could not be taken for further laboratory investigation. It was also shared, however, that 3 more cases presenting with similar symptoms were still alive in the same area and tested negative for Ebola and other hemorrhagic fever viruses. The RFI was summarized, documenting the lack of laboratory results for the first cases and reporting that additional cases tested negative for hemorrhagic fever viruses, hence quelling the fears of Ebola [20].

Case Study 3: Tracking an Unknown Pathogen

In an event that garnered the world’s attention, local media reported on December 31, 2019, that health authorities in Wuhan (Hubei, China) were investigating an unexplained cluster of severe respiratory disease that had sickened at least 27 people within a short period. A viral pathogen was suspected, raising rumors of severe acute respiratory syndrome on several media outlets. It was also suggested that some cases were linked to a seafood market in Wuhan city. The same cluster of an unknown illness was noted in a web-based report by Hong Kong’s Centre for Health Protection (December 31, 2019). EpiCore responders in proximity to Wuhan (n=36) were sent an RFI to provide any additional details to document the reported event and provide information on the level of local spread. Initial responses were provided within 7 hours, confirming the event’s occurrence. Figure 3 illustrates the timeline of key responses received for this RFI. Among the received responses, 1 responder reported that the seafood market in Wuhan also sold wild animals, noting that the risks of zoonotic disease were significant. On January 2, 2020, responders mentioned that local authorities suspected that the responsible pathogen was likely a known coronavirus able to cross species (spillover). In the following days, the network noted that around 60 pneumonia cases had been identified at Jinyintan Hospital in Wuhan. The network also provided preliminary information on at least 15 laboratory-diagnosed cases with a new strain of coronavirus (influenza, avian influenza, adenovirus, SARS-CoV-1, and MERS [Middle East Respiratory Syndrome] were ruled out). On January 10, members shared information about the first occurrence of cases in a different city (Jingmen, the same province as Wuhan). In addition, information was shared about the first cases with no known or substantiated exposure to the Wuhan market. Responders also noted a particular concern about the proximity of this outbreak to the Chinese New Year with the vast exodus of people from Wuhan to rural areas to come in the following weeks. The RFI was summarized for the public dashboard and further shared on EpiCore’s first quarterly newsletter in 2020 [21,22].

Table 2. EpiCore key statistics on all RFIs with a response by year between 2018 and 2022.

<table>
<thead>
<tr>
<th>Time period</th>
<th>RFIs with a response, n</th>
<th>RFIs verified, n</th>
<th>Verification rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>134</td>
<td>68</td>
<td>50.7</td>
</tr>
<tr>
<td>2019</td>
<td>100</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>2020</td>
<td>61</td>
<td>37</td>
<td>60.7</td>
</tr>
<tr>
<td>2021</td>
<td>70</td>
<td>57</td>
<td>81.4</td>
</tr>
<tr>
<td>2022</td>
<td>33</td>
<td>28</td>
<td>84.8</td>
</tr>
</tbody>
</table>

аRFI: request for information.
Case Study 4: Supporting Decision-Making

In August 2018, MSF Operational Center in Spain [23] received a report about several human deaths and an illness from an unknown disease in Douna, Mali. All cases experienced an extreme burning or heating feeling in the legs with edema often followed by “blackening” of the skin. MSF issued an RFI on EpiCore to obtain details to help with their risk assessment. Responders confirmed the occurrence of this event, which had been observed for several months. They also shared the investigation report from local health authorities that indicated that the cases had likely had acute severe malnutrition (scurvy and beriberi) due to a significant food shortage in the affected communities. MSF used the information provided by the responders for risk assessment purposes [24].

Case Study 5: New and Evolving Presentation of Disease

EpiCore has also proven pertinent for monitoring diseases involving multiple countries. In May 2021 for example, the incidence of a rare fungal infection, mucormycosis, was observed to be increasing across India in patients with COVID-19; EpiCore responders from countries with historically reported mucormycosis were asked to provide their professional opinion and share clinical details observed in their own countries. Several responders from around the world provided details on the incidence of COVID-19–associated mucormycosis (CAM) and shared their clinical perspective on the increased incidence in certain parts of the world. Some responders in India noted that this was likely a result of a misuse of steroids during the treatment of such patients. This theory was also supported by responders from other countries. In Pakistan, several cases of mucormycosis and other fungal infections in patients with COVID-19 were attributed to the combination of viral infection, use of steroids, uncontrolled diabetes, and a long stay in the intensive care unit. In Australia, mucormycosis increases were not documented; however, several cases of another fungal infection were reported in patients with COVID-19. In the United Kingdom, sporadic CAM cases had been reported in patients with COVID-19. EpiCore member reports revealed the incidence of CAM was increasing in geographies across the globe and not limited to the occurrence in India as initially noted by the media [25,26].

Discussion

Principal Results

EpiCore provides an innovative mechanism to increase the specificity of signals of events detected across the globe through epidemic intelligence activities from several sources. Over half of all RFIs received a response, and of those, virtually all received responses within 24 hours. Response rates and verification rates increased year by year during the observed period.

Limitations

Although EpiCore has representation in 161 countries, several areas of the globe are without any responders or have suboptimal numbers. The limitation of responder coverage consequently reduces the number of responses an RFI may receive. To circumvent this situation, RFIs are sent to responders in neighboring geographies when none are in the geography related to the RFI or it has very low numbers of EpiCore members. Currently, RFI generation is limited by the number of requesters, their availability at the time, and their subjective assessment of ongoing multiple global threats. Increasing the number of
requestors can help alleviate some of these challenges. Additionally, because most EpiCore responders have human expertise, we recognize that this may hinder the speed with which RFIs about animal and environmental health events are verified. Concerted efforts will continue to recruit responders with animal and environmental health expertise to ensure EpiCore can provide timely and useful responses to RFIs across all One Health sectors. The timely dissemination of information provided via the RFI summary on the public dashboard is a priority for EpiCore. However, the tradeoff between the speed at which the RFI summary is published on the dashboard and the quantity and specificity of information received from responders by leaving the RFI open longer is recognized. To counter this issue, verified RFIs that are made available on the public dashboard can be updated on an ongoing basis with additional information rather than delaying the summary until all details have been received.

Looking Ahead

EpiCore continues to foster a collaborative approach to support the verification of potential early warning signals and provide information relevant for risk assessment purposes to the growing number of organizations focused on the use of data for epidemic and pandemic intelligence in the future. The ability of all requestors on the EpiCore platform to view all requests and responses regardless of who sends the requests is one way to reduce the duplication of efforts and siloes of information that are common among groups engaged in epidemic and pandemic intelligence.

EpiCore aims to achieve an average time of response to an RFI of less than 24 hours in all regions of the world. Increasing the number of members and their geographic distribution will further enable this goal. Recruitment efforts will continue in the coming years with a particular focus on regions of the world with the fewest EpiCore members. Mechanisms to motivate responder engagement will be considered, including nonfinancial incentives such as increased training opportunities and access to curated web-based trainings and scientific publications. EpiCore will continue to share quarterly digital newsletters to document and disseminate the value of member contributions and provide updates on recently verified events. Additionally, EpiCore will expand its incorporation into human, animal, and environmental training curricula to ensure the next generation of those on the front lines are better positioned to expedite global pandemic and epidemic intelligence.

Certain functionalities within EpiCore may benefit from the application of generative AI tools. The use of machine learning capabilities will be explored and tested against human-curated RFIs to ensure validity and accuracy. Digital disease-detection platforms using machine learning models to identify and aggregate reports of health events through informal sources (eg, social media) [27], can also be leveraged to identify health events for which an RFI should be sent.

Conclusion

The global demand for data to improve disease prediction, prevention, and control continues to grow. A recent paper authored by the WHO underlines that being better prepared for future pandemics and epidemics will require increased collaboration among stakeholders and investment in collective abilities to detect and understand public health risks [3]. The use of EpiCore by multiple sectors can contribute toward furthering a multi-sectoral approach to epidemics and pandemics (ie, One Health Intelligence). With members spanning human, animal, and environmental health, EpiCore can help the expeditious verification of any health event. EpiCore will continue to provide information for the verification of health events to complement traditional systems and will evolve as needed to better address the global need for situational awareness and risk assessment.

Acknowledgments

We would like to recognize the efforts of all of our requestors and responders without whom EpiCore would not be possible. We also recognize the efforts and early partnership with ProMED, HealthMap, and TEPHINET in the early development of EpiCore. We thank Carrie McNeil, Jay Atanda, and Elizabeth Aaron for reviewing a draft of this paper. All work by the authors was supported by the Ending Pandemics, a project of the Tides Center. No generative AI was used for the writing of this paper or analysis of the data.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References


8. Who we are. Training Programs in Epidemiology and Public Health Interventions Network. URL: https://www.tephinet.org/who-we-are [accessed 2023-08-03]


14. Radio Distress-Signaling and Infocommunications-Emergency and Disaster Information Service. URL: https://rsoe-edis.org [accessed 2023-08-03]

15. Medicos Sin Fronteras. URL: https://www.msf.es [accessed 2023-08-03]


17. The WHO Hub for Pandemic and Epidemic Intelligence. URL: https://pandemichub.who.int [accessed 2023-08-03]


Abbreviations

CAM: COVID-19–associated mucormycosis
MERS: Middle East respiratory syndrome
MSF: Médecins Sans Frontières
ProMED: Program for Monitoring Emerging Diseases
RFI: request for information
The Role of Trust as a Driver of Private-Provider Participation in Disease Surveillance: Cross-Sectional Survey From Nigeria

Ellen MH Mitchell1*, MA, PhD; Olusola Adedeji Adejumo2,3*, MPH, MD; Hussein Abdur-Razzaq4, MPH, MD; Chidubem Ogbudebe5, MS, MPH; Mustapha Gidado6, MSc, MPH, MD, PhD

1Mycobacterial Diseases and Neglected Tropical Diseases Unit, Department of Public Health, Institute for Tropical Medicine, Antwerp, Belgium
2Mainland Hospital, Yaba Lagos, Nigeria
3Department of Community Health and Primary Health Care, Lagos State University Teaching Hospital, Ikeja, Lagos, Nigeria
4Health Research Unit, Directorate of Planning, Research, and Statistics, Lagos Ministry of Health, Lagos, Nigeria
5KNCV Tuberculosis Foundation Nigeria, Abuja, Nigeria
6KNCV Tuberculosis Foundation, The Hague, Netherlands
*these authors contributed equally

Corresponding Author:
Ellen MH Mitchell, MA, PhD
Mycobacterial Diseases and Neglected Tropical Diseases Unit
Department of Public Health
Institute for Tropical Medicine
155 Nationalestraat
Roche
Antwerp, 2000
Belgium
Phone: 31 634790509
Email: emitchell@itg.be

Abstract

Background: Recognition of the importance of valid, real-time knowledge of infectious disease risk has renewed scrutiny into private providers’ intentions, motives, and obstacles to comply with an Integrated Disease Surveillance Response (IDSR) framework. Appreciation of how private providers’ attitudes shape their tuberculosis (TB) notification behaviors can yield lessons for the surveillance of emerging pathogens, antibiotic stewardship, and other crucial public health functions. Reciprocal trust among actors and institutions is an understudied part of the “software” of surveillance.

Objective: We aimed to assess the self-reported knowledge, motivation, barriers, and TB case notification behavior of private health care providers to public health authorities in Lagos, Nigeria. We measured the concordance between self-reported notification, TB cases found in facility records, and actual notifications received.

Methods: A representative, stratified sample of 278 private health care workers was surveyed on TB notification attitudes, behavior, and perceptions of public health authorities using validated scales. Record reviews were conducted to identify the TB treatment provided and facility case counts were abstracted from the records. Self-reports were triangulated against actual notification behavior for 2016. The complex health system framework was used to identify potential predictors of notification behavior.

Results: Noncompliance with the legal obligations to notify infectious diseases was not attributable to a lack of knowledge. Private providers who were uncomfortable notifying TB cases via the IDSR system scored lower on the perceived benevolence subscale of trust. Health care workers who affirmed “always” notifying via IDSR monthly reported higher median trust in the state’s public disease control capacity. Although self-reported notification behavior was predicted by age, gender, and positive interaction with public health bodies, the self-report numbers did not tally with actual TB notifications.

Conclusions: Providers perceived both risks and benefits to recording and reporting TB cases. To improve private providers’ public health behaviors, policy makers need to transcend instrumental and transactional approaches to surveillance to include building trust in public health, simplifying the task, and enhancing the link to improved health. Renewed attention to the “software” of health systems (eg, norms, values, and relationships) is vital to address pandemic threats. Surveys with private providers may overestimate their actual participation in public health surveillance.
Introduction
Nigeria experienced five major infectious disease outbreaks during 2017, representing an unprecedented crisis for the public health system that laid bare many of the intersectoral collaboration gaps that hamper an effective public health response [1]. The existence of parallel reporting systems, authorities, and periods, along with variable case definitions challenge even the most well-intentioned and highly motivated health care workers to comply [2,3]. The globally networked, economically and culturally dynamic hub of Lagos, Nigeria, has long been identified as a place where timely information on emerging pathogens, pharmacovigilance, and infectious disease surveillance is crucial to the country’s public health [4].

Tuberculosis (TB) notification and cohort analysis are illustrative of the classical surveillance practice of a stigmatized condition worldwide and a bellwether of a country’s capacity for public health surveillance. Although TB service provision and reporting by the private sector is long recognized as an essential component of an effective TB program, it is still poorly theorized [5,6]. Many efforts to engage the private sector have been directive or transactional, driven by an incomplete or simplistic understanding of how private providers think [7]. The COVID-19 and mPox pandemics have heightened attention to the issues of trust and mistrust in public health authorities, whereas their centrality in TB surveillance has yet to be quantified [8,9].

Efforts to incentivize the private sector to render quality TB care and contribute to TB notification have intensified in recent years [10-13]. Many models are designed around logical inferences, but often without compelling evidence of efficacy to distinguish them [14]. Many public-private mix (PPM) models are transactional and/or directive, with a focus on resource transfer and regulatory oversight [7]. Although advocacy, additional professional society engagement, subsidized drugs, coordinating bodies, the introduction of advanced TB diagnostics, financial incentives, stricter penalties, and supervision interventions have been shown to improve notification initially, the gains are often modest and challenging to sustain [12]. Local stakeholders thus requested a study to generate insights into how to best set priorities among the diverse solutions and how best to distribute scarce drugs, diagnostics, staffing, and supervision resources.

We posited that the act of notifying a TB case is predicated on a set of expectations about systems, risks, rewards, penalties, and costs [15]. Surveillance assumes that specific types of patient information are available to be recorded (eg, test results and treatment outcomes) and that certain data formats (eg, registers and electronic platforms) exist that are accessible and intelligible. Notification obligations assume a certain knowledge of the legal and technical process of recording and reporting. Crucially, such an integrated approach assumes a set of values and willingness to contribute to surveillance as a public service [2]. For the desired public health participation to occur, potential contributors to systems may need to perceive the benefits as outweighing the risks [15-17]. As shown in Figure 1, we grouped the hypothesized influences on infectious disease reporting behavior into four categories: structural context, hardware, tangible software, and intangible software. This framework highlights how an enabling environment includes both mechanical and instrumental elements (ie, the “hardware”) and more relational or perspectival elements (ie, the “software”) [7,18].
Insights into the underlying reasons for facility noncompliance with surveillance requirements are needed if systems are to be reengineered in ways that make it worthwhile for providers to participate in TB notification. Indeed, early warning systems and antibiotic stewardship efforts are equally reliant upon the voluntary contribution of time-sensitive facility data. Mutual perceptions of competence, benevolence, and integrity form a crucial part of a surveillance system’s “software,” yet rarely receive as much focus as the “hardware” [7,18-20].

To improve private providers’ willingness to contribute to surveillance systems in Lagos, State Ministry of Health (SMOH) stakeholders sought assistance to answer the following policy questions: Given multiple systems for TB notification, how do private providers decide where and when to contribute their data to different surveillance systems? What are “unengaged” private providers’ concerns about notification? Is noncompliance with notification specific to TB or are they broadly noncompliant with infectious disease notification obligations? How does trust in the SMOH affect willingness to engage in public health surveillance?

The Nigerian Integrated Disease Surveillance and Response (IDSR) system is among one of the most dynamic and well-characterized surveillance systems in Africa [21]. Efforts to build and maintain rigorous surveillance systems to tackle emerging threats with pandemic potential have included policy, training, legislation, and validation exercises [3,22-28]. Over time, the system has evolved from a narrow Disease Surveillance and Notification Office (DSNO) to the comprehensive World Health Organization (WHO)–recommended IDSR of today.

The formal engagement of the Nigerian private sector in the PPM program for nationally recognized TB management commenced in 1993 [29]. PPM in Nigeria has been standardized, protocolized, and cautiously implemented by the national TB program [29,30]. Training and equipping health facilities with drugs and reagents were the most common forms of PPM engagement in Nigeria initially. Guidelines were updated approximately every 10 years. The use of incentives, performance-based finance, peer-led, and digital models have been employed and results have varied. Investments to increase private-provider collaboration in TB control in Lagos have been intense and involved multimethod approaches [12,30-34].

**Methods**

**Study Design**

We undertook a cross-sectional survey of the persons in charge of private health facilities without an ongoing relationship to the TB program, because these sites are considered to be those most likely to treat TB without reporting it. This assessment was part of a TB inventory study conducted to estimate the magnitude of underreporting [35,36]. Private facilities previously trained and equipped by the TB program were studied separately.
Study Setting
Lagos State is in Southwest Nigeria, the commercial heart of Nigeria and home to the country’s most dynamic private sector. Although Lagos is the smallest state in Nigeria in terms of land mass, the population estimates range from 13 to 21.5 million. The population density exceeds 5000 individuals/km², roughly 25 times the national average population density of 226 individuals/km². More than 65% of Lagos’ population lives below the poverty line. Lagos has 2.5 private health facilities per 10,000 people, making it one of the more complex health systems of any megacity [37]. The health facilities open, operate, and then cease operations or move locations dynamically [37]. Faith-based and private not-for-profit health facilities represent a smaller proportion (1.7%) of the Lagosian health sector than in other Nigerian states [38].

Sampling
Probability proportional to size sampling was conducted among private health facilities with no documented engagement with the TB program. Private facilities provided with TB registers and trained by the national TB program were recruited for a parallel study with distinct aims. Full details on health facility sampling, pilot testing, data collection, data management, and quality assurance have been described previously [35].

Recruitment
Potential participants were contacted by telephone for recruitment appointments. Written information and letters from the TB program and ethical review board were shared to establish the factual basis of the visit.

One interview was conducted at each of the participating facilities. Inclusion criteria for survey respondents were two-fold: (1) health care workers with influence over the completion of TB case notification and (2) able to provide individual informed consent (e.g., over 18 years of age). Semistructured interviews were administered on site using paper questionnaires by trained interviewers (see Multimedia Appendix 1).

At “unengaged” private facilities, the following 10 issues were explored: (1) awareness of the mandatory reporting of TB in Nigeria; (2) perceived trustworthiness, competence, and beneficence of the SMOH; (3) attitude toward specific aspects of TB notification processes; (4) attitude toward specific PPM engagement incentives (clinical training, drugs, reagents); (5) self-reported participation in the State TB & Leprosy Control Program (STBLCP) and/or the IDSR, also known as the DSNO; (6) self-reported challenges with disease reporting via IDSR (closed-ended); (7) self-reported reasons for nonreporting diseases (closed-ended); (8) self-reported participation in either reporting system (STBLCP or IDSR); (9) self-reported reasons for nonreporting diseases (closed-ended); (10) willingness to engage (and prior TB engagement experiences) with the STBLCP.

Study Measures
The scale for Citizen Trust in Government Organizations (CTGO), a validated 11-item scale of public perceptions toward government institutions, was adapted to capture Nigerian private practitioners’ trust in public health and disease surveillance systems [39]. The CTGO scale measures three dimensions: (1) perceived competence of public health authorities, indicating the extent to which a provider perceives a (government) organization to be capable, effective, skillful, and professional (4 items); (2) perceived benevolence, indicating the extent to which a private provider perceives a (government) organization to care about the welfare of the public and to be motivated to act in the public interest (3 items); and (3) perceived integrity, indicating the extent to which a provider perceives a (government) organization to be sincere, tell the truth, and fulfill its promises (4 items). Pilot testing of the questionnaire occurred in 5 facilities. Cronbach α and the intraclass correlation coefficient of the scale were used to gauge validity and internal consistency. Survey items with poor construct validity during pretesting were deleted. Survey items were reduced from a total of 76 to 53 and worded via piloting to improve validity and acceptability.

SPSS (IBM) version 25 and the R psych package were used for statistical analyses. The magnitude and variance of the responses were examined to identify central tendencies and outliers were considered for further exploration. Data were summarized as percentages, means, and medians, and Student t tests were used to compare mean scores. For statistical tests, P < .05 was considered statistically significant. A regression model was developed to predict the binary outcome of infectious disease case notification using IDSR.

Ethical Considerations
The study protocol was reviewed and approved by the Health Research and Ethics Committee of the Lagos State University Teaching Hospital (registration number 04/04/2008). Participation was voluntary and providers could consent to zero, partial, or full patient data access. Noncompliance with notification obligations was kept confidential and is described in ways to preclude deductive disclosure. After the data sets were linked, all personal and geographic identifiers were removed. A small monetary incentive (US $5) was offered for participation.

Results
Sample Characteristics
There were 278 representatives surveyed from private facilities that did not report TB. They ranged in age from 21 to 81 years, with a mean age of 46 (SD 0.8) years and an average of 18.2 (SD 0.8) years of clinical practice. Among the 278 respondents, 62.9% (n=175) were men and 36.3% (n=101) were women. Among the private facilities represented, 40.6% were at the primary level, 53.6% at the secondary level, and 4.7% unclassified; 94.6% (n=263) of the facilities were for-profit and 5% (n=14) were faith-based (see Table 1). Among the total 294 representatives contacted, 278 (84.2%) consented to participate. The recruited sample of health care workers was in line with the intended sample in terms of the total sample size, local government area distribution, and facility level [35].
Table 1. Sociodemographic characteristics of participants (N=278).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>26 (9.4)</td>
</tr>
<tr>
<td>30-39</td>
<td>58 (20.9)</td>
</tr>
<tr>
<td>40-49</td>
<td>53 (19.1)</td>
</tr>
<tr>
<td>50-59</td>
<td>73 (26.3)</td>
</tr>
<tr>
<td>&gt;60</td>
<td>67 (24.1)</td>
</tr>
<tr>
<td>No response</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>46.3 (12.9)</td>
</tr>
<tr>
<td>Range</td>
<td>12-81</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>175 (62.9)</td>
</tr>
<tr>
<td>Woman</td>
<td>101 (36.3)</td>
</tr>
<tr>
<td>No response</td>
<td>2 (0.7)</td>
</tr>
<tr>
<td><strong>Type of practice, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>General practitioner</td>
<td>247 (88.7)</td>
</tr>
<tr>
<td>Specialist</td>
<td>31 (11.2)</td>
</tr>
<tr>
<td><strong>Facility level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>113 (40.6)</td>
</tr>
<tr>
<td>Secondary</td>
<td>149 (53.6)</td>
</tr>
<tr>
<td>Unspecified</td>
<td>13 (4.7)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td><strong>Type of facility, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>For-profit</td>
<td>263 (94.6)</td>
</tr>
<tr>
<td>Faith-based</td>
<td>14 (5.0)</td>
</tr>
<tr>
<td>No response</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td><strong>Years of practice, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>79 (28.4)</td>
</tr>
<tr>
<td>10-19</td>
<td>47 (16.9)</td>
</tr>
<tr>
<td>20-29</td>
<td>58 (20.9)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>58 (20.9)</td>
</tr>
<tr>
<td>No response</td>
<td>36 (12.9)</td>
</tr>
</tbody>
</table>

Knowledge of Obligations and Self-Reported Infectious Disease Reporting Behavior

As shown in Table 2, over three-quarters of the respondents were aware of the obligation to notify TB cases. A minority (13.5%) reported having been notified a disease recently (within weeks of the survey), 58.2% had last reported a disease within months of the survey, and 28.4% last reported it within years of the survey. Among those surveyed, nearly one-quarter reported on-site capacity to diagnose TB (Table 2).
Table 2. Participants’ behaviors, challenges, and recommendations for improved disease notification in Lagos, Nigeria.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility had capacity to diagnose TB&lt;sup&gt;a&lt;/sup&gt; (n=276)</td>
<td>203 (73.6)</td>
</tr>
<tr>
<td>Facility had capacity to provide TB treatment (n=276)</td>
<td>126 (45.7)</td>
</tr>
<tr>
<td>Aware of obligation to report TB (n=277)</td>
<td>215 (77.6)</td>
</tr>
<tr>
<td>Evidence of provision of TB treatment in the last 12 months (N=278)</td>
<td>33 (11.9)</td>
</tr>
<tr>
<td><strong>Most recent notification (n=141)</strong></td>
<td></td>
</tr>
<tr>
<td>Weeks ago</td>
<td>19 (13.5)</td>
</tr>
<tr>
<td>Months ago</td>
<td>82 (58.2)</td>
</tr>
<tr>
<td>Years ago</td>
<td>40 (28.4)</td>
</tr>
<tr>
<td>Comfortable reporting TB patients to the LGA&lt;sup&gt;b&lt;/sup&gt; (n=267)</td>
<td>242 (90.6)</td>
</tr>
<tr>
<td>Facility has ever notified about TB (n=258)</td>
<td>132 (51.2)</td>
</tr>
<tr>
<td><strong>Entity notified (n=151)</strong></td>
<td></td>
</tr>
<tr>
<td>IDSR&lt;sup&gt;c&lt;/sup&gt; (DSNO&lt;sup&gt;d&lt;/sup&gt;)</td>
<td>74 (49.0)</td>
</tr>
<tr>
<td>TB program</td>
<td>75 (49.7)</td>
</tr>
<tr>
<td>Can’t remember</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>State TB program ever offered TB training (n=270)</td>
<td>63 (23.3)</td>
</tr>
<tr>
<td>Any training by another TB organization (n=272)</td>
<td>37 (13.6)</td>
</tr>
<tr>
<td><strong>Self-reported monthly IDSR (DSNO) notification (n=261)</strong></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>128 (49.0)</td>
</tr>
<tr>
<td>Sometimes</td>
<td>35 (13.4)</td>
</tr>
<tr>
<td>Never</td>
<td>98 (37.5)</td>
</tr>
<tr>
<td><strong>Frequency of challenges to IDSR (DSNO) notification (n=187)</strong></td>
<td></td>
</tr>
<tr>
<td>Seldom</td>
<td>116 (62.0)</td>
</tr>
<tr>
<td>Often</td>
<td>33 (17.6)</td>
</tr>
<tr>
<td>Don’t know</td>
<td>38 (20.3)</td>
</tr>
<tr>
<td><strong>Type of challenges in notification</strong></td>
<td></td>
</tr>
<tr>
<td>Lack of time to fill out forms (n=146)</td>
<td>35 (24.0)</td>
</tr>
<tr>
<td>Unavailability of forms (n=136)</td>
<td>20 (14.7)</td>
</tr>
<tr>
<td>No data to fill out forms (n=136)</td>
<td>15 (11.0)</td>
</tr>
<tr>
<td>Form design very confusing (n=135)</td>
<td>10 (7.4)</td>
</tr>
<tr>
<td><strong>Suggestions to improve the IDSR (DSNO) notification (n=62)</strong></td>
<td></td>
</tr>
<tr>
<td>Provide more training in notification</td>
<td>15 (24.2)</td>
</tr>
<tr>
<td>Make reporting electronic</td>
<td>11 (17.7)</td>
</tr>
<tr>
<td>Supportive supervision</td>
<td>11 (17.7)</td>
</tr>
<tr>
<td>Simplify forms</td>
<td>9 (14.5)</td>
</tr>
<tr>
<td>Provide feedback</td>
<td>2 (3.2)</td>
</tr>
<tr>
<td>Provide incentives</td>
<td>2 (3.2)</td>
</tr>
</tbody>
</table>

<sup>a</sup>TB: tuberculosis.

<sup>b</sup>LGA: local government area.

<sup>c</sup>IDSR: Integrated Disease Surveillance Response.

<sup>d</sup>DSNO: Disease Surveillance and Notification Office.

The vast majority (90.6%) of respondents reported being hypothetically comfortable with notifying TB patients to the local government. In contrast to high levels of comfort with notification as a norm, only about half (51.2%) reported ever...
having notified a TB case. Of those who self-reported ever having notified a TB case, roughly half (49.0%) reported doing so via the IDSR (DSNO) system, while 49.7% reported doing so via the TB program and 1.3% could not recall which system they used. Among the 33 health facilities with records of treating 156 TB cases, none had been notified [36]. Self-reported participation in disease notification and disease surveillance varied and did not often align with the findings of independent verification of notification by the facility [36].

As shown in Figure 2, a majority of unnotified TB cases (31/33) were discovered in the records of health facilities of respondents who self-reported comfort with notification. TB cases were also found in the records of health facilities where respondents stated that they lacked the capacity to treat TB (8/23).

Self-reported notification behavior differed according to sociodemographic characteristics. Older health care workers with more experience were more likely to self-report comfort with notification of TB cases, having previously notified of a case, and participation in monthly disease surveillance. Women reported less comfort with notification than men (87.2% vs 92.4%) and were significantly less likely to report “always” contributing to monthly surveillance (41.5% vs 52.7%) (Table 3).

Table 3. Health care worker self-reported participation in disease surveillance by sociodemographic characteristics (N=276).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total</th>
<th>As a private medical practitioner, I am comfortable in notifying TB cases to the local government</th>
<th>This facility has previously notified a TB case to any government entity</th>
<th>This facility provides monthly reports to DSNO using the 003 form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td>Always</td>
<td>Sometimes</td>
<td>Never</td>
</tr>
<tr>
<td>Men</td>
<td>175 (63.9)</td>
<td>158 (92.4)</td>
<td>81 (50)</td>
<td>87 (52.7)</td>
</tr>
<tr>
<td>Women</td>
<td>101 (38.1)</td>
<td>82 (87.2)</td>
<td>50 (53.2)</td>
<td>39 (41.5)</td>
</tr>
<tr>
<td>Age (years), mean (SE)</td>
<td>46.3 (0.8)</td>
<td>47.0 (0.9)</td>
<td>49.2 (1.1)</td>
<td>50.4 (1.1)</td>
</tr>
<tr>
<td>Years of health care practice, mean (SE)</td>
<td>18.2 (0.8)</td>
<td>18.9 (0.9)</td>
<td>21.9 (1.2)</td>
<td>22.2 (1.1)</td>
</tr>
</tbody>
</table>

aTB: tuberculosis.

Perceived Competence, Benevolence, and Integrity of Public Health Authorities

The adapted trust items had a Cronbach α of 0.91, suggesting good internal consistency. However, the scale showed ceiling effects and a bimodal distribution, necessitating reciprocal transformation (see Figure S1 in Multimedia Appendix 2). Initially, the gender difference in trust was not explained by age or years of experience (see Figure S2 in Multimedia Appendix 2). After transformation, the distribution of the trust scores...
varied by gender and self-reported notification behavior; men tended to report more trust and more monthly IDSR notification behavior than women (Figure S3 in Multimedia Appendix 2).

Neither the trust scale nor the subscales (competence, benevolence, and integrity) were normally distributed according to the Kolmogorov-Smirnov test. The Nigerian public health authority trust scale had a 3-factor structure with loading of four competence items, three benevolence items, and four integrity items (see Figure S4 in Multimedia Appendix 2). Scores on the trust scale ranged from 11 to 55, with higher values implying greater trust. The median value was 44 (IQR 39-48) (Table 4).

A minority of health care workers expressed doubts about the trustworthiness of the SMOH to conduct disease surveillance (Figure 3). Private providers who were not comfortable notifying TB scored the SMOH lower on the benevolence subscale of trust. Health care workers who affirmed always notifying via IDSR monthly reported higher median trust in the state’s public disease control capacity and had higher median scores on all three subscales compared to those of health care workers who indicated never reporting (Table 4).

The minority of private providers who reported being uncomfortable reporting TB cases to the state were also less likely to report that the SMOH was benevolent, acting in their interests (Table 4). Private providers who reported that they did not participate in the IDSR monthly reporting system were slightly less likely to report that the SMOH was competent and effective in providing health services (91% vs 98%; P=.03) and were less likely to view the SMOH as a capable regulatory agency (81% vs 95%; P=.02). Nonparticipants in disease surveillance reported lower median scores of SMOH competence, benevolence, and integrity than those who reported “always” submitting monthly reports. There were no significant differences in attitudes among those who reported ever notifying a TB case and those who did not (Figure 3).

Among those who were not comfortable notifying TB as mandated (n=156), the reasons for discomfort with disease notification included practical, logical, strategic, and economic concerns. The most common reason (58/156, 28.9%) given for noncompliance was lack of access to the surveillance “hardware” (eg, notification forms and registers). Approximately one-quarter of TB providers were doubtful that the low volume of TB patients they treated in their facility merited mastery of the TB notification forms and procedures. In addition, approximately 16% of the respondents incorrectly believed that notification to the TB program was unnecessary if they participated in the IDSR system (Table 5).

Table 4. Private providers’ attitudes toward State Ministry of Health trustworthiness according to tuberculosis (TB) and Integrated Disease Surveillance Response (IDSR) reporting behavior.

<table>
<thead>
<tr>
<th>Scale item</th>
<th>As a private medical practitioner, are you comfortable in notifying your TB patients to the local government?</th>
<th>Self-reported monthly IDSR notification behaviora</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n=242), median (95% CI)</td>
<td>No (n=25), median (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Private providers’ trust in public health authorities (reciprocal transformation)</td>
<td>0.023 (0.023-0.023)</td>
<td>0.025 (0.023-0.029)</td>
<td>.02</td>
</tr>
<tr>
<td>Perceived competence subscale</td>
<td>4.5 (4.0-5.0)</td>
<td>4.0 (4.0-7.0)</td>
<td>.11</td>
</tr>
<tr>
<td>Perceived benevolence subscale</td>
<td>3.0 (3.0-4.0)</td>
<td>2.0 (1.0-3.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Perceived integrity subscale</td>
<td>4.0 (4.0-6.0)</td>
<td>2.0 (0.0-4.0)</td>
<td>.12</td>
</tr>
</tbody>
</table>

aParticipants responding “sometimes” were classified as missing.
Figure 3. Comparison of mean trust scores by self-reported notification intention and behaviors (N=278). IDSR: Integrated Disease Surveillance Response; TB: tuberculosis.

Table 5. Private providers’ rationales for discomfort with tuberculosis (TB) notification (n=156).

<table>
<thead>
<tr>
<th>Reason</th>
<th>Respondents, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I lack access to TB forms and/or TB registers</td>
<td>58 (28.9)</td>
</tr>
<tr>
<td>The number of TB patients treated here is too small to merit mastery of so many TB forms and procedures</td>
<td>31 (19.5)</td>
</tr>
<tr>
<td>Notifying TB to the state TB program is unnecessary because TB cases are reported to IDSRb</td>
<td>26 (15.9)</td>
</tr>
<tr>
<td>Notification does not align with my business interests</td>
<td>19 (12.0)</td>
</tr>
<tr>
<td>My reputation for patient confidentiality would be at risk</td>
<td>16 (10.2)</td>
</tr>
<tr>
<td>Lack of time to fill out reporting forms</td>
<td>14 (8.9)</td>
</tr>
<tr>
<td>I am unconvinced of the purpose/value of TB notification</td>
<td>12 (7.6)</td>
</tr>
<tr>
<td>TB forms and TB registers are confusing and complex</td>
<td>11 (7.1)</td>
</tr>
<tr>
<td>Reporting TB cases is not required in this state</td>
<td>8 (5.2)</td>
</tr>
<tr>
<td>Supervision by the TB program would be a burden</td>
<td>7 (4.5)</td>
</tr>
</tbody>
</table>

aRespondents could give more than one answer.

bIDSR: Integrated Disease Surveillance Response.

Discussion

Principal Findings

Private providers who did not notify TB cases via the disease surveillance system were younger and more likely to doubt the competence and effectiveness of public health authorities. Younger health care workers and women perceived fewer incentives to notify and indicated many obstacles to compliance. The summary of our hypothesis testing results is provided in Table S1 of Multimedia Appendix 2.

In contrast to the conclusion of a 2018 rapid assessment of TB surveillance in Nigeria, our findings show that over three-quarters (77.6%) of the “unengaged” providers fully understood that TB notification is required by law. More than half (51%) had notified TB in the past [40]. However, some private providers seemed confused by the seemingly duplicative notification mandates of the IDSR and the STBLCP.

A plurality of Lagos providers remained skeptical of the value of notification to the SMOH. Providers had low levels of trust in the state’s disease control and surveillance efforts. Our study demonstrates that lower trust scores correlate with lower participation. Such a perceived lack of effectiveness and
regulatory competence of Nigerian state public health authorities is not without precedent. Oleribe et al [16,17] also found that clinicians throughout Nigeria reported limited faith in the governmental commitment to public health and health care workers. Uchenna et al [3] identified a “bad attitude” as a barrier to completion of IDSR in Enugu state. Lafond et al [26] showed that providers often lack confidence that notification would make a positive difference. Our study extends this body of work by showing that the “software” of public health in Nigeria (ie, the values, norms, relationships, power, and mutually defined aims) are associated with mission-critical behaviors of key actors.

The development of trust needs reciprocal strengthening. Nigerian TB program stakeholders also express ambivalence regarding the competence, integrity, and motives of the private for-profit sector in TB care [29,41]. Distrust in the ethics and altruism of private providers is similarly well documented [42]. In India, Nair et al [43] found that lack of trust on the part of health authorities was a barrier to public-private TB collaboration. Some policy makers assume that private practitioners only respond to financial incentives; however, the evidence for this hypothesis is often overstated [41,44-46]. While providers who participated in this study reported a willingness to notify in exchange for free or subsidized drugs, diagnostic commodities, and patient materials, they also expressed support for nonmonetary incentives. Peer norming and social network recognition can be powerful forces in the health professions [44,45]. A majority (57.9%) of Lagos private providers stated that they would contribute to TB notification if they received professional recognition of their contributions from their medical peers and they were certain their peers were also engaged. Implementation research on these lower-cost peer-norming means to improve behavior is needed.

This study sought to understand attitudes toward participation in infectious disease surveillance using TB notification as an illustration. A unique feature of the study is the holistic measurement of complex governance trust constructs such as public health competence, benevolence, and integrity using validated scales.

Private provider “noncompliance” with public health obligations is a complex, multicausal behavioral phenomenon. Providers are influenced by peer norms around recording and reporting but also harbored doubts about the purpose and value of participation in public health surveillance. In contrast to earlier studies, Nigerian private providers did not lack knowledge of the notification obligation, nor did transactional “engagement” prompt full compliance. Although 17% of the respondents stated that electronic systems would be favored, 1 in 10 private providers had concerns about safeguarding patient confidentiality, which may impact their willingness to partner in WhatsApp groups, notification apps, and electronic registers.

In India, Thomas et al [47] found that patients’ confidentiality concerns (24%) and fear of offending patients (11%) were barriers to notification. El Emam et al [15] also found that private providers worry that disclosure of case counts could have adverse legal, ethical, financial, and regulatory consequences.

Effective engagement of private sector providers in Lagos and elsewhere will require mutual trust building, compromise, and respect. Achieving this will involve trade-offs, especially early on. Greater attention to end-user acceptability and the design of surveillance systems is paramount. While women were more likely to state that they were uncomfortable with disease reporting, none of the 156 TB cases found in the 33 facilities had been notified, highlighting the limits of PPM surveys as a method to understand notification behavior. Going forward, ethnographic research and inventory studies are needed to explore the gap between the rhetoric of self-reported notification behavior and actual notification behaviors.

As private providers’ motivations and TB capacity-building needs varied widely, offering a menu of incentives and enablers to this heterogeneous group would be a strategic approach to gain broad compliance. Some of the strategies that are being trialed to engage private sites include continuing medical education credits, vouchers for subsidized rapid molecular testing, and computerized chest-x-ray imaging, among others. Although these traditional PPM incentives (eg, training; free informational, educational, and communication materials; medical commodities; and free diagnostic tests) make sense for private providers with substantial TB caseloads, for small practitioners who will only ever treat limited TB patient volumes, sustainable options might include simpler, less-onerous, anonymous notification systems [2].

Simultaneous PPM initiatives funded by the Global Fund, United States Agency for International Development, Directorate-General for International Cooperation of the Netherlands, and the Centers for Disease Control and Prevention were implemented in Lagos during 2017-2020 [2,12,34,41,48]. Simplified private-sector TB surveillance systems have been introduced, including an Android app (STARRTB) and peer-to-peer WhatsApp notifications [13]. Efforts to improve the interoperability and integration of the IDSR and STBLCP TB notification systems have also been implemented by the WHO [2,13,49]. Although no interventions have yet addressed the issues of perceived regulatory effectiveness of public health agencies explicitly, efforts to streamline and simplify reporting burdens for private providers could contribute to increases in the perceived beneficence of public health agencies.

Nationally, the proportion of TB notifications by private providers increased from 11% in 2015 to 22% in 2021, but the addition of multiple reporting modalities raises the possibility of double counting and complicates attribution [2,12,13,41,48]. In the National Strategic Plan of the TB program, an additional US $35 million was foreseen for improving reporting in 2022. Continued appraisal of the return on investment of these diverse strategies is warranted.

One of the critical hurdles in gaining private-sector cooperation in antibiotic stewardship, surveillance, pandemic response, and other vital public health efforts is building the reputation of state public health institutions as credible, competent, and committed stewards of data. Training and deploying of apps are necessary, but likely insufficient to substantially and sustainably increase notification in the long term [13]. A possible strategy to build trust in the SMOH should be part of
any approach to boost stakeholders’ motivation to comply with notification obligations. Timely provision of valid scientific information and appreciation of the complementary strengths of private providers can help ministries of health earn the respect of private providers. Few contributors to the IDSР in Nigeria receive regular feedback or are aware of how the data are used [50]. Going forward, this “software” of surveillance needs as much attention as the equipment and human resources required to perform it.

Sensitivity on the part of TB policy makers will be required to craft a minimalist TB surveillance system that is easy to use and appropriately concise to be acceptable to reluctant providers. Private providers are numerous, but their individual TB caseloads tend to be small; therefore, it is unsurprising that they may not wish to invest time in mastery of the complex TB registers common in infectious disease surveillance. This is particularly true when providers are unsure how their TB data are to be used or whether public health gains accrue to their communities via participation in these systems. A plurality of private providers (49%) stated that they are willing to contribute to a TB notification system if the data submitted were anonymous. As unique IDs are instituted via electronic recording and reporting, the necessity of the collection of patient names should be revisited. Ways to satisfy providers’ anonymity preferences for case notification should be explored [15].

An infectious disease surveillance system acceptable to private providers would collect fewer variables, report less frequently, and integrate the task with the existing IDSР obligations. TB stakeholders at the national and international levels would need to be willing to accept less granular information from the private sector in exchange for higher adherence, fidelity, and completeness. Surveillance systems based on semitrusted partners that protect privileged and proprietary information are possible [15].

Limitations
Given the sensitivity of the topics in the study, the methods have certain caveats and design choices that need to be taken into consideration when interpreting the findings. Scales to detect social desirability were not included. A bimodal distribution with ceiling effects was observed; use of a structured survey did not allow us to probe all underlying rationales for noncompliance and the model did not explain all variances. Survey refusals were more common among facilities that treated TB but did not notify it (22% vs 4%), suggesting that the findings may underrepresent the full diversity of rationales for noncompliance. Sampling quotients overestimated TB treatment provision in the unengaged private sector, meaning that the majority (88%) of those interviewed were not faced with decisions about whether or not to notify TB cases and thus their responses may reflect historical or hypothetical choices. However, a strength of the study design is that it measured both providers’ self-reported reporting behavior and their actual reporting behavior so that socially desirable response bias is revealed in the juxtaposition as a finding.

This exploratory study endeavored to adapt and test a scale of public health authority trust as a possible contributor to understanding infectious disease surveillance behavior. While the trust scale proved robust, trust alone was insufficient to explain the variability in notification behavior. Mistrust in public authorities proved difficult to disentangle from mistrust of surveys. Additional methodological innovation may be required to overcome the influences of social desirability and acquiescence. The greater mistrust reported by those with more seniority may be confounded by the greater candor afforded by stature.

Conclusion
New forms of public-private collaboration in surveillance are needed that align with the varied interests of private providers, reflecting their varied caseloads and capacity for recording and reporting. While spurring desired provider behavior may seem a matter of assembling an enticing package of carrots and sticks, achieving this is demonstrably difficult without addressing the underlying governance and trust considerations that blunt private providers’ willingness. Moreover, our study suggests that it may be more effective to adapt the TB notification system to make it more responsive to end-user needs than to modify providers’ attitudes and behavior. The lessons learned are relevant for the design of other surveillance systems, including postmarketing pharmacovigilance of new health technologies, patient safety reporting, antibiotic stewardship, and early warning systems for emerging pathogens.

Acknowledgments
We wish to acknowledge the contributions of Nkem Chukwueme (New York University), Samson Bamidele Olorunju (Department of Epidemiology and Medical Statistics, Faculty of Public Health, University of Ibadan), Adebukola Agnes Adegbola (Koninklijke Centrale Vereniging tot bestrijding der Tuberculose [KNCV] tuberculosis [TB] Foundation Lagos; Challenge TB), Kingsley Umahoin (KNCV TB Foundation Lagos; Challenge TB), Christina Mengenthaler (Royal Tropical Institute, Amsterdam, the Netherlands), Victor Adepoju (KNCV TB Foundation Lagos; Directorate-General for International Cooperation of the Netherlands Project), Olanike Idris (Lagos State TB and Leprosy Control Program, Lagos State Ministry of Health), Rupert Enugu (United States Agency for International Development [USAID]), and Silke Heumann (Institute for Social Studies, The Hague, the Netherlands) for their contributions to the study and suggestions on the manuscript. This study was made possible by the generous support of taxpayers in the United States via the USAID. The Global Health Bureau, Office of Health, Infectious Disease and Nutrition, USAID financially supported this study through Challenge TB under the terms of agreement AID-OAA-A-14-00029. The funder played no role in the design, analysis, or reporting of this study.
Data Availability
Deidentified data are available for research purposes upon request to the corresponding author.

Authors' Contributions
EMHM, OAA, and CO conceived of the study methodology. OAA and CO curated the data and administered the study. EMHM and OAA conducted the formal analysis. MG acquired the funding. EMHM created data visualizations and wrote the original draft. Manuscript writing, review, and editing were conducted by all authors.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Survey Instrument.

Multimedia Appendix 2
Distribution of CTGO scores according to monthly IDSR (Figure S1); scatter plot of trust according to years of experience in the health care field by gender (Figure S2); distribution of public authority trust scores according to gender and self-notification behavior (Figure S3); factor structure of the adapted trust in public health authorities scale (Figure S4); hypothesis test summary (Table S1). CTGO: Citizen Trust in Government Organizations; IDSR: Integrated Disease Surveillance Response.

References


46. Rupani MP, Shah CJ, Dave JD, Trivedi AV, Mehta KG. 'We are not aware of notification of tuberculosis': a mixed-methods study among private practitioners from western India. Int J Tuberc Lung Dis 2015 Mar 03;19(3):298-304. [Medline: 25889952]


Abbreviations

CTGO: Citizen Trust in Government Organizations
DSNO: Disease Surveillance and Notification Office
IDSR: Integrated Disease Surveillance and Response
PPM: public-private mix
STBLCP: State Tuberculosis and Leprosy Control Program
TB: tuberculosis
WHO: World Health Organization
Giving a Voice to Patients With Smell Disorders Associated With COVID-19: Cross-Sectional Longitudinal Analysis Using Natural Language Processing of Self-Reports

Nick S Menger¹, MSc; Arnaud Tognetti²,³, PhD; Michael C Farruggia⁴, MSc; Carla Mucignat⁵, PhD; Surabhi Bhutani⁶, PhD; Keiland W Cooper⁷, MSc; Paloma Rohlf Domínguez⁸, PhD; Thomas Heinbockel⁹, PhD; Vonnie D C Shields¹⁰, PhD; Anna D'Errico¹¹, PhD; Veronica Pereda-Loth¹², PhD; Denis Pierron¹², PhD; Sachiko Koyama¹³, PhD; Ilja Croijmans¹⁴,¹⁵, PhD

¹Institute of Medical Psychology and Behavioural Neurobiology, University of Tübingen, Tübingen, Germany
²Department of Clinical Neuroscience, Division of Psychology, Karolinska Institutet, Stockholm, Sweden
³Centre d'Economie de l'Environnement Montpellier, Centre National de la Recherche Scientifique, Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement, Institut Agro, Université de Montpellier, Montpellier, France
⁴Interdepartmental Neuroscience Program, Yale University, New Haven, CT, United States
⁵Department of Molecular Medicine, University of Padova, Padua, Italy
⁶School of Exercise and Nutritional Sciences, San Diego State University, San Diego, CA, United States
⁷Department of Neurobiology and Behavior, University of California, Irvine, CA, United States
⁸Department of Developmental and Educational Psychology, University of Basque Country, Leioa, Spain
⁹Department of Anatomy, Howard University College of Medicine, Washington, DC, United States
¹⁰Biological Sciences Department, Fisher College of Science and Mathematics, Towson University, Towson, MD, United States
¹¹Goethe University of Frankfurt, Frankfurt am Main, Germany
¹²Laboratoire Évolution et Santé Orale, Université Toulouse III, Toulouse, France
¹³Department of Medicine, School of Medicine, Indiana University, Indianapolis, IN, United States
¹⁴Language and Communication Department, Faculty of Arts, Radboud University, Nijmegen, Netherlands
¹⁵Centre for Language Studies, Radboud University, Nijmegen, Netherlands

Corresponding Author:
Carla Mucignat, PhD
Department of Molecular Medicine
University of Padova
Via Francesco Marzolo
Padua, 35131
Italy
Phone: 39 0498275304
Email: carla.mucignat@unipd.it

Abstract

Background: Smell disorders are commonly reported with COVID-19 infection. The smell-related issues associated with COVID-19 may be prolonged, even after the respiratory symptoms are resolved. These smell dysfunctions can range from anosmia (complete loss of smell) or hyposmia (reduced sense of smell) to parosmia (smells perceived differently) or phantosmia (smells perceived without an odor source being present). Similar to the difficulty that people experience when talking about their smell experiences, patients find it difficult to express or label the symptoms they experience, thereby complicating diagnosis. The complexity of these symptoms can be an additional burden for patients and health care providers and thus needs further investigation.

Objective: This study aims to explore the smell disorder concerns of patients and to provide an overview for each specific smell disorder by using the longitudinal survey conducted in 2020 by the Global Consortium for Chemosensory Research, an international research group that has been created ad hoc for studying chemosensory dysfunctions. We aimed to extend the existing knowledge on smell disorders related to COVID-19 by analyzing a large data set of self-reported descriptive comments by using methods from natural language processing.
Methods: We included self-reported data on the description of changes in smell provided by 1560 participants at 2 timepoints (second survey completed between 23 and 291 days). Text data from participants who still had smell disorders at the second timepoint (long-haulers) were compared with the text data of those who did not (non–long-haulers). Specifically, 3 aims were pursued in this study. The first aim was to classify smell disorders based on the participants’ self-reports. The second aim was to classify the sentiment of each self-report by using a machine learning approach, and the third aim was to find particular food and nonfood keywords that were more salient among long-haulers than those among non–long-haulers.

Results: We found that parosmia (odds ratio [OR] 1.78, 95% CI 1.35-2.37; \( P < .001 \)) as well as hyposmia (OR 1.74, 95% CI 1.34-2.26; \( P < .001 \)) were more frequently reported in long-haulers than in non–long-haulers. Furthermore, a significant relationship was found between long-hauler status and sentiment of self-report (\( P < .001 \)). Finally, we found specific keywords that were more typical for long-haulers than those for non–long-haulers, for example, fire, gas, wine, and vinegar.

Conclusions: Our work shows consistent findings with those of previous studies, which indicate that self-reports, which can easily be extracted online, may offer valuable information to health care and understanding of smell disorders. At the same time, our study on self-reports provides new insights for future studies investigating smell disorders.

(JMIR Public Health Surveill 2024;10:e47064) doi:10.2196/47064

KEYWORDS
parosmia; anosmia; phantosmia; hyposmia; hyperosmia; long-hauler; sentiment classification; web-based study; COVID-19; smell disorders

Introduction

Chemosensory dysfunctions are among the distinguishing symptoms of COVID-19 [1-3]. Although many infected patients recover within weeks, a large percentage of patients experience long-term olfactory dysfunction even after recovery from the acute phase [4,5]. These individuals are known as smell long-haulers. Since smell impairment is often hard to notice and even harder to describe by patients (eg, patients commonly confuse smell disorders with taste disorders), information that can clarify the process of smell impairment or recovery can be useful to describe, understand, and track the phenomenon [6]. It is also unclear how distorted chemosensory perception relates to well-being or changes in behavior such as those related to food intake or avoidance.

Smell disturbances, categorized as quantitative (alterations in odor intensity) or qualitative (changes in perceived odor quality) [7,8], are common among COVID-19 long-haulers [1,9]. Symptoms can range from increased perceived intensity (hyperosmia) to faint perception of odors (hyposmia) or even complete loss of olfactory perception (anosmia). Qualitative changes may occur when the quality of the perceived odor is altered [7,8], which occurs, for instance, in case of parosmia (ie, distorted smell perception typically leading to perceived smells as bad odors, eg, the smell of food in good condition is perceived as garbage) or phantosmia (ie, perceiving smells without an olfactory stimulus, eg, smoke when there is no fire). Although these distortions are prevalent, individuals often lack technical terminology awareness, leading to challenges in accurate self-reporting and potential impacts on emotional well-being. Confusion may, for example, arise from terms such as “flavor” and “taste” [10], where individuals may report that they do not taste anything while only their smell is affected and not their taste. Similarly, different types of smell changes may cause confusion, as the quantity (complete or partial loss of smell) and quality of the changes indicate different impairments. Moreover, people are often inaccurate in reporting their olfactory performances, which may affect their emotional well-being and their awareness of olfactory dysfunction [11]. Properly classifying symptoms is crucial for clinical and research purposes, emphasizing the need for open-ended survey formats to capture nuanced experiences [12]. The difficulties in symptom description make it hard to capture a person’s particular olfactory experience in close-ended survey questions [9,13,14]. In our study, we employed open-ended surveys to collect comments [2], providing valuable insights like “smell is back to normal,” “my sense of smell fluctuates and is not as good as before,” or “the smell of coffee and onions has changed.”

The strong connection between smell and emotional-cognitive states involves neural pathways connecting the amygdala, thalamus, orbitofrontal cortex, and hippocampus [15-18]. Olfactory dysfunctions, often preceding neurodegenerative disorders, are evident in conditions such as Alzheimer disease, Parkinson disease, autism, and depression [16,19,20]. The detrimental effect of chemosensory dysfunction on emotional well-being is well-recognized [21-23] but is not fully understood with respect to olfaction in long-haulers. Patients have reported altered mental status as well as frustrations with COVID-19–related olfactory dysfunction [24]. Furthermore, olfactory dysfunction predicts the development of depression in adults [20]. Therefore, a negatively affected well-being or emotional tone when describing their symptoms would be expected in patients with COVID-19 and olfactory dysfunctions such as anosmia or parosmia, particularly in long-haulers.

In this study, we analyzed open-ended questions that were included in surveys conducted by the Global Consortium for Chemosensory Research (GCCR) [2,3]. As an initial validation of the informative value of the comments, a comparison was made between symptoms coded from open-ended comments and from the multiple-choice answers alone, administered during COVID-19 (survey 1) and after recovery (survey 2). This sets the stage to address the following 3 questions: (1) What are the frequencies of parosmia, phantosmia, and other olfactory dysfunctions (ie, hyposmia and anosmia) as reported in open-ended comments? (2) What is the well-being or emotional tone of people experiencing these symptoms as reported in...
open-ended comments? (3) What specific food-related experiences are related to these symptoms? Open-ended questions allow participants to voice their concerns that may not be covered by the other type of questions and are closer to how patients may report these symptoms to their general practitioner or health care worker [25]. The questions addressed smell loss while participants experienced COVID-19 (survey 1) and during a follow-up survey (survey 2). Analyzing these comments and their content contributes to better understanding, in a more ecologically valid way, of how long hauling might affect emotional well-being as it relates to olfactory experiences and the frequency and severity of symptoms compared with analyzing close-ended survey questions alone. On the basis of previously reported information, the following hypotheses have been formulated in this study.

1. Hypothesis aim 1: Recovery from smell loss is often accompanied by parosmia and phantosmia and is considered a sign of olfactory mucosa regeneration [1]. Considering that some smell-related symptoms may remain in COVID-19 long-haulers, we predict that long-haulers will have a greater occurrence of parosmia and phantosmia in addition to other potential chemosensory dysfunctions compared with non–long-haulers based on their own description of their olfactory symptom progression in survey 2.

2. Hypothesis aim 2: Using a machine learning aspect–based sentiment analysis, we predict that long-haulers will report significantly more emotional and psychological distress compared to non–long-haulers.

3. Hypothesis aim 3: We hypothesize that long-haulers reporting parosmia and phantosmia will exhibit avoidance behavior, resulting in omission of certain food and nonfood items. This will be apparent from a qualitative semantic analysis of the comments in survey 2.

**Methods**

**Study Design**
To investigate the study hypotheses, we used data previously available to the GCCR [2,3] that analyzed closed-ended responses. More specifically, we used data acquired by means of open-ended questions included in those surveys, not analyzed before in those studies. The data collection has been designed and planned following the structure of a prospective cohort study with cross-sectional characteristics, in which participants experiencing smell disorders completed a survey at timepoint 1 (survey 1) and were invited to complete a follow-up survey again at timepoint 2 (survey 2). Participants who had recovered at timepoint 2 were classified as non–long-haulers, whereas participants still experiencing smell disorders at timepoint 2 were classified as long-haulers. The differences between these 2 groups were analyzed and described.

**Study Participants**
The initial survey (survey 1) was completed between April and September 2020 by 12,313 participants [2,3], of which 3386 participants also completed the follow-up survey (survey 2) at timepoint 2 (between September 2020 and February 2021). Participants self-selected to participate in survey 1. They were invited via email to participate in survey 2 if they previously agreed to be recontacted; provided an email address; completed survey 1 in English, Spanish, Italian, Dutch, or French; and reported a change in smell, taste, and flavor (via a symptom checkbox) in survey 1. Participants completed the second survey between 23 and 291 (median 200) days after the first survey. From this, data from 1560 participants were included in this study. The participants were classified as either non–long-haulers (n=673) or long-haulers (n=887) based on their self-reported smell ability at survey 2 relative to survey 1 (Table 1). We refer the reader to [1] for a detailed overview of the data collection. Dutch, French, Italian, and Spanish comments were translated to English for our analyses. Translations were conducted by native speakers.

**Table 1.** Sample characteristics of the second Global Consortium for Chemosensory Research web-based survey on COVID-19 that was administered globally between September 2020 and February 2021.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Long-haulers (n=887)</th>
<th>Non–long-haulers (n=673)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>43.8 (12.3)</td>
<td>43.9 (12.2)</td>
</tr>
<tr>
<td>Gender (female), n (%)</td>
<td>701 (79)</td>
<td>464 (68.9)</td>
</tr>
<tr>
<td>Language, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>348 (39.2)</td>
<td>231 (34.3)</td>
</tr>
<tr>
<td>Spanish</td>
<td>114 (12.9)</td>
<td>76 (11.3)</td>
</tr>
<tr>
<td>Dutch</td>
<td>92 (10.4)</td>
<td>51 (7.6)</td>
</tr>
<tr>
<td>Italian</td>
<td>70 (7.9)</td>
<td>34 (5.1)</td>
</tr>
<tr>
<td>French</td>
<td>263 (29.7)</td>
<td>281 (41.8)</td>
</tr>
</tbody>
</table>

**Ethics Approval**
This preregistered cross-sectional web-based study was approved by the Office of Research Protections of the Pennsylvania State University in the United States (STUDY00014904). This study was approved as an exempt study. The protocol complies with the revised Declaration of Helsinki and is compliant with privacy laws in the United States and European Union. Data reported here were collected from the GCCR core questionnaire. All participants provided an
online consent to participate in the survey prior to proceeding with answering the survey questions. Only those participants who previously agreed to be recontacted and provided their emails were approached to complete the second survey. The study data were deidentified at the Pennsylvania State University before these were provided to other researchers for analysis. All participants volunteered to participate in this study, and no compensation was provided.

Procedure

**Aim 1: Occurrence of Smell Disorders in Long-Haulers**

Our first aim was to determine the incidence of anosmia, hyposmia, phantosmia, and parosmia in long-haulers versus that in non–long-haulers by analyzing the free-text comments. At variance with [1], these complaints were categorized and counted based on descriptive comments and not on the participants’ self-reports by means of closed questions. The self-report question in [1] asking for changes in smell was meant to capture quantitative changes and was not always sensitive to capture individual experiences. The question that prompted the free-text comment was “Please describe any current changes in smell. Type ‘none’ if this is not applicable.”

The data were first processed by means of a concept-driven quantitative content analysis, as described by [26]. Comments recorded at both timepoints (n=2543) were coded manually by 8 different coders for the presence of anosmia, hyposmia, parosmia, phantosmia; whether the person indicated that they had recovered, not at all, partially or fully; and for mentioning food and nonfood odorous items according to a predetermined coding scheme (see coding scheme in Multimedia Appendix 1). Cases that did not meet any of the symptoms were excluded from further analyses. The responses that contained “none” entries (n=469) were also excluded from further analyses.

As an additional validation measure, the overlap between coded symptom prevalence of the free-text comments was compared to the self-reported prevalence on the multiple-choice question that was used in [1] (Multimedia Appendices 2-4). In these multiple-choice questions, participants were asked to rate whether they had noticed anosmia, hyposmia, parosmia, or phantosmia by asking the following questions, respectively: “I cannot smell at all/smells smell less strong than they did before my impairment” (anosmia/hyposmia), “smells smell different than they did before my impairment (the quality of smell has changed)” (parosmia), and “I can smell things that aren’t there (eg, I smell burning when nothing is on fire)” (phantosmia). Since multiple-choice questions for anosmia and hyposmia were not provided separately, we could only validate parosmia, phantosmia, as well as anosmia and hyposmia together.

**Aim 2: Emotional Distress in Long-Haulers**

For this aim, we trained a sentiment classification algorithm with LCF-ATEPC (local context focus–aspect term extraction and polarity classification) from PyABSA [27] (version 1.16.15; Heng Yang) on a data set of restaurant reviews available in PyABSA (Restaurant16) and then used the trained algorithm to first extract the so-called aspects, that is, keywords in each comment. Each aspect in each comment was classified as being negative, neutral, or positive (numerically coded as 1, 2, and 3, respectively). Comments such as “mustard smells really pungent, and I can’t be around it” or “I smell things a little less well” were classified as negative, while comments such as “I am back to normal in smelling” or “smell has come back completely after recovering” were classified as positive. To validate the results of the first model, we trained a second model with the same parameters on a data set of laptop reviews also available in PyABSA (Laptop14) and used that model to classify the comments and conduct the same analyses.

**Aim 3: Salience of Food and Nonfood Items in Long-Haulers**

Once the incidence and impact of smell disorders were established, we determined whether specific foods, drinks, or other objects were associated with long hauling, with the goal to examine whether specific items were more salient in long-haulers than in non–long-haulers. We first extracted all food and nonfood items from the comments that had been coded for aim 1. Qualitative relationships between items were visualized using word clouds. To quantify the use of specific words across groups, we then conducted a relative frequency analysis on the extracted food and nonfood words from comments.

Statistical Analyses

**Aim 1: Occurrence of Smell Disorders in Long-Haulers**

We used logistic regressions to approach aim 1 (glm function with a binomial error structure of the stats package in R) and assessed whether the 2 categories of participants (long-haulers vs non–long-haulers) differed in terms of the reported disorders that they respectively experienced. The dependent variables were each of the 4 smell disorders studied, namely, parosmia, phantosmia, hyposmia, or anosmia (0 for absence, 1 otherwise). Our explanatory variable was “smell-long-hauler status” (long-hauler vs non–long-haulers). We included age and gender of the participants and whether the comments were translated in English (0 for untranslated comments vs 1 for translated comments) as control variables as well as their interactions with the explanatory variable. We included all the main effects and interaction terms in the initial model, which were then simplified by removing the nonsignificant interaction terms to achieve the minimal adequate model. We centered “age” in all the models in order to make the effects more easily interpretable. Statistical analyses were performed using R software (version 4.1.3; R Foundation for Statistical Computing).

**Aim 2: Emotional Distress in Long-Haulers**

A chi-square test was conducted to examine the relationship between sentiment classifications and long-hauler status of the aspect-based sentiment classification algorithm. A mixed-effects logistic regression model was used to examine the relationship among sentiment, long-hauler status, and olfactory dysfunction type. All analyses were conducted with Python (version 3.8.8; Python Software Foundation) and the packages Scipy (version 1.9.3; Community library project) and Statsmodels (version 0.13.5; Jonathan Taylor).
Aim 3: Salience of Food and Nonfood Items in Long-Haulers

This aim was approached initially by creating word clouds and subsequently conducting a relative frequency analysis. For the word clouds, the extracted words describing food and nonfood objects were converted to lower case unigrams, bigrams, and trigrams by using the R package RWeka (version 0.4.44; Kurt Hornik) and TM (version 0.7.8; Ingo Feinerer). The R package wordcloud2 (version 0.2.1; Dawei Lang) and RColorBrewer (version 1.1.3; Erich Neuwirth) were used to create word clouds, where the frequency of an n-gram determines the size within the cloud. For the relative frequency analysis, we preprocessed and aggregated each group’s comments into a corpus based on the long-hauler status. Preprocessing steps included splitting the plain text comments into tokens. Tokens were lowercased, and numerals and punctuations were removed. Commonly used stop words were removed, and the text was lemmatized using Wordnet [28]. We then computed the frequency lists for each corpus in the comparison based on our preprocessed comments. The log-likelihood statistic was calculated for each word in the 2 frequency lists by constructing a contingency table based on word frequencies within and across corpora as per the method in [29]. Given that log-likelihood is a statistical significance measure, it does not compute the size of the difference between corpora; rather, it provides the words we have the most evidence for. Thus, to determine the influence of each word in each of the corpora, the relative frequency [30] method was used. By comparing the normalized frequencies for each word, this method returns a value (−1, 1). In our case, 1 indicates that the word is overused in the long-hauler corpus, and −1 in the non–long-hauler corpus. Using these metrics, we determined the food and nonfood words in the corpora by manually coding words from each category with log-likelihood values greater than 3.84 (a significance threshold of $P<.05$ or lower) and selected for further analysis.

Results

Aim 1: Occurrence of Smell Disorders in Long-Haulers

The logistic regression examining the association between anosmia and long-hauler status could not be performed, as the number of participants reporting anosmia was too small in both long-haulers (17/750) and non–long–haulers (0/338; see Table 2; Multimedia Appendix 5). For the other smell disorders, the minimal adequate models that were run to obtain the results reported below were obtained by removing the nonsignificant interaction terms between smell long-hauler status and the 3 control variables (age, gender, and translation) in all models (.06<$P<.98). The logistic regressions revealed a significant effect of the smell long-hauler status (long-hauler vs non–long-haulers) in terms of reported disorders (Multimedia Appendix 6). Long-hauler participants were significantly more likely to report symptoms interpreted as parosmia ($β=.58$, SE .14; $z=4.06$; odds ratio [OR] 1.78, 95% CI 1.35-2.37; $P<.001$; Multimedia Appendix 7) and hyposmia ($β=.55$, SE .13; $z=4.16$; OR 1.74, 95% CI 1.34-2.26; $P<.001$; Multimedia Appendix 8) compared to non–long-haulers. However, long-hauler status did not affect the likelihood to report symptoms associated with phantosmia ($β=.09$, SE .19; $z=0.49$; OR 1.10, 95% CI 0.76-1.62; $P=.63$; Multimedia Appendix 9). To obtain in-depth information on the relationship between smell disorders and the control variables (eg, sex, age), please consult Multimedia Appendices 6-9.

Table 2. Sample size of reported olfactory dysfunctions by smell long-hauling status and gender as coded from self-reports in the Global Consortium for Chemosensory Research survey on COVID-19 that was administered globally between September 2020 and February 2021.

<table>
<thead>
<tr>
<th>Dysfunction</th>
<th>Long-hauler (female), n/N (%)</th>
<th>Non–long-hauler (female), n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parosmia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>331/430 (77)</td>
<td>165/238 (69.3)</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>68/122 (55.7)</td>
<td>266/298 (89.3)</td>
</tr>
<tr>
<td>Phantosmia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>497/637 (78)</td>
<td>196/294 (66.7)</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>100/113 (88.5)</td>
<td>37/44 (84.1)</td>
</tr>
<tr>
<td>Hyposmia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>233/286 (81.5)</td>
<td>114/176 (64.8)</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>364/464 (78.4)</td>
<td>119/162 (73.5)</td>
</tr>
<tr>
<td>Anosmia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>584/733 (79.7)</td>
<td>233/338 (68.9)</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>13/17 (76.5)</td>
<td>0</td>
</tr>
</tbody>
</table>

Aim 2: Emotional Distress in Long-Haulers

The classifier that was trained on restaurant reviews was unable to classify the sentiment of 854 out of 1560 comments, and 838 out of 1560 could not be classified by the model trained on laptop reviews. These comments were removed from further analyses. There was a relationship between the classified sentiment and long-hauler status based on the model that was...
trained on restaurant reviews ($\chi^2 = 15.1; P < .001$; Table 3) and laptop reviews ($\chi^2 = 30.3; P < .001$).

In addition to comparing the sentiments of long-haulers and non–long-haulers, an additional analysis examined the effect of specific olfactory dysfunction on the comment’s sentiment. We therefore split the data set based on each smell disorder (Table 4; Multimedia Appendix 10) and compared how the sentiment changed when a smell disorder was reported versus when it was not reported. It is important to note that the split of the data was not exclusive, that is, participants who, for example, did not report parosmia, could still have reported hyposmia, phantosmia, or anosmia.

Within the comments of long-haulers, there was an effect of parosmia ($\beta = -1.13$, SE .35; OR 0.32, 95% CI 0.16-0.62; $P = .001$) and hyposmia ($\beta = -.76$, SE .33; OR 0.47, 95% CI 0.24-0.90; $P = .02$). No significant effects were found for phantosmia ($\beta = -.53$, SE .50; OR 0.62, 95% CI 0.36-1.06; $P = .08$), hyposmia ($\beta = .16$, SE .20; OR 1.17, 95% CI 0.80-1.72; $P = .43$), or anosmia ($\beta = .27$, SE .62; OR 1.31, 95% CI 0.37-4.39; $P = .66$). For the non–long-hauler comments (Multimedia Appendix 13), a significant effect was found for parosmia ($\beta = -.71$, SE .30; OR 0.49, 95% CI 0.27-0.89; $P = .02$) and hyposmia ($\beta = -.55$, SE .28; OR 0.58, 95% CI 0.33-1.00; $P = .05$) but not phantosmia ($\beta = .01$, SE .51; OR 1.01, 95% CI 0.37-2.81; $P = .99$).

As a validation of these results, the same analyses were conducted for the model that was trained on laptop reviews (Multimedia Appendix 12). The sentiment classification of long-hauler comments showed no significant effects for parosmia ($\beta = -.22$, SE .19; OR 0.80, 95% CI 0.56-1.15; $P = .24$), phantosmia ($\beta = -.47$, SE .28; OR 0.62, 95% CI 0.36-1.06; $P = .08$), hyposmia ($\beta = 1.17$, SE .20; OR 1.17, 95% CI 0.80-1.72; $P = .43$), or anosmia ($\beta = .27$, SE .62; OR 1.31, 95% CI 0.37-4.39; $P = .66$). For the non–long-hauler comments (Multimedia Appendix 13), a significant effect was found for parosmia ($\beta = -.71$, SE .30; OR 0.49, 95% CI 0.27-0.89; $P = .02$) and hyposmia ($\beta = -.55$, SE .28; OR 0.58, 95% CI 0.33-1.00; $P = .05$) but not phantosmia ($\beta = .01$, SE .51; OR 1.01, 95% CI 0.37-2.81; $P = .99$).

Table 3. Sample size (n=706) of classified sentiments by smell long-hauling status for the model trained on restaurant reviews. Responses are smell-related self-reports from a web-based survey on COVID-19 that was administered globally between September 2020 and February 2021.

<table>
<thead>
<tr>
<th>Hauling status responses</th>
<th>Negative, n</th>
<th>Neutral, n</th>
<th>Positive, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responses from long-haulers</td>
<td>172</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>Responses from non–long-haulers</td>
<td>435</td>
<td>1</td>
<td>51</td>
</tr>
</tbody>
</table>

Table 4. Proportion of long-haulers’ comments and their sentiments, as classified by the model trained on restaurant reviews. The comments are smell-related self-reports from a web-based survey on COVID-19 that was administered globally between September 2020 and February 2021.

<table>
<thead>
<tr>
<th>Dysfunction</th>
<th>Negative sentiment</th>
<th>Neutral sentiment</th>
<th>Positive sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parosmia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>0.44</td>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>0.46</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>Phantosmia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>0.76</td>
<td>0</td>
<td>0.09</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>0.13</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Hyposmia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>0.33</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>0.56</td>
<td>0</td>
<td>0.06</td>
</tr>
<tr>
<td>Anosmia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>0.87</td>
<td>0</td>
<td>0.10</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 5. Proportion of non–long-haulers’ comments and their sentiments, as classified by the model trained on restaurant reviews. The comments are smell-related self-reports from a web-based survey on COVID-19 that was administered globally between September 2020 and February 2021.

<table>
<thead>
<tr>
<th>Dysfunction</th>
<th>Negative sentiment</th>
<th>Neutral sentiment</th>
<th>Positive sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parosmia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>0.45</td>
<td>0.01</td>
<td>0.17</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>0.34</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Phantosmia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>0.72</td>
<td>0.01</td>
<td>0.17</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>0.07</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Hyposmia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No dysfunction reported</td>
<td>0.36</td>
<td>0</td>
<td>0.15</td>
</tr>
<tr>
<td>Dysfunction reported</td>
<td>0.42</td>
<td>0</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Aim 3: Salience of Food and Nonfood items in Long-Haulers

Word clouds were generated to visually explore differences in food and nonfood items mentioned differently in the comments of long-haulers and non–long-haulers (Multimedia Appendices 14 and 15). Superficially, the word clouds appeared similar for both groups. Non–long-haulers appeared to mention cheese, urine, and sweat somewhat more often than long-haulers. Both groups most often mentioned coffee, onion, and food, and for the nonfood items, they mentioned perfume and smoke.

The relative frequency analysis, reported in Table 6, revealed that lemon was mentioned more often by long-haulers, whereas wine, cheese, vinegar, and mustard were mentioned more often by non–long-haulers. For the nonfood items, long-haulers more often mentioned weird (presumably for weird smells), fire, gas, and eucalyptus among other smelling objects. This is in line with the finding that long-haulers more often report parosmia and thus report more of these foul-smelling objects in their comments, whereas non–long-haulers might report the objects that they can smell and taste again in their comments (eg, wine, cheese).
Table 6. Results of a relative frequency analysis showing a list of words that were reported significantly ($P<.05$) more often by long-haulers or non–long-haulers. These words were extracted from smell-related self-reports in a web-based survey on COVID-19 that was administered globally between September 2020 and February 2021.

<table>
<thead>
<tr>
<th>Food items</th>
<th>Log-likelihood</th>
<th>Occurrence$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lemon</td>
<td>7.62</td>
<td>0.78</td>
</tr>
<tr>
<td>Vinegar</td>
<td>9.90</td>
<td>-0.50</td>
</tr>
<tr>
<td>Cheese</td>
<td>6.12</td>
<td>-0.52</td>
</tr>
<tr>
<td>Wine</td>
<td>12.60</td>
<td>-0.63</td>
</tr>
<tr>
<td>Mustard</td>
<td>4.22</td>
<td>-0.76</td>
</tr>
<tr>
<td>Red wine</td>
<td>9.46</td>
<td>-0.86</td>
</tr>
<tr>
<td>Nonfood items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weird</td>
<td>7.62</td>
<td>0.78</td>
</tr>
<tr>
<td>Fire</td>
<td>7.62</td>
<td>0.78</td>
</tr>
<tr>
<td>Gas</td>
<td>10.20</td>
<td>0.72</td>
</tr>
<tr>
<td>Eucalyptus</td>
<td>4.75</td>
<td>0.71</td>
</tr>
<tr>
<td>Detergent</td>
<td>6.14</td>
<td>0.64</td>
</tr>
<tr>
<td>Air</td>
<td>9.14</td>
<td>0.60</td>
</tr>
<tr>
<td>Scent</td>
<td>8.07</td>
<td>0.32</td>
</tr>
<tr>
<td>Burning</td>
<td>4.90</td>
<td>0.27</td>
</tr>
<tr>
<td>Smoke</td>
<td>4.05</td>
<td>0.24</td>
</tr>
<tr>
<td>Thing</td>
<td>9.40</td>
<td>0.19</td>
</tr>
<tr>
<td>Very</td>
<td>5.90</td>
<td>0.13</td>
</tr>
<tr>
<td>Smell</td>
<td>4.08</td>
<td>0.03</td>
</tr>
</tbody>
</table>

$^a$A larger occurrence value indicates that words were reported more frequently by long-haulers, and a lower (negative) value indicates a higher frequency for non–long-haulers.

Discussion

Principal Findings

First, our study shows that there is a strong relationship between long-hauler status and the incidence of both quantitative and qualitative smell changes, as apparent from our analysis of the open-ended comments. In addition, our analysis shows that there is a strong relationship between the smell disorder status as extracted from open-ended comments compared to those extracted from multiple-choice questions. Second, the sentiment analysis revealed that long-haulers are more negative in tone, underscoring the socioemotional impact of smell disorders on the individual. Third, we found that specific smell objects were mentioned in the free-text comments, differentiated by long-haulers and non–long-haulers, and specific smell dysfunction symptoms. We will elucidate these results more in the remainder of this discussion.

Our data show that, according to our first hypothesis, long-haulers reported a significantly higher occurrence of olfactory dysfunctions, in particular, parosmia ($P<.001$) and hyposmia ($P<.001$), than non–long-haulers did. This agrees with previous research that shows that olfactory dysfunctions, especially parosmia, appear to be part of long COVID [1]. This marks the need for accurately defining the sensory experience by introducing a clinical routine testing that may highlight the recovery pathway, even if partial, as well as to give feedback to patients on their clinical course. It also points toward a need for better sensory education among the population that often has no means of describing their perception accurately. Moreover, it is relevant that clinicians become aware of the relative inaccuracy of patients when reporting their chemosensory symptoms. Olfactory unawareness has been reported several times [31], in particular, in older and at-risk populations. However, even if reported to health care professionals, olfactory dysfunctions are seldom objectively tested [32]. Reduced or simplified tests may be of value [33,34]; yet, the clinical practice largely relies on self-reporting. In the case of patients with COVID-19, guided classification of odors revealed a perceptual landscape different from controls [35].

Our second aim was to link emotional distress to long-hauler status by using a machine learning model, thereby eliminating human bias. The sentiment classifier revealed a more negative tone in long-haulers’ comments, as has also been shown by other studies [36-38]. Clustering by symptoms indicated that parosmia or hyposmia was associated with more negative words for both long-haulers and non–long-haulers. This raises the question why this was not the case for phantosmia. The reason might be that parosmia and its stronger relationship with food...
intake results in, for example, loss of appetite [37,39] and therefore stronger emotional impact. However, ambiguity exists in the interpretation of these results as is also mentioned by [40], who argue that there is much room for improvement of machine learning models in the domain of health. The tendency of these results is interesting nonetheless, and follow-up analyses coupled to the analysis of food intake habits [41] that more specifically delves into the quality of life could shed a clearer light on the matter.

A semantic analysis of the objects mentioned within the comments suggested that the most frequently mentioned objects were similar for long-haulers and non–long-haulers concerning nonfood items (eg, perfume, smoke), while for food items, the 2 most frequent words were the same but switched between long-haulers and non–long-haulers. The word “onion” was used most frequently by non–long-haulers and “coffee” by long-haulers. Concerning differentially reported food words, “lemon” was mostly reported by long-haulers and “red wine” by non–long-haulers. When looking at the specific olfactory dysfunction, the largest problem seemed to arise in the nonfood words mentioned by participants with phantosmia. They seemed to often mention words such as “smoke” and “burning.” Objects reported by both groups may represent the most salient odor objects. Long-haulers may no longer smell these, while non–long-haulers may perceive them as the first smells they regain. However, aim 3 words were only analyzed on their occurrence and thus did not provide a means to examine whether their occurrence elicited positive or negative feelings. Exploring whether certain foods were avoided or approached during smell dysfunction and investigating compounds triggering parosmia could provide valuable indicators, such as pyrazines for coffee, disulfides for onion, thiols for garlic, and methoxypyrazines for wine [42].

The mentioning of words that describe a burning sensation in participants with phantosmia is consistent with those mentioned in previous findings [43]. It provides confirmation of these findings in a large sample and offers an overview of other words and sensations that may be associated with phantosmia, which may be helpful for clinicians. Although multiple hypotheses have been proposed on the causes of phantosmia, it may very well be that phantasmic sensations have varying causes [44]. Our data consist of a wide range of self-reports of these phantasmic sensations and are therefore suitable for a follow-up analysis on specific contexts that evoke a phantasmic sensation to further the understanding of phantasmic experiences and the mechanisms that underlie them. This also highlights the value of internet in medical research, which allows for a collection of large data sets on patients’ self-reports that may add to the understanding of olfactory dysfunctions.

The difficulty people face in discussing smells and olfactory experiences is a well-known phenomenon [11]. In this study, we analyzed open-ended comments to gain insight into the experiences of individuals experiencing olfactory dysfunctions that are not entirely understood or correctly named by patients experiencing these. First, we found a discrepancy between the information in the open-ended comments and the close-ended multiple-choice question that asks for olfactory dysfunction. This suggests that the comments contain valuable information that differs between groups and symptoms, highlighting the importance of this approach. This study emphasizes the importance of considering open-ended comments to gain a more holistic understanding of participants’ experiences and perceptions. The approach presented here, that is, manually coding short open-ended responses for different symptoms, may be used in combination with machine learning classification paradigms, to better understand patients’ concerns voiced in web-based settings such as in web-based survey research, online patient council groups, or by using brief digital notes from general practitioners as input. This additional source of information could lead to better identification of different diagnoses, and along the line, better understanding of the different types of smell disorders.

Finally, few people mentioned that months after developing COVID-19–induced smell dysfunction, their smell did improve, even to a level of functioning that was better than before the onset of COVID-19. We identified 13 cases of hyperosmia. It is currently unknown in what percentage this phenomenon occurs and what the mechanisms underlying this phenomenon are, and hyperosmia has been seldom reported after COVID-19 [45,46]. Speculatively, an increase in the awareness of smells after having lost one’s sense of smell for a brief period could potentially drive attentional experience with odors—in an “you don’t know what you’ve got until you lose it” way [47]. In the future, large-scale prevalence studies on smell disorders could investigate the percentage of people who report an improved sense of smell after recovering from a smell dysfunction.

We considered the validity of our analyses of open-ended responses, as opposed to the analyses of close-ended questions in [1]. Concurrent with their results, we found that long-haulers were more likely to experience parosmia and hyposmia, but in contrast to their findings for phantosmia, we did not find that long-haulers were more likely to experience phantosmia. This may result from the lack of overlap between our coding and participants’ self-reports. Generally, coders underestimated the prevalence of smell disorders. This effect was the highest for phantosmia, and may, therefore, have led to the discrepancy between [1] and our study. Possible explanations for the high underestimation of prevalence by coders are that participants did not report their whole experience in the comments or that participants overestimated their experiences in the multiple-choice questions. This is a downside to our approach, but overall, the overlap between the coded and self-reported smell disorders seemed to be in a good range and should therefore allow for a useful interpretation of the results that our natural language processing measures yielded.

Limitations in Our Study

Our study has several limitations. There was a selection bias in the participant group. As in [1], we are aware that there might be a self-selection bias in completing these surveys, as participants experiencing severe symptoms may be more motivated to also complete the second survey. At the same time, participants who felt that they had severe smell disorders may select additional answer options on the multiple-choice questions and clarify their symptoms in the open-comment field. Thus, this selection bias, as compared with [1], could go either way.

https://publichealth.jmir.org/2024/1/e47064
This provides an opportunity to compare both ways of asking about symptoms from participants, which we included in aim 1. Notwithstanding, we interpret the results with caution and phrase future suggestions and implications cautiously.

**Conclusion**

The use of web-based surveys proved of value during COVID-19 since it allowed the rapid collection of data for monitoring infection-related chemosensory deficits. In addition to multiple-choice answers, a thorough description of symptoms could be extracted from open-ended answers. The surveys in this study made it possible to describe the qualitative changes in the chemosensory functions and explore their frequency. The distress of long-haulers could be investigated using a nonhuman unbiased algorithm for sentiment classification. Lastly, it was possible to highlight the different use of words related to food and nonfood items, possibly relating to the different perceptual experiences. In conclusion, we proved the validity of our approach, based on the analysis of open-ended questions to better understand the perceptual world of patients, described using their own words. Our analysis provides a new perspective on olfactory disturbances following COVID-19 that cannot be captured through closed-ended questions. On a more general level, this data science perspective can advance web-based survey-based patient research studies.

**Data Availability**

The data sets and scripts generated and analyzed during this study are available in the Open Science Framework repository “Giving a voice to adults with COVID-19: An analysis of the open-ended comments from COVID-19 smell long-haulers and non-long-haulers on the Global Consortium for Chemosensory Research survey” [48].

**Authors' Contributions**

IC, SB, KWC, PRD, MCF, TH, SK, NSM, CM, VDCS, and AD conceptualized this study. IC, SB, MCF, NSM, AT, and VP-L curated the data for this study. IC, KWC, PRD, TH, NSM, AT, and VDCS performed the formal analysis in this study. CM acquired the funds. IC, SB, KWC, NSM, and AT devised the methodology. IC, SB, and SK administered the project. IC, SB, KWC, PRD, MCF, TH, NSM, CM, DP, AT, VDCS, AD, and VP-L provided the resources. IC, KWC, NSM, and AT designed the software. IC, SB, NSM, and AT supervised this study. IC, KWC, MCF, NSM, and AT visualized this study. IC, PRD, NSM, and AT contributed to the writing of the original draft. IC, SB, KWC, PRD, MCF, TH, SK, NSM, CM, DP, AT, VDCS, and AD reviewed and edited this paper.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Guidelines that were used in aim 1 to code the comments regarding olfactory dysfunction and a description of the practical implementation.

[PDF File (Adobe PDF File), 78 KB - publichealth_v10i1e47064_app1.pdf ]

Multimedia Appendix 2
Confusion matrix showing the agreement between the coders (open-ended comments) and the participants’ self-report (multiple-choice question) for parosmia (survey 1 and survey 2 combined).

[PDF File (Adobe PDF File), 151 KB - publichealth_v10i1e47064_app2.pdf ]

Multimedia Appendix 3
Confusion matrix showing the agreement between the coders (open-ended comments) and the participants’ self-report (multiple-choice question) for phantosmia (survey 1 and survey 2 combined).

[PDF File (Adobe PDF File), 153 KB - publichealth_v10i1e47064_app3.pdf ]

Multimedia Appendix 4
Confusion matrix showing the agreement between the coders (open-ended comments) and the participants’ self-report (multiple-choice question) for smell loss, that is, anosmia and hyposmia (survey 1 and survey 2 combined).

[PDF File (Adobe PDF File), 155 KB - publichealth_v10i1e47064_app4.pdf ]

Multimedia Appendix 5

[PDF File (Adobe PDF File), 541 KB - publichealth_v10i1e47064_app5.pdf ]
Multimedia Appendix 6
Odds ratios from the logistic regressions examining the association between smell long-hauling and prevalence of olfactory disorders.

[PDF File (Adobe PDF File), 543 KB - publichealth_v10i1e47064_app6.pdf ]

Multimedia Appendix 7
Logistic regression investigating whether smell long- vs. non-long haulers differed in terms of reported parosmia they respectively experienced.

[PDF File (Adobe PDF File), 512 KB - publichealth_v10i1e47064_app7.pdf ]

Multimedia Appendix 8
Logistic regression investigating whether smell long- vs. non-long haulers differed in terms of reported hyposmia they respectively experienced.

[PDF File (Adobe PDF File), 511 KB - publichealth_v10i1e47064_app8.pdf ]

Multimedia Appendix 9
Logistic regression investigating whether smell long- vs. non-long haulers differed in terms of reported phantosmia they respectively experienced.

[PDF File (Adobe PDF File), 511 KB - publichealth_v10i1e47064_app9.pdf ]

Multimedia Appendix 10
Proportion of long-haulers’ comments and their sentiments, as classified by the model trained on restaurant reviews. The comments are smell-related self-reports from a web-based survey on COVID-19 that was administered globally between September 2020 and February 2021.

[PDF File (Adobe PDF File), 516 KB - publichealth_v10i1e47064_app10.pdf ]

Multimedia Appendix 11
Proportion of non-long-haulers’ comments and their sentiments, as classified by the model trained on restaurant reviews. The comments are smell-related self-reports from a web-based survey on COVID-19 that was administered globally between September 2020 and February 2021.

[PDF File (Adobe PDF File), 517 KB - publichealth_v10i1e47064_app11.pdf ]

Multimedia Appendix 12
Sentiment classification from the model that was trained on laptop reviews. Showing the proportion of comments from long-haulers that were classified as negative, neutral, and positive across all smell disorders.

[PDF File (Adobe PDF File), 488 KB - publichealth_v10i1e47064_app12.pdf ]

Multimedia Appendix 13
Sentiment classification from the model that was trained on laptop reviews. Showing the proportion of comments from non-long-haulers that were classified as negative, neutral, and positive across all smell disorders.

[PDF File (Adobe PDF File), 489 KB - publichealth_v10i1e47064_app13.pdf ]

Multimedia Appendix 14
Word clouds of words that were extracted from long-hauler and non-long-hauler comments, where the size of each word is determined by its frequency.

[PDF File (Adobe PDF File), 595 KB - publichealth_v10i1e47064_app14.pdf ]

Multimedia Appendix 15
Word clouds of words that were manually extracted from comments, grouped by olfactory dysfunction, where the size of the word represents its frequency.

[PDF File (Adobe PDF File), 575 KB - publichealth_v10i1e47064_app15.pdf ]

References


Abbreviations

GCCR: Global Consortium for Chemosensory Research
LCF-ATEPC: local context focus–aspect term extraction and polarity classification
OR: odds ratio
Modeling the Regional Distribution of International Travelers in Spain to Estimate Imported Cases of Dengue and Malaria: Statistical Inference and Validation Study

Abstract

Background: Understanding the patterns of disease importation through international travel is paramount for effective public health interventions and global disease surveillance. While global airline network data have been used to assist in outbreak prevention and effective preparedness, accurately estimating how these imported cases disseminate locally in receiving countries remains a challenge.

Objective: This study aimed to describe and understand the regional distribution of imported cases of dengue and malaria upon arrival in Spain via air travel.

Methods: We have proposed a method to describe the regional distribution of imported cases of dengue and malaria based on the computation of the “travelers’ index” from readily available socioeconomic data. We combined indicators representing the main drivers for international travel, including tourism, economy, and visits to friends and relatives, to measure the relative appeal of each region in the importing country for travelers. We validated the resulting estimates by comparing them with the reported cases of malaria and dengue in Spain from 2015 to 2019. We also assessed which motivation provided more accurate estimates for imported cases of both diseases.

Results: The estimates provided by the best fitted model showed high correlation with notified cases of malaria (0.94) and dengue (0.87), with economic motivation being the most relevant for imported cases of malaria and visits to friends and relatives being the most relevant for imported cases of dengue.

Conclusions: Factual descriptions of the local movement of international travelers may substantially enhance the design of cost-effective prevention policies and control strategies, and essentially contribute to decision-support systems. Our approach contributes in this direction by providing a reliable estimate of the number of imported cases of nonendemic diseases, which could be generalized to other applications. Realistic risk assessments will be obtained by combining this regional predictor with the observed local distribution of vectors.

(JMIR Public Health Surveill 2024;10:e51191) doi:10.2196/51191

KEYWORDS
epidemiology; imported infections; modeling; surveillance system; vector-borne diseases
Introduction

Throughout history, human mobility has been a key determinant for the spread of infectious diseases. From the 14th century bubonic plague pandemic to the 1918 Spanish flu as well as the more recent Ebola epidemic and COVID-19 pandemic, the way individuals travel across the globe has shaped the evolution and geographical dynamics of infectious diseases [1-3].

International mobility flows are especially relevant for the spread of vector-borne diseases (VBDs), which often receive less attention in routine epidemiological surveillance plans in countries where they are not endemic. In recent years, global warming and intensified urbanization processes have favored the establishment of previously foreign species around the globe, such as Aedes, Anopheles, and Culex mosquitoes [4-7]. These are vectors for malaria, dengue, yellow fever, West Nile virus, Zika, and chikungunya, and thus, they pose a significant public health risk that needs adequate preparedness [8-10]. Air travel plays a central role in the diffusion of most of these diseases, allowing their spread through imported cases at nonendemic locations [11-13] with vector presence, and it must be incorporated in decision-support systems to achieve operational preparedness and risk prediction [14-16]. Reliable descriptions and predictions of migration flows have been proven to be valuable tools for the design of more effective public health policies [17-20]. However, new developments are always needed as complex dynamics are expected to arise from the multi-step life cycle of VBDs [21-23].

A wide range of approaches have been used to model risks of imported cases of dengue [24-27] and malaria [28-30] from endemic to nonendemic regions. Several studies [31-33] have incorporated data on the global airline network to assist outbreak prevention and public health preparedness. However, accurate estimations of how such imported cases disseminate locally after arrival in the receiving countries are harder to devise. While fine-grained data on local mobility are available from cellular networks [34-36], it is not clear how the specific behavior of travelers can be differentiated from the local population dynamics. Moreover, travel-related data are usually obtained through coarse-grained spatial statistics, thus involving large territories. A reason for this is the inherent complexity and range of scales involved in human flows. While it is possible to record information at designated locations (e.g., airports), it is much more difficult to reliably collect detailed movement data of target groups over larger geographical regions. In these cases, insights on the nature of human traveling behaviors and motivations that uncover hidden patterns in these processes are crucial as they may be used to sidestep the need for excessively detailed and thus unreachable data [37-39].

We aimed to provide accurate descriptions of how infected travelers may distribute in a territory, which could be valuable input for local authorities in the design of cost-effective VBD prevention and control strategies. For this, we approximated the local distribution of travelers arriving at a specific country (or any other territory) in terms of readily available indicators, rather than considering travel information that is usually not quantified at a local scale. These statistics gauge the appeal of each region to foreign travelers, quantifying the number of imported cases each region may receive. We calibrated our model with the number of imported cases of dengue and malaria at each province in Spain from 2015 to 2018 and then performed validation by comparing our model’s estimates for the number of imported cases in 2019 with official data.

Methods

Travelers’ Distribution

We first developed a theoretical framework to estimate how travelers distribute throughout the territory after their arrival in the country. See Figure 1 for a schematic description of this approach.
Figure 1. Summary of the rationale behind our approach. Infected travelers arrive at the importing country from $L$ exporting countries across the world by means of air travel. For each country $k$, the prevalence of the disease in the country ($p_k$) and the number of travelers arriving from it to the importing country ($a_k$) are combined to obtain an estimate of the number of imported cases ($p_k a_k$). These spread across $R$ regions or spatial units of the importing country following a distribution that can be estimated by means of the “travelers’ index.” The index is computed from local statistics concerning economic and touristic activity and the number of foreign residents at each region. The travelers’ index $v_{ki}$ measures the proportion of imported cases from country $k$ moving to local region $i$ upon arrival at the importing country. The total number of cases expected to arrive at the region ($c_i$) is then obtained as the sum of the estimated number of imported cases over all exporting countries: $c_i = \sum p_k a_k v_{ki}$. AENA: Aeropuertos Españoles y Navegación Aérea (Spanish Airports and Aerial Navigation); GHDx: Global Health Data Exchange; INE: Instituto Nacional de Estadística (Spanish National Statistics Institute).

Input Data

We used the following yearly statistics from 2015 to 2019 to compute the relative appeal of each province in Spain to international travelers, which are publicly available and curated by the Spanish National Statistics Institute [40]:

- **Tourist indicators:** For both hotels and tourist apartments, we used the variables total capacity, number of national travelers, number of foreign travelers, number of overnight stays by national travelers, and number of overnight stays by foreign travelers.

- **Economic indicators:** We considered each province’s population, gross domestic product (GDP), GDP per capita, number of private limited companies (Sociedades Limitadas), and number of public limited companies (Sociedades Anónimas).

- **Indicators for visits to friends and relatives:** For each country and province, we used the number of foreign residents by nationality, number of foreign residents by birthplace, and number of national residents by birthplace (other than Spain).

The following 3 additional data inputs were used in our approach:

- **Arrival data:** We computed the yearly number of travelers arriving in Spain from 2015 to 2019 at each of the 100 airports with the largest flows of incoming travelers and aggregated publicly available monthly data provided by the public entity in charge of the Spanish Airports and Aerial Navigation (Aeropuertos Españoles y Navegación Aérea, AENA) [41].

- **Disease data:** We used yearly prevalence estimates for malaria and dengue from 2015 to 2019 provided by Global Health Data Exchange (GHDx) [42]. Data on malaria were supplied by the Malaria Atlas Project [43].

- **Cases in Spain:** The number of imported cases of malaria and dengue (including any type of infection by the dengue virus) is reported by each province to the Spanish National Surveillance System (Red Nacional de Vigilancia Epidemiológica, RENAVE). These data were used.

Travelers’ Index

We have described the rationale used to combine the above statistics in an informed indicator to estimate the propensity of international travelers to move to a specific region. We considered a country that receives travelers from other countries of the world (these are denoted as importing and exporting countries, respectively). We assumed that the importing country is divided into regions, which may represent geographical regions or administrative units, for instance.

Motivations for international travel are usually classified into 3 major categories: tourism, business, and visits to friends and relatives [44,45]. Based on this principle, we computed the relative importance of each region in the importing country in
terms of each of these drivers as follows: \( t_i \) measures the relative importance of region \( i \) in terms of tourism, \( e_i \) measures the relative importance of region \( i \) in terms of economy, and \( r_{ik} \) measures the relative importance of region \( i \) in terms of visits to friends and relatives for travelers arriving from country \( k \).

The last indicator depends on the exporting country \( k \). For instance, when computing the travelers’ index with \( k=\text{Brazil} \), only the number of Brazilian residents is considered in the computation. The 3 indicators were computed from several available statistics concerning the regions of the country (see the Input Data subsection) as the relative contribution of each region to the country’s total. For instance, when using GDP to measure economic status, the relative importance of region \( i \) in terms of economy would be as follows:

\[ t_i = \frac{\text{GDP of region } i}{\text{total GDP of the importing country}} \tag{1} \]

We then computed the travelers’ index for each region \( i \) of the importing country and any exporting country \( k \) as the average of the 3 indicators as follows:

\[ v_{ik} = \frac{1}{3} \left( t_i + e_i + r_{ik} \right) \tag{2} \]

It follows from equations 1 and 2 that for each exporting country \( k \), the sum of the travelers’ index of the country over all the regions \( i \) of the importing country is as follows:

\[ \sum_{i} v_{ik} \]

Therefore, given an exporting country \( k \), the travelers’ index \( v_{ik} \) may be understood as an estimate of the portion of travelers \( a_k \) arriving from this country to each province \( i \). Hence, we can estimate the total number of travelers arriving at a given region \( i \) as follows:

\[ \sum_{k} a_k p_k v_{ik} \]

where the sum runs over all exporting countries \( k \), and \( a_k \) is the total number of travelers arriving from any such country \( k \). Accordingly, we estimated the number of imported cases of a disease at each province \( i \) as the sum of the imported cases from each of the exporting countries \( k \) as follows:

\[ \sum_{k} a_k p_k v_{ik} \]

where \( p_k \) denotes the prevalence of the disease in the exporting country \( k \). Plainly speaking, we estimated the total number of cases arriving from country \( k \) as the product of the total number of travelers arriving from the country and the prevalence of the disease in the country. We then estimated how these cases disseminate across regions by means of the travelers’ index, which assigns to travelers a relative importance or preference for each of the regions. Adding these local distributions over all the exporting countries resulted in the total number of expected cases at each of the regions in the importing country. See Figure 1 for a visual representation of this reasoning.

**Model Calibration and Validation**

In order to test the validity of our approach, we followed the pipeline depicted in Figure 2. The steps involved in this process were (1) input variables, (2) model fitting, (3) model selection, and (4) model validation and assessment.

**Input Variables**

We considered data from Spain (importing country) and its 52 provinces (regions over which the imported cases disseminate). We computed the relative importance of \( t_i, e_i, \) and \( r_{ik} \) using each of the statistics listed in the Input Data subsection for each of the 3 drivers (ie, tourism, economy, and visits to friends and family).

**Model Fitting**

We used equation 2 to construct the travelers’ index (1 for each possible combination of the indicators). We combined these with arrivals and prevalence data to obtain estimates for the expected number of cases at each province for 2015-2018 (equation 4), resulting from simple averages of the indicators \( t_i, e_i, \) and \( r_{ik} \). We also considered a generalized version of the travelers’ index by replacing the average in equation 2 with a weighted average as follows:

\[ t_i = \text{GDP of region } i / \text{total GDP of the importing country} \tag{1} \]

**Model Selection**

To identify which travelers’ index best approximates the number of imported cases per province, we computed the correlation

\[ 0 \leq a_j \leq 1 \text{ and } a_1 + a_2 + a_3 = 1. \]
between the estimates provided by each of our models (1 for each combination of indicators) and the actually reported cases of malaria and dengue at each province for the years 2015 to 2018. The model reporting the highest value for this correlation was selected as the best model. We followed the same procedure for the case of the weighted averages, and a different estimate was obtained for each choice of the indicator and weight $a_j$.

**Model Validation and Assessment**

As a final test for accuracy, we computed the correlation between the best model’s estimates for 2019 (data not used during the fitting and selection process) and the officially reported cases for this year, both for the simple and weighted averages. In case this correlation was high, we considered the model as validated and proceeded to the next step.

We assessed 3 features of the resulting model. First, we fit a linear model explaining the estimated number of imported cases at each province in 2019 in terms of the officially reported cases. The coefficient of the linear model may be understood as the number of cases predicted by the model per officially reported case, thus informing of the overestimation or underestimation of the model’s prediction of actually reported cases. This refers only to the raw number of cases as the accuracy of the distribution is captured by the validated correlation. Second, we ranked the contribution of each of the statistics considered in the model by computing the average correlation with 2019 official data of those estimates obtained from models including each particular variable. We also computed the average loss of accuracy associated with each variable as the difference in the average correlation of those models including and not including each statistic. This allowed us to identify which statistics among the choices made for each indicator $t_i, e_i$, and $r_{ij}$ provided more reliable predictors of the disease. Third, we followed an analogous procedure for the assessment of the weighted averages and computed the average correlation of those models built from each choice of the indicator and weight $a_j$. Those indicators scoring higher for larger values of the corresponding weight were expected to inform about the motivations of the travelers carrying each disease among business, tourism, and visits to friends and relatives.

**Human Mobility Model**

Finally, we tested the validity of our results against a well-established model for human mobility [46,47], assuming that the movement of travelers does not follow motivations different from those of resident populations. For this, we assumed that, on arrival to their destination airport in Spain, a proportion $q$ of imported cases stay at that destination and the rest move to a different province following a well-known, generic, and random human mobility pattern [47,48]. The probability of these travelers moving to each province in Spain is assumed to follow a decaying power law with exponent $\gamma$ on the distance $d$ between the origin and destination province centroids ($p(d) = d^{-\gamma}$).

We then grouped the total number of expected cases at each province as the sum of those arriving at the province according to their final flight destination and those arriving from any other province by means of other transport modes reflected in a geographically bounded power law distribution. This model was then fitted to the data on the officially reported cases of dengue and malaria from 2015 to 2018 for values of $q$ between 0 and 1 (proportion of travelers who leave their destination province upon arrival) and values of $\gamma$ between 1 and 5 (exponent of the power law, with lower values favoring longer-ranged movement and higher values favoring shorter-ranged movement).

We followed an analogous philosophy for the model building and assessment process as in the travelers’ index model. The parameters leading to the highest correlation with the reported cases for the period of 2015 to 2018 were used to compute an estimate for 2019, and the correlation between this estimate and the 2019 official record was then computed to allow for comparison with the travelers’ index model. A linear model between the human mobility model’s estimate and the official 2019 record was also fitted to assess the underestimation or overestimation of the model.

**Ethical Considerations**

Our study used publicly available aggregated secondary data with no characteristics that allowed for individual identification. There are no relevant data protection and privacy issues to report.

**Results**

**Input Data**

A preliminary analysis showed that among all statistics used, those concerning the same drivers were usually highly co-dependent, with some exceptions (eg, GDP per capita; see Figures S1 and S2 in Multimedia Appendix 1 [40]). The relative importance appearing in the computation of the travelers’ index (equation 2) showed little variation over the years (see Figure S3 in Multimedia Appendix 1). This temporal stability has been observed before in the distribution of international [37] and national [49] human mobility flows across destinations, which have been obtained from expressions analogous to equation 1.

The 100 airports with the highest number of incoming travelers were located in 49 countries and accounted for 99.75% of the total incoming travelers to Spain from 2015 to 2019. Out of these countries, 10 were removed from our study as neither malaria nor dengue was present during the time span under study (according to prevalence data from GHDx), resulting in 39 exporting countries. Table 1 shows the number of incoming travelers from each of these countries and the average prevalences of dengue and malaria from 2015 to 2019 as provided by GHDx.
<table>
<thead>
<tr>
<th>Country</th>
<th>Incoming travelers</th>
<th>Malaria prevalence (/100,000 population)</th>
<th>Dengue prevalence (/100,000 population)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>6,845,337</td>
<td>0</td>
<td>0.54</td>
</tr>
<tr>
<td>Brazil</td>
<td>3,472,397</td>
<td>78.96</td>
<td>65.40</td>
</tr>
<tr>
<td>Colombia</td>
<td>3,053,635</td>
<td>197.47</td>
<td>54.97</td>
</tr>
<tr>
<td>Argentina</td>
<td>2,972,990</td>
<td>0</td>
<td>15.47</td>
</tr>
<tr>
<td>Peru</td>
<td>2,101,286</td>
<td>323.49</td>
<td>41.37</td>
</tr>
<tr>
<td>Mexico</td>
<td>2,063,577</td>
<td>7.26</td>
<td>37.29</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>1,611,336</td>
<td>1.50</td>
<td>49.11</td>
</tr>
<tr>
<td>Algeria</td>
<td>1,562,772</td>
<td>4.92</td>
<td>0</td>
</tr>
<tr>
<td>Venezuela</td>
<td>1,241,154</td>
<td>1065.28</td>
<td>48.17</td>
</tr>
<tr>
<td>Cuba</td>
<td>1,062,531</td>
<td>0</td>
<td>37.93</td>
</tr>
<tr>
<td>Cape Verde</td>
<td>1,054,106</td>
<td>135.17</td>
<td>36.41</td>
</tr>
<tr>
<td>Ecuador</td>
<td>1,027,244</td>
<td>86.97</td>
<td>37.84</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>810,214</td>
<td>0</td>
<td>67.74</td>
</tr>
<tr>
<td>Senegal</td>
<td>630,192</td>
<td>2464.71</td>
<td>32.43</td>
</tr>
<tr>
<td>Panama</td>
<td>566,765</td>
<td>115.70</td>
<td>56.47</td>
</tr>
<tr>
<td>Bolivia</td>
<td>556,092</td>
<td>181.43</td>
<td>60.84</td>
</tr>
<tr>
<td>Gambia</td>
<td>522,809</td>
<td>4237.62</td>
<td>33.78</td>
</tr>
<tr>
<td>Egypt</td>
<td>466,705</td>
<td>0</td>
<td>10.88</td>
</tr>
<tr>
<td>Thailand</td>
<td>365,318</td>
<td>68.10</td>
<td>58.10</td>
</tr>
<tr>
<td>Singapore</td>
<td>360,619</td>
<td>0</td>
<td>68.44</td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td>360,393</td>
<td>32981.86</td>
<td>39.18</td>
</tr>
<tr>
<td>China</td>
<td>359,372</td>
<td>0.14</td>
<td>24.63</td>
</tr>
<tr>
<td>Pakistan</td>
<td>306,321</td>
<td>536.27</td>
<td>41.90</td>
</tr>
<tr>
<td>Mauritania</td>
<td>300,521</td>
<td>4095.70</td>
<td>24.57</td>
</tr>
<tr>
<td>El Salvador</td>
<td>293,976</td>
<td>8.95</td>
<td>130.85</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>254,457</td>
<td>15.80</td>
<td>0</td>
</tr>
<tr>
<td>Nigeria</td>
<td>200,813</td>
<td>18792.47</td>
<td>38.73</td>
</tr>
<tr>
<td>Jordan</td>
<td>182,431</td>
<td>0</td>
<td>12.45</td>
</tr>
<tr>
<td>Angola</td>
<td>176,690</td>
<td>11182.79</td>
<td>27.97</td>
</tr>
<tr>
<td>Guatemala</td>
<td>174,030</td>
<td>169.12</td>
<td>45.25</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>166,730</td>
<td>3.98</td>
<td>15.52</td>
</tr>
<tr>
<td>Ghana</td>
<td>149,374</td>
<td>18512.96</td>
<td>40.65</td>
</tr>
<tr>
<td>Guinea</td>
<td>78,545</td>
<td>30131.12</td>
<td>38.35</td>
</tr>
<tr>
<td>The Bahamas</td>
<td>49,010</td>
<td>0</td>
<td>39.30</td>
</tr>
<tr>
<td>Gabon</td>
<td>38,016</td>
<td>15756.90</td>
<td>44.25</td>
</tr>
<tr>
<td>Jamaica</td>
<td>14,802</td>
<td>0</td>
<td>45.43</td>
</tr>
<tr>
<td>South Africa</td>
<td>12,394</td>
<td>36.43</td>
<td>0</td>
</tr>
<tr>
<td>Cameroon</td>
<td>8994</td>
<td>19904.66</td>
<td>34.96</td>
</tr>
<tr>
<td>Mali</td>
<td>1165</td>
<td>16024.21</td>
<td>31.19</td>
</tr>
</tbody>
</table>

aCountries with no malaria or dengue prevalence have been removed from the list.
bTotal incoming travelers and average malaria and dengue prevalences (total cases per 100,000 inhabitants) from 2015 to 2019 for the 39 exporting countries.
Estimates and Model Assessment

High correlation values were found for both malaria (0.94) and dengue (0.86) between the best model’s estimates for 2019 and the notified cases. The models that provided the most accurate estimates included public limited companies, foreign travelers at hotels, and foreign residents by birthplace in the computation of the travelers’ index. The same variables led to the best estimates for both malaria and dengue. While considering weighted averages in the construction of the travelers’ index did not improve the accuracy of the models, different motivations were obtained for travelers carrying each of the diseases: economy seemed to best capture the appeal of each region for imported cases of malaria (relative weight of 0.7, with GDP being the most accurate indicator) and visits to friends and relatives seemed to be the main motivation for travelers with dengue (relative weight of 0.9, assigned to the number of foreign residents in the province by birthplace). Different proportions of overestimation were found for each disease (99% for malaria and 86.5% for dengue). A summary of the relevant features of the models provided by the fitting and selection process is presented in Table 2.

Table 2. Summary of the models that most accurately approximated the reported cases in 2015-2018.

<table>
<thead>
<tr>
<th>Disease (model)</th>
<th>Economic indicator (weighta)</th>
<th>Tourist indicator</th>
<th>Visits to friends and relatives indicator (weighta)</th>
<th>Pearson correlation of model’s estimate with 2019 data</th>
<th>Overestimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malaria (simple)</td>
<td>Public limited companies</td>
<td>Foreign travelers at hotels</td>
<td>Foreign residents by birthplace</td>
<td>0.94</td>
<td>98.9%</td>
</tr>
<tr>
<td>Malaria (weighted)</td>
<td>GDP (0.7)</td>
<td>Foreign travelers at hotels (0.1)</td>
<td>Foreign residents by birthplace (0.2)</td>
<td>0.94</td>
<td>99.0%</td>
</tr>
<tr>
<td>Dengue (simple)</td>
<td>Public limited companies</td>
<td>Foreign travelers at hotels</td>
<td>Foreign residents by birthplace</td>
<td>0.86</td>
<td>86.5%</td>
</tr>
<tr>
<td>Dengue (weighted)</td>
<td>No contribution (0)</td>
<td>Foreign travelers at hotels (0.1)</td>
<td>Foreign residents by birthplace (0.9)</td>
<td>0.87</td>
<td>86.7%</td>
</tr>
</tbody>
</table>

aFor the models including weighted averages, the weight $a_i$ of each indicator is included in parenthesis. If the weight of a given indicator is 0, no contribution to the estimate is provided by the corresponding indicator.

bEach row shows the statistics that provide the best estimate of imported cases of each disease, the correlation with the actually reported data in 2019, and the approximation for the proportion of overestimation as obtained from the linear models.

cGDP: gross domestic product.

Figure 3 shows the fit of the weighted models and their estimates for 2019, together with the officially reported number of cases of each disease (malaria and dengue) at each province in Spain. Upon visual inspection, 2 provinces seemed to have a high influence on the fit of the models. These corresponded to Madrid and Barcelona, which hosted a much larger number of reported cases of both diseases. We excluded these provinces from the input data set and repeated the analysis (Table S1 and Figure S4 in Multimedia Appendix 1). While a decrease in correlation was found overall (approximately 0.12 over all models), the resulting estimates still showed high agreement with the official report for 2019 (above 0.74 correlation with 2019 data; see Figure 3, and Table S1 and Figure S4 in Multimedia Appendix 1).
Figure 3. Summary of the best linear models for 2019 imported cases of malaria (top row) and dengue (bottom row). The left column shows the predictions of the models (in red), together with the number of reported cases (in blue) for 2019 at each province in Spain. The right column shows the fit between the estimates of the models and the official records (inset figures correspond to the fit after removing Madrid and Barcelona from the data set).

We performed a residual analysis to check for normality and autocorrelation of the residuals of the models. The malaria model showed close-to-normal residuals with no autocorrelation (statistically significant $W=0.67$ and $DW=2.03$ in the Shapiro-Wilk and Durbin-Watson tests, respectively). For the dengue model, a relevant deviation was caused by the estimate for Barcelona (Figure 3). Exclusion of this outlier resulted in normally distributed and not autocorrelated residuals (statistically significant $W=0.94$ and $DW=1.87$). See Table S3 in Multimedia Appendix 1 for complete details on the residual analysis.

Variable Performance

The models constructed using simple averages provided a unanimous choice of indicators associated with tourism, economy, and visits to friends and relatives. On the contrary, the best weighted models included different economic indicators. GDP provided the best estimate for imported malaria cases, while no influence of the economic indicator was considered in the best dengue model. In addition, different drivers were the most important ones for each disease, as shown by the much higher relative weight for economic motivations in malaria cases and for visits to friends and relatives in dengue cases (Table 2).

When ranking the contribution of each of the variables to the accuracy of the models, similar results were found for both diseases, with some minor variations across variables (Table 3). Several statistics concerning tourism ranked the highest in this classification, although several others ranked in a low position, indicating that appropriate choices of indicators may be important and may need careful examination. All economic indicators provided an improvement (or absence of a decrease) in correlation, except for GDP per capita, which resulted in less accurate estimates (average decrease of approximately 0.08 in correlation with 2019 data). Indicators corresponding to visits to friends and relatives had mild average effects on the outputs of the models (the largest variation in correlation with 2019 data was $-0.02$).
Table 3. Contribution of each variable to model accuracy.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Malaria&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Malaria (loss)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Dengue&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Dengue (loss)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>National travelers in hotels</td>
<td>0.83</td>
<td>0.08</td>
<td>0.74</td>
<td>0.05</td>
</tr>
<tr>
<td>Overnight stays by national travelers in hotels</td>
<td>0.83</td>
<td>0.07</td>
<td>0.73</td>
<td>0.05</td>
</tr>
<tr>
<td>Foreign travelers in hotels</td>
<td>0.82</td>
<td>0.07</td>
<td>0.76</td>
<td>0.08</td>
</tr>
<tr>
<td>Total hotel capacity</td>
<td>0.81</td>
<td>0.05</td>
<td>0.74</td>
<td>0.05</td>
</tr>
<tr>
<td>Public limited companies (Sociedades Anónimas)</td>
<td>0.81</td>
<td>0.05</td>
<td>0.72</td>
<td>0.04</td>
</tr>
<tr>
<td>National travelers in tourist apartments</td>
<td>0.79</td>
<td>0.03</td>
<td>0.70</td>
<td>0.01</td>
</tr>
<tr>
<td>Overnight stays by national travelers in tourist apartments</td>
<td>0.79</td>
<td>0.03</td>
<td>0.70</td>
<td>0.02</td>
</tr>
<tr>
<td>GDP&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.78</td>
<td>0.02</td>
<td>0.71</td>
<td>0.02</td>
</tr>
<tr>
<td>Private limited companies (Sociedades Limitadas)</td>
<td>0.78</td>
<td>0.02</td>
<td>0.70</td>
<td>0.01</td>
</tr>
<tr>
<td>Foreign residents by country of nationality</td>
<td>0.77</td>
<td>0.01</td>
<td>0.70</td>
<td>0.01</td>
</tr>
<tr>
<td>Foreign residents by country of birth</td>
<td>0.77</td>
<td>0.01</td>
<td>0.70</td>
<td>0.01</td>
</tr>
<tr>
<td>Population</td>
<td>0.76</td>
<td>0.00</td>
<td>0.69</td>
<td>0.00</td>
</tr>
<tr>
<td>Overnight stays by foreign travelers in hotels</td>
<td>0.75</td>
<td>-0.01</td>
<td>0.69</td>
<td>0.00</td>
</tr>
<tr>
<td>National residents by country of birth</td>
<td>0.75</td>
<td>-0.02</td>
<td>0.68</td>
<td>-0.02</td>
</tr>
<tr>
<td>Total tourist apartment capacity</td>
<td>0.71</td>
<td>-0.06</td>
<td>0.64</td>
<td>-0.06</td>
</tr>
<tr>
<td>GDP per capita</td>
<td>0.69</td>
<td>-0.09</td>
<td>0.63</td>
<td>-0.07</td>
</tr>
<tr>
<td>Foreign travelers in tourist apartments</td>
<td>0.67</td>
<td>-0.10</td>
<td>0.62</td>
<td>-0.08</td>
</tr>
<tr>
<td>Overnight stays by foreign travelers in tourist apartments</td>
<td>0.63</td>
<td>-0.15</td>
<td>0.58</td>
<td>-0.13</td>
</tr>
</tbody>
</table>

<sup>a</sup>The average correlation of the estimates of the models including each variable in their fit with the officially reported 2019 data.

<sup>b</sup>The average difference in correlation between models including each variable in their fit and models not including each of the variables (variables ranked by the average correlation for predictions).

<sup>c</sup>GDP: gross domestic product.

A similar procedure was followed for the weighted models. The average correlation between the models including each variable and the 2019 official data was computed in this case with stratification by the weight assigned to the variable (Figure 4). In addition to the ranking of variables (similar to data in Table 3), this provided a measure of the variability of each variable’s contribution to the accuracy of the model in terms of the weight assigned to it. Smaller overall variations in model accuracy were identified for variables measuring visits to friends and relatives, for instance, while much larger variability was recorded for some tourist indicators. This shows the higher potential loss in accuracy that would result from including these variables in the models than including other variables.
**Figure 4.** Summary of each input variable’s performance on the estimates for malaria (A) and dengue (B). Each square in the figure is colored according to the average correlation between the official 2019 reports and the estimates provided by the weighted models including each of the variables, with the associated weight ranging from 0 (no contribution from the variable is assumed in the model) to 1 (the model only includes that variable). The variables are ranked from top to bottom according to the overall average correlation with 2019 data of the estimates of the models including each variable. GDP: gross domestic product; SA: Sociedades Anónimas; SL: Sociedades Limitadas.

### Comparison With a Generic Mobility Model

For both dengue and malaria, the human mobility models ranked higher in terms of correlation with 2019 data for higher values of the assumed proportion of travelers who do not move from their destination province upon arrival ($q$; see Figures S5 and S6 in Multimedia Appendix 1). Models also favored the choice of smaller values of the exponent of the power law distribution (highest average correlation with 2019 data for $\gamma=1$), indicating that longer movements may take place if a displacement occurs after arrival. Much higher variability in the correlation with...
Discussion

We computed estimates for the number of imported cases of malaria and dengue at each province in Spain based on simple methodological assumptions. Our approach makes use of readily available data and provides approximations of the actually declared number of cases of the disease. This advance may contribute to the adequate modeling and monitoring of VBDs, which might be relevant for effective outbreak prevention strategies. More efficient resource allocation strategies for both vector control and disease prevention can be designed if reliable predictions of the geographical locations of imported cases are available. By circumventing the need for detailed large-scale data on human mobility or traveler behavior, this methodology is accessible and suitable to be used in countries lacking more exhaustive data infrastructure, for instance [39,50]. The reasoning presented here could also be generalized to other choices of territories.

The high correlation found between our estimates and real data support the validity of our approach based on a priori theoretical conceptualization. This agreement in trend suggests that our estimates are reliable enough for the elaboration of scale-less risk indicators, for instance. On the other hand, our estimates of the raw number of imported cases were simplistic (product of yearly prevalence and total number of travelers), which resulted in substantial overestimation of the number of imported cases. For the case of malaria, this is coherent with the epidemiology of the disease, being more severe unless treatment is available and having a higher incidence in economically deprived populations [51]. These factors may prevent individuals with malaria from engaging in international displacements. For dengue, however, the identified overestimation (8 predicted cases per notified case) lies relatively close to previously obtained estimates of the underreporting of cases in other contexts [52]. This suggests that our approach could also provide a valid method for assessing the sensibility of epidemiological systems. In any case, our focus was on assessing the validity of the travelers’ index as a method to improve risk analysis, rather than developing a predictive model for imported cases of the diseases.

The proposed computation of the key indicators involved in our model (the travelers’ index \(v_B\)) has the advantage of being partially robust considering errors in declaration or incomplete data collection. Indeed, as these only involve the relative importance of each region in the country, correction factors are unnecessary in our approach, and incomplete data will yield equally valid estimates as long as the underreporting can be assumed to be comparable for all regions. Moreover, the little variation in time shown by these quantities (see Figure S3 in Multimedia Appendix 1) could allow for reliable estimates even when only past statistics are available.

A key finding in this direction is that while the impact of each particular indicator in the quality of the estimate was similar for both diseases, the relevant drivers for case importation were different (economic motivations for malaria cases and visits to friends and relatives for dengue cases). This may be due to the different nature of the motivation for international travel across countries in the world. Most malaria cases were imported from African countries, while travelers carrying dengue usually arrived from America or Asia (see Table S1 in Multimedia Appendix 1). Travelers arriving from these continents are expected to follow different motivations for international displacement. Actually, malaria cases imported to Spain in the pre-pandemic era were mainly due to visits to friends and relatives or migration in almost 75% of cases, corresponding to travelers following economic motivations [53]. On the other hand, dengue cases were imported mainly by tourist travelers or visits to friends and relatives [54].

Further evidence of the appropriateness of our approach was provided by a comparison with the human mobility model. While the validity of this model has been established in many contexts and is widely acknowledged [47], it provided much

---

Table 4. Summary of the human mobility model that most accurately approximated the reported cases from 2015 to 2018 (including all provinces).

<table>
<thead>
<tr>
<th>Model</th>
<th>Proportion of cases that do not move after arrival (q)</th>
<th>Exponent of the power law distribution (γ)</th>
<th>Correlation with 2019 data</th>
<th>Overestimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malaria(^a)</td>
<td>1</td>
<td>Any</td>
<td>0.59</td>
<td>99.5%</td>
</tr>
<tr>
<td>Dengue(^a)</td>
<td>1</td>
<td>Any</td>
<td>0.66</td>
<td>95.2%</td>
</tr>
</tbody>
</table>

\(^a\)Each row shows the parameters of the model that provide the best estimate of imported cases of each disease, the correlation with the actually reported data in 2019, and the overestimation of the models as obtained from the linear fit with official records.
less accurate estimates and was highly sensitive to data retrieval from provinces with a larger number of disease cases. This demonstrates the need to consider specific designs that take into account travelers’ behaviors and differentiate them from general resident population dynamics.

Future developments of our approach should cover the following improvements:

- Coupling with postimportation dynamics: Our framework could be integrated into more complicated models incorporating transmission dynamics that involve the life cycle of the disease within the vector and the host [55-58]. Well-developed approaches, such as compartmental models, could benefit from more precise estimates on the expected location of arrival of imported cases of diseases.

- Refining the gross estimate of imported cases: As mentioned above, we computed simple estimates for the total number of cases arriving to the importing country (product of yearly prevalence and total number of travelers). We focused on how these cases distribute over the regions of the importing country. Consideration of more elaborate estimates of these quantities or the local distribution of the disease in the exporting countries would probably yield more precise final estimates.

- Extending the scope of the model to other diseases: A particular feature of the treated examples is that virtually all incoming streams of travelers into Spain from regions where malaria and dengue are endemic, which may result in transmission, may be assumed to be associated with air travel. This may not be the case for other diseases and countries, for which detailed data on the total traveler flow or further development of the proposed methodology could be necessary. Similarly, other importation phenomena that may depend on human behavior or allocation of resources could be analyzed under our assumptions, such as passive mobility of vectors by human means of transportation [59] or migratory flows [60,61].

It should be noted that our model was focused on countries with high dengue and malaria prevalences, and hence, they were likely to export these diseases to Spain. However, this concept could be generalized to other types of risk-related importation scenarios like the transport of new vectors or exotic species (invasion biology), which is another crucial process in the spread of VBDs.

Several factors may be limiting the extent of our results. First, both malaria and dengue are diseases known to be subject to high underreporting [51,52]. Second, we validated our models with annual data, as data on the number of monthly reported cases were too noisy. In any case, model predictions could be generated at a higher temporal resolution by incorporating monthly numbers of arriving travelers in the country, for instance. Third, our model was designed to address the motivations of international travelers; however, a significant number of imported cases may correspond to national travelers returning to the country or individuals from other nonendemic regions, especially for dengue. It would be desirable to devise an accurate method to differentiate between these 2 types of travelers and incorporate both motivations in the model. Finally, geographical borders are not always the best spatial human structure [62], and therefore, the availability of data with a finer geographical resolution could result in significant improvements in our estimates. We also note that our model has been validated with data obtained prior to the COVID-19 pandemic, and differences may arise in the postpandemic era. Therefore, further validation with future data is desirable.

We have shown the validity of the travelers’ index as a method to estimate the distribution of imported cases of malaria and dengue from endemic regions. This is an appropriate way to improve disease risk prediction on the basis of human mobility patterns. Our methodology adds value to available socioeconomic information relevant to public health. Nonetheless, human mobility is just 1 component of VBD risk models. The other key components that need to be added are vector (mosquito) distribution and suitability. Our work will be combined with multi-sourced presence/absence and suitability vector data in Spain, including both authoritative and citizen science data collections [63], and integrated into the Spanish National Surveillance System for VBDs. Pairing the risk of importation of cases and the risk of local transmission through the presence of vectors will provide a more comprehensive evaluation of the threats posed by VBDs to public health.

Acknowledgments

The project leading to the present results has received funding from “la Caixa” Foundation (ID 100010434) under agreement HR-18-0036.

Data Availability

Indicators for tourism, economy, and visits to friends and relatives are available from the Spanish National Institute for Statistics [40]. Arrival data are available from the Spanish Aviation and Airports Office [41]. Disease prevalence data for exporting countries are available from Global Health Data Exchange [42]. Reported cases of dengue and malaria in Spain are collected by the National Surveillance System and managed by the Spanish National Center for Epidemiology, and the data are available upon reasonable request (contact email: vigilancia.cne@isciii.es).

Conflicts of Interest

None declared.
References


40. INE Instituto Nacional de Estadística. URL: https://www.ine.es/ [ accessed 2022-03-15 ]


Abbreviations

GDP: gross domestic product
GHDX: Global Health Data Exchange
VBD: vector-borne disease
Modeling the Regional Distribution of International Travelers in Spain to Estimate Imported Cases of Dengue and Malaria: Statistical Inference and Validation Study

JMIR Public Health Surveill 2024;10:e51191

URL: https://publichealth.jmir.org/2024/1/e51191
doi:10.2196/51191
PMID:38801767
Updated Surveillance Metrics and History of the COVID-19 Pandemic (2020-2023) in Europe: Longitudinal Trend Analysis

Alexander L Lundberg¹,², PhD, MSc; Scott A Wu³, BA; Alan G Soetikno³, BS; Claudia Hawkins⁴,⁵, MD, MPH; Robert L Murphy⁴,⁶, MD; Robert J Havey⁶,⁷, MD; Egon A Ozer⁴,⁸, MD, PhD; Charles B Moss⁹, PhD; Sarah B Welch¹,², MPH; Maryann Mason¹,², PhD; Yingxuan Liu¹,², MA; Lori A Post¹,², PhD

¹Buehler Center for Health Policy and Economics, Robert J. Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States
²Department of Emergency Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
³Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
⁴Department of Medicine, Division of Infectious Diseases, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
⁵Center for Global Communicable and Emerging Infectious Diseases, Robert J. Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States
⁶Robert J. Havey, MD Institute for Global Health, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
⁷Department of Medicine, General Internal Medicine and Geriatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
⁸Center for Pathogen Genomics and Microbial Evolution, Robert J. Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States
⁹Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States

Corresponding Author:
Lori A Post, PhD
Buehler Center for Health Policy and Economics
Robert J. Havey, MD Institute for Global Health
Northwestern University
420 E. Superior
Chicago, IL, 60611
United States
Phone: 1 3125031706
Email: lori.post@northwestern.edu

Abstract

Background: In this study, we built upon our initial research published in 2020 by incorporating an additional 2 years of data for Europe. We assessed whether COVID-19 had shifted from the pandemic to endemic phase in the region when the World Health Organization (WHO) declared the end of the public health emergency of international concern on May 5, 2023.

Objective: We first aimed to measure whether there was an expansion or contraction in the pandemic in Europe at the time of the WHO declaration. Second, we used dynamic and genomic surveillance methods to describe the history of the pandemic in the region and situate the window of the WHO declaration within the broader history. Third, we provided the historical context for the course of the pandemic in Europe in terms of policy and disease burden at the country and region levels.

Methods: In addition to the updates of traditional surveillance data and dynamic panel estimates from the original study, this study used data on sequenced SARS-CoV-2 variants from the Global Initiative on Sharing All Influenza Data to identify the appearance and duration of variants of concern. We used Nextclade nomenclature to collect clade designations from sequences and Pangolin nomenclature for lineage designations of SARS-CoV-2. Finally, we conducted a 1-tailed t test for whether regional weekly speed was greater than an outbreak threshold of 10. We ran the test iteratively with 6 months of data across the sample period.

Results: Speed for the region had remained below the outbreak threshold for 4 months by the time of the WHO declaration. Acceleration and jerk were also low and stable. While the 1-day and 7-day persistence coefficients remained statistically significant, the coefficients were moderate in magnitude (0.404 and 0.547, respectively; P<.001 for both). The shift parameters for the 2 weeks around the WHO declaration were small and insignificant, suggesting little change in the clustering effect of cases on
future cases at the time. From December 2021 onward, Omicron was the predominant variant of concern in sequenced viral samples. The rolling test of speed equal to 10 became insignificant for the first time in April 2023.

**Conclusions:** While COVID-19 continues to circulate in Europe, the rate of transmission remained below the threshold of an outbreak for 4 months ahead of the WHO declaration. The region had previously been in a nearly continuous state of outbreak. The more recent trend suggested that COVID-19 was endemic in the region and no longer reached the threshold of the pandemic definition. However, several countries remained in a state of outbreak, and the conclusion that COVID-19 was no longer a pandemic in Europe at the time is unclear.

(JMIR Public Health Surveill 2024;10:e53551) doi:10.2196/53551

**KEYWORDS**

Europe; COVID-19; history of the pandemic; method of the moments; Arellano-Bond estimators; Albania; Andorra; Austria; Belarus; Belgium; Bosnia and Herzegovina; Bulgaria; Croatia; the Czech Republic; Denmark; Estonia; Finland; France; Germany; Greece; Greenland; Hungary; Iceland; Ireland; the Isle of Man; Italy; Latvia; Liechtenstein; Lithuania; Luxembourg; Moldova; Monaco; Montenegro; the Netherlands; Norway; Poland; Portugal; Romania; San Marino; Serbia; Slovakia; Slovenia; Spain; Sweden; Switzerland; Ukraine; the United Kingdom; the Vatican City

**Introduction**

**Background**

COVID-19, the disease caused by the virus SARS-CoV-2, was first detected in Wuhan, China, in the fall of 2019 [1-5]. The first European [6] case of COVID-19 was reported in France on January 24, 2020, with additional cases reported in Germany and Finland soon afterward [7,8]. Our research team conducted an analysis of the pandemic in Europe 1 year into the pandemic [9]; this study provides 2 additional years of updated surveillance and analysis for the region.

We adopt the World Bank’s definition of Europe, which is based on economic development and geographical proximity, encompassing Albania, Andorra, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Greenland, Hungary, Iceland, Ireland, the Isle of Man, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Moldova, Monaco, Montenegro, the Netherlands, Norway, Poland, Portugal, Romania, San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, the United Kingdom, and the Vatican City [6].

The World Health Organization (WHO) and Director-General Tedros Ghebreyesus declared the end of COVID-19 as a public health emergency of international concern on May 5, 2023 [10-12], based on the recommendation of the COVID-19 Emergency Committee [12]. To that end, we compared how the pandemic was progressing before and after the declaration.

**Empirical Definition of Pandemic Versus Epidemic Versus Outbreak Versus Endemic**

Epidemiological terms, such as pandemic, epidemic, outbreak, and endemic, are used to describe the occurrence and spread of diseases [13,14]. The distinctions between these terms lie in their scope, geographic extent, and severity. An epidemic refers to a sudden increase in the number of disease cases in a specific population or region. If the epidemic spreads across several countries or continents, it becomes a pandemic. An outbreak, on the other hand, describes a sudden increase in a concentrated setting, usually involving a more limited geographic area than an epidemic. Endemic refers to the constant presence of a disease in a particular geographic region or population, with no sudden increases in case volume [15,16].

**Traditional Surveillance Versus Enhanced Surveillance**

Public health surveillance is the “ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning and evaluation of public health practice” [17]. Surveillance not only explains the burden of death and disease due to a virus but also generates research questions and guides researchers on topics that require further investigation [18-32]. Surveillance allows us to compare the burden of disease between geographical regions and to understand which regions are most impacted. The impact can be measured through rates of how many people contract a disease and how many die, as well as the affiliated costs.

However, traditional surveillance carries several limitations that this study had addressed. Traditional surveillance provides a snapshot of what has already happened [18-32], meaning surveillance is static and only considers the past. In the middle of a burgeoning pandemic, policy makers and public health practitioners also need to understand what is about to happen. Is an outbreak increasing? Will growth switch from linear to exponential? Are more people dying from that particular condition in one place than another? To inform health policy and practice, knowledge of what is about to happen is often more valuable than knowledge of what did happen. To that end, we have developed enhanced surveillance metrics that reflect the dynamics of a pandemic and inform imminent growth, most importantly, where along the epidemiological outbreak curve a particular region is situated. We have also included dynamic metrics about the speed of the pandemic at the national, regional, and global levels and measured how the acceleration of speed this week compared to last week, as well as how novel infections last week would predict new cases this week. We can think of the latter measure as the echoing forward of cases. These metrics were tested and validated in prior research [9,33-43].

The novel metrics add acceleration, jerk, and 1-day and 7-day persistence to the traditional measure of speed. The rate of new COVID-19 cases per 100,000 population is the “speed” of the pandemic. Acceleration is the difference in speed from one unit to the next. We define the rolling test of speed equal to 10 as becoming insignificant for the first time in April 2023.
of time to the next. A positive (negative) acceleration means cases are rising (falling), and an acceleration of 0 indicates an inflection point or stable speed. From physics nomenclature, “jerk” is the change in acceleration from one time interval to the next. A positive jerk may indicate explosive growth in a disease. Finally, 1-day and 7-day persistence measures capture the impact of the 1-day and 7-day lag of speed on current speed. These measures derive from an Arellano-Bond dynamic panel data model, and they capture the echo-forward effect of COVID-19 cases on future cases either 1 or 7 days later [44].

This research team used these metrics to effectively analyze the role of economic reopening on COVID-19 transmissions [42]. These metrics also provided the status of the pandemic in global regions, including Europe [9]. Finally, they helped quickly identify the emergence of the Omicron variant, and they were used in policy briefs throughout the pandemic [45].

For the purpose of this study, standard surveillance metrics explain what has already happened in Europe, while enhanced surveillance metrics speak to what is about to happen or where along an epidemiological curve a country may sit. We used both types of metrics to analyze the possible end to the pandemic.

Objectives

This study has 3 objectives. First, we aimed to measure whether there was an expansion or contraction in the pandemic in Europe when WHO declared the end of the COVID-19 pandemic as a public health emergency of international concern on May 5, 2023. At both the region and country levels, we used advanced surveillance and analytical techniques to describe the status of the pandemic in a 2-week window around the WHO declaration. From a public health perspective, we need to know whether the rate of new COVID-19 cases was increasing, decreasing, or stable from week to week and whether any changes in the transmission rate indicated an acceleration or deceleration of the pandemic. Statistical insignificance is significant; it can signal the epidemiological “end” to the pandemic if the rate of new cases is 0 (or very low) and stable, meaning the number of new cases is neither accelerating nor decelerating.

Second, we used dynamic and genomic surveillance methods to describe the history of the pandemic in the region and situate the time window around the WHO declaration within the broader history. We included the ratio of COVID-19 deaths to the number of transmissions as a proxy for the mortality risk from infection at the population level. We also included a historical record of genomic surveillance from sequenced viral specimens to identify the appearance and spread of variants of concern in the region.

Third, we aimed to provide historical context for the course of the pandemic in Europe. We addressed several questions. How did countries respond to the pandemic? How did the region fare in terms of disease burden? Furthermore, what social, economic, and political factors shaped the course of COVID-19 in the region? This context can provide important lessons for disease prevention and mitigation in future pandemics.

Methods

Overview

This study conducted trend analyses with longitudinal COVID-19 data from Our World in Data [46]. This study provides updates of traditional surveillance data and dynamic panel estimates from the original study by Post et al [9,41,42,47,48]. For the region of Europe, the data comprised an unbalanced panel of 44 countries and territories, running from August 14, 2020, to May 12, 2023. Because a number of countries around the world switched from daily to weekly reports at various points in 2023, we used a cubic spline to interpolate daily new cases and deaths if any country had 4 consecutive periods of nonzero new cases interspersed by 6 days of 0 new cases.

To identify the appearance and duration of variants of concern, we also used data on sequenced SARS-CoV-2 variants from the Global Initiative on Sharing All Influenza Data (GISAID), which is an effective and trusted web-based resource for sharing genetic, clinical, and epidemiological COVID-19 data [49-52]. We used Nextclade nomenclature [53] to collect clade designations from sequences and Pangolin nomenclature for lineage designations of SARS-CoV-2 [54,55]. Metadata for the study period, which add geographic location to the clade designations, were collected on June 22, 2023. To avoid low-frequency or potentially erroneous samples, the data set was further filtered to exclude months with <100 available samples, variant groups with <5 samples in a month, and variant groups representing <0.5% of the total samples in a month. The final data set consisted of 184,386 total samples available on GISAID [49-52].

We analyzed the potential “statistical end” to the pandemic with a 1-tailed t-test for whether the mean of speed was equal to or greater than the outbreak threshold of 10. We ran the test on a rolling 6-month window over weekly speed for the region, and we plotted the P values from the test over time. All statistical analyses were conducted in R (version 4.2.1; R Foundation for Statistical Computing) with the plm package (version 2.6-2) [47,56].

Ethical Considerations

This study does not constitute research with human participants (as defined by 45CFR46.102) because all data are publicly available and contain no identifiable private information. The institutional review board’s review was therefore unsolicited.

Results

Dynamic Panel Estimates

Table 1 presents the dynamic panel estimates for the week of May 5, 2023.

<table>
<thead>
<tr>
<th>Region</th>
<th>t-statistic</th>
<th>Significance</th>
<th>Intercept</th>
<th>Speed</th>
<th>Acceleration</th>
<th>Deceleration</th>
<th>Jerk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Arellano-Bond dynamic panel data estimates for COVID-19 dynamics at the country level in Europe for the week of May 5, 2023.  

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-day lag coefficient</td>
<td>0.404</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>7-day lag coefficient</td>
<td>0.547</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Shift parameter week of April 28</td>
<td>0.032</td>
<td>.42</td>
</tr>
<tr>
<td>Shift parameter week of May 5</td>
<td>0.075</td>
<td>.48</td>
</tr>
<tr>
<td>Weekend effect</td>
<td>−0.272</td>
<td>.01</td>
</tr>
</tbody>
</table>

*Wald: χ²₆=6104; P<.001; Sargan: χ²₅₄₀=40; P>.99.*

While the 1-day and 7-day lag coefficients were positive and statistically significant (P<.001 for both), they were moderate in magnitude (0.404 and 0.547, respectively). For example, the 7-day coefficient suggests a cluster effect in which 1 case on a given day predicts 0.547 cases 1 week later. The shift parameters for the weeks of April 28, 2023, and May 5, 2023, were small and statistically insignificant (P=.42 and P=.48, respectively), however, suggesting the cluster effect of cases remained stable around the window of the WHO declaration.

The dynamic panel estimates have several advantages over the basic reproductive number, R₀, which estimates the average number of people a contagious person will infect [57]. Foremost, R₀ depends on many variables, such as social distancing, vaccination rates, demographics, and the transmissibility of a pathogen. Because the SARS-CoV-2 virus has mutated over time, so has its R₀, but rapidly updated estimates for R₀ are difficult to obtain. Vaccination campaigns and public health mitigation efforts have also evolved and thereby shaped R₀. The dynamic panel estimates are based on a recent, 120-day window, so they can quickly adjust to new circumstances. The Arellano-Bond model is also robust to time-invariant, unobservable factors (in the application, stable differences between countries); corrects for autocorrelation; and allows for statistical tests of various model parameters [42].

The Wald and Sargan tests can assess the validity of the dynamic panel model. The Wald hypothesis test checks whether the independent variables have explanatory power for the dependent variable. From Table 1, the Wald test was highly significant (P<.001), rejecting the null hypothesis of no explanatory power. The Sargan test checks the validity of the overidentifying restrictions of the model. A rejection of the null would be evidence against the validity, but the test failed to reject the null with P>.99.

**Statis Surveillance Metrics**

Static surveillance metrics for the weeks of April 28 and May 5, 2023, are provided in Tables 2 and 3.
Table 2. Static COVID-19 surveillance metrics for European countries for the week of April 28, 2023.

<table>
<thead>
<tr>
<th>Country</th>
<th>New COVID-19 cases, n</th>
<th>Cumulative COVID-19 cases, n</th>
<th>7-day moving average of new cases</th>
<th>Weekly transmission rate per 100,000 population</th>
<th>New weekly deaths</th>
<th>Cumulative deaths</th>
<th>7-day moving average of deaths</th>
<th>Death rate per 100,000 population</th>
<th>Condition death rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>0</td>
<td>334,090</td>
<td>0</td>
<td>0</td>
<td>3604</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Andorra</td>
<td>0</td>
<td>48,015</td>
<td>4.57</td>
<td>0</td>
<td>159</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>952</td>
<td>6,067,780</td>
<td>614.43</td>
<td>10.65</td>
<td>3</td>
<td>22,456</td>
<td>2.14</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>Belgium</td>
<td>317</td>
<td>4,795,866</td>
<td>256.43</td>
<td>2.72</td>
<td>5</td>
<td>34,280</td>
<td>4.57</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>14</td>
<td>402,906</td>
<td>5.86</td>
<td>0.43</td>
<td>1</td>
<td>16,338</td>
<td>0.14</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>155</td>
<td>1,305,030</td>
<td>132.57</td>
<td>2.29</td>
<td>3</td>
<td>38,339</td>
<td>1.86</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>Croatia</td>
<td>62</td>
<td>1,272,886</td>
<td>69.86</td>
<td>10.77</td>
<td>0</td>
<td>18,180</td>
<td>3.57</td>
<td>0.79</td>
<td>0.01</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>93</td>
<td>4,640,818</td>
<td>86.43</td>
<td>0.89</td>
<td>0</td>
<td>42,778</td>
<td>1.71</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Denmark</td>
<td>20</td>
<td>3,412,109</td>
<td>72</td>
<td>0.34</td>
<td>7</td>
<td>8572</td>
<td>5</td>
<td>0.12</td>
<td>0</td>
</tr>
<tr>
<td>Estonia</td>
<td>23</td>
<td>618,297</td>
<td>32.14</td>
<td>1.73</td>
<td>0</td>
<td>3001</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Finland</td>
<td>279</td>
<td>1,475,841</td>
<td>246</td>
<td>5.04</td>
<td>8</td>
<td>9612</td>
<td>10.57</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>France</td>
<td>6161</td>
<td>38,946,430</td>
<td>4,651.43</td>
<td>9.09</td>
<td>57</td>
<td>163,076</td>
<td>35</td>
<td>0.08</td>
<td>0</td>
</tr>
<tr>
<td>Germany</td>
<td>1496</td>
<td>38,409,945</td>
<td>1,058.29</td>
<td>1.79</td>
<td>9</td>
<td>174,086</td>
<td>16.86</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>Greece</td>
<td>1763</td>
<td>6,031,868</td>
<td>1,870.86</td>
<td>118.81</td>
<td>0</td>
<td>36,811</td>
<td>8.14</td>
<td>0.53</td>
<td>0.01</td>
</tr>
<tr>
<td>Hungary</td>
<td>56</td>
<td>2,201,824</td>
<td>64.43</td>
<td>3.92</td>
<td>0</td>
<td>48,778</td>
<td>1.14</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>Iceland</td>
<td>0</td>
<td>209,191</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>260</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ireland</td>
<td>183</td>
<td>1,710,808</td>
<td>60.29</td>
<td>3.64</td>
<td>3</td>
<td>8914</td>
<td>2.43</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>Italy</td>
<td>3580</td>
<td>25,809,208</td>
<td>2,974.43</td>
<td>6.06</td>
<td>31</td>
<td>189,904</td>
<td>23.71</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Latvia</td>
<td>11</td>
<td>977,734</td>
<td>17.43</td>
<td>0.59</td>
<td>0</td>
<td>6351</td>
<td>0.86</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>0</td>
<td>21,465</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lithuania</td>
<td>80</td>
<td>1,318,871</td>
<td>70.14</td>
<td>2.91</td>
<td>1</td>
<td>9671</td>
<td>1</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0</td>
<td>319,959</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1232</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Malta</td>
<td>9</td>
<td>118,524</td>
<td>12</td>
<td>11.94</td>
<td>0</td>
<td>835</td>
<td>0.19</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Moldova</td>
<td>17</td>
<td>620,333</td>
<td>21</td>
<td>3.73</td>
<td>0</td>
<td>12,113</td>
<td>0.14</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>Monaco</td>
<td>2</td>
<td>16,763</td>
<td>1.71</td>
<td>5.48</td>
<td>0</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Montenegro</td>
<td>24</td>
<td>291,702</td>
<td>26.71</td>
<td>3.83</td>
<td>0</td>
<td>2826</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0</td>
<td>8,610,372</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22,992</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Norway</td>
<td>77</td>
<td>1,483,939</td>
<td>54.29</td>
<td>1.42</td>
<td>0</td>
<td>5476</td>
<td>4.57</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Poland</td>
<td>251</td>
<td>6,514,536</td>
<td>136.29</td>
<td>0.63</td>
<td>2</td>
<td>119,559</td>
<td>1.14</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Portugal</td>
<td>238</td>
<td>5,582,561</td>
<td>193.57</td>
<td>2.32</td>
<td>9</td>
<td>26,628</td>
<td>5.29</td>
<td>0.09</td>
<td>0</td>
</tr>
<tr>
<td>Romania</td>
<td>504</td>
<td>3,393,902</td>
<td>522.29</td>
<td>17.93</td>
<td>0</td>
<td>68,089</td>
<td>6</td>
<td>0.16</td>
<td>0.02</td>
</tr>
<tr>
<td>San Marino</td>
<td>15</td>
<td>24,189</td>
<td>10.71</td>
<td>44.52</td>
<td>0</td>
<td>125</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Serbia</td>
<td>472</td>
<td>2,536,159</td>
<td>355.71</td>
<td>6.87</td>
<td>0</td>
<td>18,022</td>
<td>0.57</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Slovakia</td>
<td>27</td>
<td>1,866,470</td>
<td>14.43</td>
<td>0</td>
<td>0</td>
<td>21,167</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Slovenia</td>
<td>40</td>
<td>1,343,646</td>
<td>31.43</td>
<td>1.89</td>
<td>2</td>
<td>9336</td>
<td>3.14</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Spain</td>
<td>1692</td>
<td>13,845,803</td>
<td>1,461.14</td>
<td>24.91</td>
<td>184</td>
<td>120,917</td>
<td>26.29</td>
<td>0.39</td>
<td>0.01</td>
</tr>
<tr>
<td>Sweden</td>
<td>140</td>
<td>2,707,171</td>
<td>147.71</td>
<td>9.31</td>
<td>0</td>
<td>24,174</td>
<td>10.71</td>
<td>0.71</td>
<td>0.01</td>
</tr>
<tr>
<td>Switzerland</td>
<td>134</td>
<td>4,403,491</td>
<td>142.29</td>
<td>10.75</td>
<td>0</td>
<td>14,004</td>
<td>0.71</td>
<td>0.08</td>
<td>0</td>
</tr>
<tr>
<td>Ukraine</td>
<td>1150</td>
<td>5,538,357</td>
<td>1,258.57</td>
<td>20.27</td>
<td>0</td>
<td>112,210</td>
<td>11.71</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td>Country</td>
<td>New COVID-19 cases, n</td>
<td>Cumulative COVID-19 cases, n</td>
<td>7-day moving average of new cases</td>
<td>Weekly transmission rate per 100,000 population</td>
<td>New weekly deaths</td>
<td>Cumulative deaths</td>
<td>7-day moving average of deaths</td>
<td>Death rate per 100,000 population</td>
<td>Conditional death rate</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------</td>
<td>------------------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1670</td>
<td>24,594,965</td>
<td>1,615.57</td>
<td>2.47</td>
<td>0</td>
<td>226,278</td>
<td>14.43</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Country</td>
<td>New COVID-19 cases, n</td>
<td>Cumulative COVID-19 cases, n</td>
<td>7-day moving average of new cases</td>
<td>Weekly transmission rate per 100,000 population</td>
<td>New weekly deaths</td>
<td>Cumulative deaths</td>
<td>7-day moving average of deaths</td>
<td>Death rate per 100,000 population</td>
<td>Conditional death rate</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>-----------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Albania</td>
<td>0</td>
<td>334,090</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3604</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Andorra</td>
<td>0</td>
<td>48,015</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>159</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Austria</td>
<td>436</td>
<td>6,071,124</td>
<td>477.71</td>
<td>4.88</td>
<td>2</td>
<td>22,479</td>
<td>3.29</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>Belgium</td>
<td>184</td>
<td>4,797,400</td>
<td>219.14</td>
<td>1.58</td>
<td>1</td>
<td>34,300</td>
<td>2.86</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>0</td>
<td>402,924</td>
<td>2.57</td>
<td>0</td>
<td>0</td>
<td>16,340</td>
<td>0.29</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>86</td>
<td>1,305,755</td>
<td>103.57</td>
<td>1.27</td>
<td>1</td>
<td>38,351</td>
<td>1.71</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Croatia</td>
<td>43</td>
<td>1,273,256</td>
<td>50.43</td>
<td>7.39</td>
<td>0</td>
<td>18,213</td>
<td>4.71</td>
<td>0.61</td>
<td>0.01</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>74</td>
<td>4,641,292</td>
<td>67.71</td>
<td>0.71</td>
<td>2</td>
<td>42,788</td>
<td>1.43</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Denmark</td>
<td>109</td>
<td>3,412,750</td>
<td>91.57</td>
<td>1.85</td>
<td>5</td>
<td>8619</td>
<td>6.71</td>
<td>0.09</td>
<td>0</td>
</tr>
<tr>
<td>Estonia</td>
<td>21</td>
<td>618,492</td>
<td>27.86</td>
<td>1.58</td>
<td>0</td>
<td>3001</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Finland</td>
<td>205</td>
<td>1,477,339</td>
<td>214</td>
<td>3.70</td>
<td>6</td>
<td>9659</td>
<td>6.71</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>France</td>
<td>4657</td>
<td>38,973,507</td>
<td>3,868.14</td>
<td>6.87</td>
<td>28</td>
<td>163,242</td>
<td>23.71</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>Germany</td>
<td>1056</td>
<td>38,417,047</td>
<td>1,014.57</td>
<td>1.27</td>
<td>13</td>
<td>174,169</td>
<td>11.86</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>Greece</td>
<td>2048</td>
<td>6,044,517</td>
<td>1,878.29</td>
<td>138.05</td>
<td>0</td>
<td>36,863</td>
<td>7.43</td>
<td>0.50</td>
<td>0.01</td>
</tr>
<tr>
<td>Hungary</td>
<td>47</td>
<td>2,202,179</td>
<td>50.14</td>
<td>3.29</td>
<td>0</td>
<td>48,782</td>
<td>0.57</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Ireland</td>
<td>54</td>
<td>1,711,287</td>
<td>68.43</td>
<td>1.08</td>
<td>0</td>
<td>8917</td>
<td>0.43</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Italy</td>
<td>2592</td>
<td>25,828,252</td>
<td>2,720.57</td>
<td>4.39</td>
<td>24</td>
<td>190,080</td>
<td>25.14</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Latvia</td>
<td>9</td>
<td>977,805</td>
<td>10.14</td>
<td>0.49</td>
<td>1</td>
<td>6358</td>
<td>1</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>1</td>
<td>21,468</td>
<td>0.43</td>
<td>0</td>
<td>0</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lithuania</td>
<td>73</td>
<td>1,319,395</td>
<td>74.86</td>
<td>2.65</td>
<td>0</td>
<td>9672</td>
<td>0.14</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0</td>
<td>319,959</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1232</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Malta</td>
<td>7</td>
<td>118,574</td>
<td>7.29</td>
<td>9.71</td>
<td>0</td>
<td>835</td>
<td>0</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>Moldova</td>
<td>12</td>
<td>620,443</td>
<td>14.29</td>
<td>2.58</td>
<td>0</td>
<td>12,115</td>
<td>0.29</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Monaco</td>
<td>2</td>
<td>16,773</td>
<td>1.43</td>
<td>5.48</td>
<td>0</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Montenegro</td>
<td>5</td>
<td>291,817</td>
<td>16.43</td>
<td>0.80</td>
<td>0</td>
<td>2827</td>
<td>0.14</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>0</td>
<td>8,610,372</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22,992</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Norway</td>
<td>59</td>
<td>1,484,287</td>
<td>49.71</td>
<td>1.09</td>
<td>0</td>
<td>5497</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Poland</td>
<td>113</td>
<td>6,515,449</td>
<td>130.43</td>
<td>0.28</td>
<td>6</td>
<td>119,581</td>
<td>3.14</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Portugal</td>
<td>231</td>
<td>5,583,979</td>
<td>202.57</td>
<td>2.25</td>
<td>3</td>
<td>26,671</td>
<td>6.14</td>
<td>0.03</td>
<td>0</td>
</tr>
<tr>
<td>Romania</td>
<td>408</td>
<td>3,397,365</td>
<td>461.43</td>
<td>14.52</td>
<td>0</td>
<td>68,117</td>
<td>4</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>San Marino</td>
<td>5</td>
<td>24,241</td>
<td>7.43</td>
<td>14.84</td>
<td>0</td>
<td>125</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Serbia</td>
<td>272</td>
<td>2,538,142</td>
<td>283.29</td>
<td>3.96</td>
<td>1</td>
<td>18,033</td>
<td>1.57</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Slovakia</td>
<td>21</td>
<td>1,866,647</td>
<td>24.57</td>
<td>0</td>
<td>0</td>
<td>21,167</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Slovenia</td>
<td>27</td>
<td>1,343,831</td>
<td>26.43</td>
<td>1.27</td>
<td>1</td>
<td>9356</td>
<td>2.86</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Spain</td>
<td>2502</td>
<td>13,845,803</td>
<td>2,143.29</td>
<td>36.82</td>
<td>0</td>
<td>120,917</td>
<td>0</td>
<td>0.25</td>
<td>0.01</td>
</tr>
<tr>
<td>Sweden</td>
<td>133</td>
<td>2,708,118</td>
<td>134.86</td>
<td>8.83</td>
<td>0</td>
<td>24,247</td>
<td>10.43</td>
<td>0.64</td>
<td>0.01</td>
</tr>
<tr>
<td>Switzerland</td>
<td>99</td>
<td>4,404,340</td>
<td>115.43</td>
<td>7.94</td>
<td>0</td>
<td>14,012</td>
<td>1.14</td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td>Ukraine</td>
<td>814</td>
<td>5,544,969</td>
<td>950.29</td>
<td>14.35</td>
<td>0</td>
<td>112,271</td>
<td>8.71</td>
<td>0.13</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Most countries had a transmission rate considered low by the US Centers for Disease Control and Prevention [9,33-43,58]. Specifically, a “low” transmission is considered to be no more than 10 cases per 100,000 people per week. “Moderate” transmission is 10 to 50 cases per 100,000 people per week. “Substantial” transmission is 50 to 100 cases [58,59]. However, a number of countries were in a state of outbreak. In particular, Greece had a speed of 119 new cases per 100,000 population in the week of April 28, and San Marino had a speed of 45. Greece was therefore in a “substantial” outbreak, while San Marino was in a “moderate” outbreak. Austria, Croatia, Malta, Romania, Spain, Switzerland, and Ukraine were also in “moderate” outbreaks.

By the following week, only Greece, Romania, San Marino, Spain, and Ukraine remained in an outbreak. While speed dropped notably for San Marino, speed alarmingly increased for Greece and Spain. We noted that speed in island nations often vacillates between high and low rates of transmission.

Overall, the status of the pandemic around the WHO declaration in Europe is consistent with an “end” to the pandemic, but the distinction is muddied by continued outbreaks in several countries. The outbreaks in Greece and San Marino comprise a small portion of the overall European population, and speed is more variable in island nations. On the basis of the definition of a pandemic or an outbreak in several countries, the data suggest a shift from pandemic to endemic COVID-19, but the continued outbreaks in 3 other countries do caution that the conclusion may be premature.

Comparing Tables 2 and 3 demonstrates a drop in transmission rates before and after the WHO declared an end to COVID-19 as a public health emergency. Overall, the United Kingdom, Italy, and Germany had the most cases of COVID-19 transmissions and deaths, but these ranks are largely a function of population size. Thus, a better measure of COVID-19 fatality risk is the number of COVID-19 cases and deaths per 100,000 people. Moreover, death is often a better proxy for the state of an outbreak than transmissions because deaths are less likely to be undercounted [60]. Undercounting may be due to poor public health infrastructure, home antigen testing, or a dearth of polymerase chain reaction testing or other resources. When we control for the risk of death given the number of COVID-19 transmissions, we find that Eastern European countries, such as Bosnia and Herzegovina, Moldova, and Poland, had the highest conditional death rates. For example, Bosnia and Herzegovina had the highest conditional death rate of 0.04 deaths per confirmed case. These disparities could be driven by differences in public health infrastructure, demographics, and the influx of refugees from the Russian invasion of Ukraine.

Enhanced Dynamic Surveillance Metrics
Tables 4 and 5 contain enhanced dynamic surveillance metrics for the weeks before and after May 5.

We note that the figures in Tables 4 and 5 are not calculated as day-over-day averages across the week, as they are in Tables 2 and 3. Thus, the magnitudes of speed differ slightly across the tables. Again, by the week of May 5, speed was low for every country except Greece, Romania, San Marino, Spain, and Ukraine. The 7-day persistence effect on speed was also relatively high for these countries but low for others. Acceleration was almost uniformly negative, with the exception of Greece and Spain, which saw their outbreaks grow somewhat from the first week to the second. Across the board, jerk tended to be very small in magnitude, suggesting little change in acceleration rates.

Table 6 compares the 7-day persistence effect on speed for the top 5 countries around the 2 weeks of the WHO declaration. These ranks largely reflect the speed in the countries with outbreaks in the prior tables.
Table 4. Novel surveillance metrics for European countries for the week of April 28, 2023.

<table>
<thead>
<tr>
<th>Country</th>
<th>Speed$^a$</th>
<th>Acceleration$^b$</th>
<th>Jerk$^c$</th>
<th>7-day persistence effect on speed$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Andorra</td>
<td>7.34</td>
<td>0</td>
<td>−2.15</td>
<td>2.81</td>
</tr>
<tr>
<td>Austria</td>
<td>6.87</td>
<td>0.42</td>
<td>1.22</td>
<td>5.08</td>
</tr>
<tr>
<td>Belgium</td>
<td>2.20</td>
<td>−0.12</td>
<td>−0.03</td>
<td>1.79</td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>0.18</td>
<td>0</td>
<td>0.06</td>
<td>0.19</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>1.95</td>
<td>0.08</td>
<td>0.05</td>
<td>1.21</td>
</tr>
<tr>
<td>Croatia</td>
<td>12.12</td>
<td>−0.37</td>
<td>−0.05</td>
<td>7.44</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.82</td>
<td>−0.07</td>
<td>−0.07</td>
<td>0.76</td>
</tr>
<tr>
<td>Denmark</td>
<td>1.22</td>
<td>−0.17</td>
<td>−0.17</td>
<td>0.84</td>
</tr>
<tr>
<td>Estonia</td>
<td>2.42</td>
<td>−0.02</td>
<td>0.11</td>
<td>1.70</td>
</tr>
<tr>
<td>Finland</td>
<td>4.44</td>
<td>−0.11</td>
<td>−0.15</td>
<td>2.73</td>
</tr>
<tr>
<td>France</td>
<td>6.86</td>
<td>−0.09</td>
<td>−0.35</td>
<td>4.92</td>
</tr>
<tr>
<td>Germany</td>
<td>1.27</td>
<td>0.02</td>
<td>0.01</td>
<td>0.88</td>
</tr>
<tr>
<td>Greece</td>
<td>126.11</td>
<td>−3.09</td>
<td>0.26</td>
<td>75.57</td>
</tr>
<tr>
<td>Hungary</td>
<td>4.52</td>
<td>−0.22</td>
<td>0.01</td>
<td>3.23</td>
</tr>
<tr>
<td>Iceland</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ireland</td>
<td>1.20</td>
<td>0.28</td>
<td>0.73</td>
<td>0.84</td>
</tr>
<tr>
<td>Italy</td>
<td>5.04</td>
<td>−0.48</td>
<td>−1.35</td>
<td>3.03</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.94</td>
<td>−0.06</td>
<td>0</td>
<td>0.55</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>0.36</td>
<td>0</td>
<td>0</td>
<td>0.39</td>
</tr>
<tr>
<td>Lithuania</td>
<td>2.55</td>
<td>0</td>
<td>0.04</td>
<td>1.76</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Malta</td>
<td>15.67</td>
<td>−1.28</td>
<td>−0.01</td>
<td>12.24</td>
</tr>
<tr>
<td>Moldova</td>
<td>4.49</td>
<td>−0.30</td>
<td>0.03</td>
<td>4.01</td>
</tr>
<tr>
<td>Monaco</td>
<td>4.70</td>
<td>0.78</td>
<td>−0.78</td>
<td>0.85</td>
</tr>
<tr>
<td>Montenegro</td>
<td>4.26</td>
<td>0</td>
<td>0.21</td>
<td>2.97</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Norway</td>
<td>1</td>
<td>0.03</td>
<td>0.04</td>
<td>0.67</td>
</tr>
<tr>
<td>Poland</td>
<td>0.34</td>
<td>−0.01</td>
<td>0.09</td>
<td>0.41</td>
</tr>
<tr>
<td>Portugal</td>
<td>1.88</td>
<td>−0.05</td>
<td>−0.32</td>
<td>0.99</td>
</tr>
<tr>
<td>Romania</td>
<td>18.59</td>
<td>−0.48</td>
<td>0.17</td>
<td>13.01</td>
</tr>
<tr>
<td>San Marino</td>
<td>31.80</td>
<td>2.97</td>
<td>5.94</td>
<td>23.16</td>
</tr>
<tr>
<td>Serbia</td>
<td>5.18</td>
<td>0.03</td>
<td>0.02</td>
<td>3.57</td>
</tr>
<tr>
<td>Slovenia</td>
<td>0.40</td>
<td>0</td>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>Slovakia</td>
<td>1.48</td>
<td>0.08</td>
<td>0.05</td>
<td>1.28</td>
</tr>
<tr>
<td>Spain</td>
<td>21.51</td>
<td>0.88</td>
<td>0.20</td>
<td>10.81</td>
</tr>
<tr>
<td>Sweden</td>
<td>9.80</td>
<td>−0.17</td>
<td>0</td>
<td>5.91</td>
</tr>
<tr>
<td>Switzerland</td>
<td>11.39</td>
<td>−0.17</td>
<td>−0.03</td>
<td>6.54</td>
</tr>
<tr>
<td>Ukraine</td>
<td>22.19</td>
<td>−0.56</td>
<td>−0.01</td>
<td>15.18</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>2.39</td>
<td>−0.02</td>
<td>−0.05</td>
<td>1.32</td>
</tr>
</tbody>
</table>

$^a$ New COVID-19 cases per 100,000 people.

$^b$ The difference in speed from one week to the next.
cThe change in acceleration from one week to the next.

dThe impact of the 7-day lag of speed on current speed (the echo-forward effect of COVID-19 cases on future cases 7 days later).
<table>
<thead>
<tr>
<th>Country</th>
<th>Speed(^a)</th>
<th>Acceleration(^b)</th>
<th>Jerk(^c)</th>
<th>7-day persistence effect on speed(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Andorra</td>
<td>0</td>
<td>0</td>
<td>7.34</td>
<td>4.25</td>
</tr>
<tr>
<td>Austria</td>
<td>5.34</td>
<td>−0.82</td>
<td>−0.70</td>
<td>3.98</td>
</tr>
<tr>
<td>Belgium</td>
<td>1.88</td>
<td>−0.16</td>
<td>0.01</td>
<td>1.27</td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>0.08</td>
<td>−0.06</td>
<td>−0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>1.53</td>
<td>−0.15</td>
<td>−0.08</td>
<td>1.13</td>
</tr>
<tr>
<td>Croatia</td>
<td>8.75</td>
<td>−0.48</td>
<td>0.02</td>
<td>7.02</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.65</td>
<td>−0.03</td>
<td>0.01</td>
<td>0.48</td>
</tr>
<tr>
<td>Denmark</td>
<td>1.56</td>
<td>0.22</td>
<td>0.17</td>
<td>0.71</td>
</tr>
<tr>
<td>Estonia</td>
<td>2.10</td>
<td>−0.02</td>
<td>0.09</td>
<td>1.40</td>
</tr>
<tr>
<td>Finland</td>
<td>3.86</td>
<td>−0.19</td>
<td>0.16</td>
<td>2.57</td>
</tr>
<tr>
<td>France</td>
<td>5.70</td>
<td>−0.32</td>
<td>0.11</td>
<td>3.97</td>
</tr>
<tr>
<td>Germany</td>
<td>1.22</td>
<td>−0.08</td>
<td>0.01</td>
<td>0.74</td>
</tr>
<tr>
<td>Greece</td>
<td>126.61</td>
<td>2.75</td>
<td>0.78</td>
<td>73.03</td>
</tr>
<tr>
<td>Hungary</td>
<td>3.50</td>
<td>−0.09</td>
<td>0.02</td>
<td>2.62</td>
</tr>
<tr>
<td>Iceland</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ireland</td>
<td>1.36</td>
<td>−0.37</td>
<td>−0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>Italy</td>
<td>4.61</td>
<td>−0.24</td>
<td>0.29</td>
<td>2.92</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.55</td>
<td>−0.02</td>
<td>0.12</td>
<td>0.55</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>1.09</td>
<td>0</td>
<td>0</td>
<td>0.21</td>
</tr>
<tr>
<td>Lithuania</td>
<td>2.72</td>
<td>−0.04</td>
<td>0.09</td>
<td>1.48</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Malta</td>
<td>9.82</td>
<td>−0.32</td>
<td>0.19</td>
<td>9.08</td>
</tr>
<tr>
<td>Moldova</td>
<td>3.04</td>
<td>−0.16</td>
<td>0.01</td>
<td>2.60</td>
</tr>
<tr>
<td>Monaco</td>
<td>3.91</td>
<td>0</td>
<td>0.78</td>
<td>2.72</td>
</tr>
<tr>
<td>Montenegro</td>
<td>2.62</td>
<td>−0.43</td>
<td>−0.43</td>
<td>2.47</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Norway</td>
<td>0.91</td>
<td>−0.05</td>
<td>−0.01</td>
<td>0.58</td>
</tr>
<tr>
<td>Poland</td>
<td>0.33</td>
<td>−0.05</td>
<td>−0.08</td>
<td>0.20</td>
</tr>
<tr>
<td>Portugal</td>
<td>1.97</td>
<td>−0.01</td>
<td>0.09</td>
<td>1.09</td>
</tr>
<tr>
<td>Romania</td>
<td>16.43</td>
<td>−0.49</td>
<td>−0.10</td>
<td>10.77</td>
</tr>
<tr>
<td>San Marino</td>
<td>22.05</td>
<td>−4.24</td>
<td>−3.39</td>
<td>18.42</td>
</tr>
<tr>
<td>Serbia</td>
<td>4.12</td>
<td>−0.42</td>
<td>−0.05</td>
<td>3</td>
</tr>
<tr>
<td>Slovakia</td>
<td>0.45</td>
<td>0</td>
<td>0</td>
<td>0.23</td>
</tr>
<tr>
<td>Slovenia</td>
<td>1.25</td>
<td>−0.09</td>
<td>0.14</td>
<td>0.86</td>
</tr>
<tr>
<td>Spain</td>
<td>31.54</td>
<td>1.70</td>
<td>0.05</td>
<td>12.46</td>
</tr>
<tr>
<td>Sweden</td>
<td>8.95</td>
<td>−0.07</td>
<td>0.02</td>
<td>5.67</td>
</tr>
<tr>
<td>Switzerland</td>
<td>9.25</td>
<td>−0.40</td>
<td>−0.03</td>
<td>6.60</td>
</tr>
<tr>
<td>Ukraine</td>
<td>16.75</td>
<td>−0.84</td>
<td>0.02</td>
<td>12.85</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>2.02</td>
<td>−0.06</td>
<td>0.03</td>
<td>1.39</td>
</tr>
</tbody>
</table>

\(^a\) New COVID-19 cases per 100,000 people.

\(^b\) The difference in speed from one week to the next.
The change in acceleration from one week to the next.

The impact of the 7-day lag of speed on current speed (the echo-forward effect of COVID-19 cases on future cases 7 days later).

Table 6. The European countries with the highest 7-day persistence estimate in the weeks of April 28 and May 5, 2023.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>7-day persistence week 1 (April 28)</th>
<th>7-day persistence week 2 (May 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Greece</td>
<td>75.57</td>
<td>73.03</td>
</tr>
<tr>
<td>2</td>
<td>San Marino</td>
<td>23.16</td>
<td>18.42</td>
</tr>
<tr>
<td>3</td>
<td>Ukraine</td>
<td>15.18</td>
<td>12.85</td>
</tr>
<tr>
<td>4</td>
<td>Romania</td>
<td>13.01</td>
<td>10.77</td>
</tr>
<tr>
<td>4</td>
<td>Spain</td>
<td>10.81</td>
<td>12.46</td>
</tr>
<tr>
<td>5</td>
<td>Malta</td>
<td>12.24</td>
<td>9.08</td>
</tr>
</tbody>
</table>

These metrics suggest that the pandemic may have ended in the region. Still, speed and persistence measures were moderate to high for several countries in outbreaks, and the region had not exited the pandemic with as much clarity as had several other global regions, such as South Asia or sub-Saharan Africa.

Figure 1 plots regional speed, acceleration, jerk, and 7-day persistence metrics from August 14, 2020, to May 12, 2023.

Figure 1. Novel surveillance metrics (speed, acceleration, jerk, and 7-day persistence) for COVID-19 transmissions in Europe from August 2020 to May 2023.

The region did see a slight bump in cases around the end of February 2023. Still, the bump did not cause the region to cross the outbreak threshold.

Europe saw one especially pronounced outbreak over the course of the pandemic. The outbreak caused speed to reach a peak of 246 novel COVID-19 cases per 100,000 population in the last week of January 2022. Figure 2 plots variant groups as a proportion of all viral specimens collected and sequenced in the region (and made available through GISAID) each month. The outbreak occurred just after the arrival of the Omicron variant. Europe, like much of the rest of the world, saw a surge in cases amid the heightened transmissibility of Omicron [45]. Earlier outbreaks were driven by the ancestral, Alpha, and Delta variants.
Another potential indication of the end of the pandemic is the continued dominance of the Omicron variant. Subclades of Omicron continue to circulate 4.5 months after WHO declared the end of the COVID-19 emergency [61]. Viral sequences have almost exclusively returned as Omicron and its subvariants ever since its arrival.

Figure 3 plots $P$ values from a series of 1-tailed $t$ tests of whether speed for the region was equal to or greater than the threshold outbreak of 10. These tests were conducted over a rolling 6-month window of weekly regional speed. The dashed gray line denotes the least restrictive conventional significance level threshold of $\alpha=0.10$. The test strongly rejected the null in favor of the alternative until the very end of April 2023. While this more recent lack of statistical significance is consistent with the end of the pandemic in the region, its relative recency around the WHO declaration suggests prematurity in the conclusion that COVID-19 had transitioned from the pandemic to endemic phase in Europe.

**Figure 2.** Variants of concern (VOCs) as a proportion of all sequenced SARS-CoV-2 specimens from April 2020 to May 2023 in Europe.

**Figure 3.** $P$ values from $t$ tests of weekly COVID-19 transmissions per 100,000 people equal to 10 over a rolling, 6-month window in Europe.
Figure 4 provides a timeline of the onset of COVID-19 in Europe as well as vaccination programs and major events that likely shaped the course of disease control, such as the Next Generation European Union (NGEU) economic recovery package and the Russian invasion of Ukraine. Millions of refugees fled Ukraine, accelerating the spread of disease in the region. Mass human migration is affiliated with increased disease transmission [62].

Figure 4. Timeline of the COVID-19 pandemic in Europe. EU: European Union; VOC: variants of concern; WHO: World Health Organization.
Discussion

**Principal Findings**

Multiple metrics suggest the pandemic had largely subsided in Europe by the time of the WHO declaration. Regional speed acceleration and jerk were both low and stable. The 1-day and 7-day persistence coefficients did remain statistically significant ($P<.001$ for both), but the coefficients were moderate in magnitude, and the shift parameters for the weeks around the WHO declaration were insignificant. This statistical insignificance implies no change in the clustering effect of cases on future cases at the time. The rolling $t$ test of speed equal to 10 became insignificant for the first time in the month ahead of the declaration. Finally, Omicron had been the predominant variant of concern in sequenced viral samples for approximately 6 months.

Still, among the 44 countries in Europe, 5 had surpassed the outbreak threshold at the time of the WHO declaration. While the region as a whole did not breach the outbreak threshold, 5 countries were classified as having epidemic status, with 4 being categorized as having mild epidemics. Our analysis of transmission rates, outbreak testing, and statistical findings suggests that the pandemic is likely concluding in Europe. Nevertheless, exercising caution, we refrain from definitively declaring the end of the pandemic at the time of the WHO declaration.

**History of Policy and Disease Burden**

For a brief history of policy and disease burden, differences in health outcomes across countries remained relatively minor despite heterogeneity in European COVID-19 mitigation efforts in terms of school closures, business restrictions, quarantines, social distancing, and mask mandates [63]. For instance, Denmark implemented the most stringent measures, while Sweden pursued a comparatively relaxed approach [63,64]. The environment and climate factors also factored into differential transmission and death rates [64,65]. Notably, widespread vaccination efforts proved effective in reducing the severity of COVID-19 cases and lowering mortality rates [66].

Age structure and environmental risk factors such as pollution affected the COVID-19 pandemic at its onset [67]. Italy reported their first cases in March 2020 [68-70]. Hospitals and cities lacked preparedness, with the initial wave claiming 35,000 lives [68]. Similar challenges were observed in France, which struggled despite having robust health care resources [7,68,71-74]. Around 6 months into the pandemic, France had one of the highest prevalence rates of the virus in Europe, with >2 million cumulative confirmed cases at the time [73]. Despite benefiting from a universal health insurance system, centralized presidential regime with a strong public administration, and a relatively high number of health care professionals and hospital beds compared to other European countries, France was still underprepared for the pandemic’s onset, with the health care system soon overwhelmed by cases and fatalities [73,75]. During the peak months in 2020, Spain, Belgium, and Ireland had the highest incidence and mortality rates [76]. European health care systems improved as vaccines became available [77]. Despite difficulties, Europe’s health care system benefited from risk recognition, treatment availability, and increased vaccine access [78-81]. Throughout the pandemic, new variants and fluctuating social protection measures have led to additional waves of infection [82,83]. Since this time, case rates have fallen despite mild rebound periods, with the COVID-19 social and financial recovery in Europe benefiting from rapidly decreased social restrictions and unprecedented monetary investments [84].

**Policies Implemented to Control and Mitigate the Transmission of COVID-19**

When the WHO declared COVID-19 a pandemic [85], Europe was the epicenter [86]. The European Union (EU) responded by closing external borders [7,87], but public reactions to social restrictions were volatile, leading to protests [88,89]. Despite this, Europe adapted to new regulations and accelerated its digital transformation [90,91].

The pandemic caused a 4% economic decline in Europe in 2020, with wide variation between countries [92]. To aid response, the EU allocated funds, including €140 million (US $159.6 million) in emergency funding [93]. Member states received continuous support through programs such as European instrument for temporary Support to mitigate Unemployment Risks in an Emergency [94,95]. In addition, in July 2020, the European Council, the European Commission’s sibling executive arm in the EU, agreed to an unprecedented €750 billion (US $855 billion) recovery package titled NGEU to be disseminated to member states through the end of 2023 [96]. NGEU, the largest stimulus package in European history, included special investments in future health care preparedness and modernization [97].

The EU’s Health Security Committee held its first meeting regarding COVID-19 on January 17, 2020, with the first mobilization of COVID-19 research funds taking place on January 31, 2020 [94]. Early efforts were made in conjunction with member states to secure medical equipment and protective gear for health care workers and ensure the free movement of critical workers to and around the EU [94]. Lockdowns and travel restrictions put in place by mid-March 2020, which is approximately when Italy’s total death toll overtook China during the first wave of the pandemic, affected >250 million Europeans [98]. As the pandemic intensified in the spring and summer of 2020, the European Commission partnered with Global Citizen and the European Investment Bank to develop tests, treatments, and vaccines with >€10 billion (US $11.4 billion) in new funds [94]. The most stringent virus mitigation orders in the United Kingdom came from Prime Minister Boris Johnson and the House of Commons in late March 2020, which granted the prime minister emergency powers to enforce widespread lockdowns and travel restrictions [99].

**Vaccination Campaigns**

Europe’s vaccination strategy featured early collective bargaining, resulting in 2.8 billion vaccine doses secured for member states through an advance purchase agreement in June 2020 [93]. The United Kingdom administered its first vaccine dose on December 8, 2020 [100], and by August 2021, the EU had fully vaccinated 70% of its adult population, exporting >1
billion vaccine doses to low- and middle-income countries by October 2021 [94]. However, some Eastern European countries, such as Hungary, Poland, Romania, and Bulgaria, lagged behind with <65% of their populations fully vaccinated by February 2022 [101].

In total, 4 vaccines received emergency use authorization and were widely used in Europe throughout the pandemic: BioNTech or Pfizer, Moderna, Oxford or AstraZeneca, and Johnson & Johnson [102-104]. Vaccine development, approval, and rollout occurred in rapid fashion due to collective efforts by regulatory agencies, private pharmaceutical companies, public universities, and public health agencies [102].

Humanitarian Crises
During the Omicron surge in early 2022, Russia’s attack on Ukraine displaced two-thirds of Ukrainians, leading to surges in Omicron and other infectious diseases, including tuberculosis. This conflict hindered COVID-19 vaccine distribution and access to essential medical services, such as HIV treatment [105-107]. Ukrainians faced obstacles in migrating without vaccination proof, as only 36% were vaccinated when the war began [108-113].

Europe absorbed many refugees during the COVID-19 pandemic, including those fleeing political unrest and violence in the Middle East and Eastern Europe [114,115]. In addition to overcrowded and unsafe living conditions and limited access to financial or medical support, many migrants were unable to access timely COVID-19 vaccination [114,115]. Some countries, like Spain, prioritized migrant vaccination, while Portugal granted temporary residence to ensure equal access [115].

Limitations
The COVID-19 data had become less frequently reported around the world by the time the WHO declared an end to the pandemic public health emergency [116]. In addition, more people began to use at-home tests as the pandemic evolved [117], and the Russian invasion of Ukraine damaged public health infrastructure, which may have reduced the accuracy of reported cases in the region. Because the enhanced surveillance metrics of speed, acceleration, jerk, and 7-day persistence are based on rates, not total counts, statistical bias caused by countries dropping in or out of the sample is mitigated, but to the extent that a nonincluded country is unrepresentative of the region in disease burden; the omission of a country or territory can still influence historical data comparisons. Viral specimen tests for variants of concern in GISAID are also dependent on testing and sequencing capacity, which varied by country across the region.

Conclusions
While there is significant evidence indicating that the pandemic in Europe has transitioned to an endemic phase, the persistent risk of new COVID-19 variants underscores the need for vigilance, robust vaccination campaigns, and international cooperation to effectively curb the spread of coronavirus in the region [40]. As the data on transmissions become less frequent [117] and as pandemic fatigue grows [118], the challenge of vigilance also evolves. The public health lessons from European policy and disease burden can inform not only the continued challenge but also responses to inevitable future pandemics.

Acknowledgments
The research reported in this publication was generously supported by the Robert J. Havey, MD Institute for Global Health’s Global Innovation Challenge Award at Northwestern University, Feinberg School of Medicine. This study was also supported by Feed the Future through the US Agency for International Development, under the terms of contract 7200LA1800003 and Feed the Future Innovation Lab for Collaborative Research on Sustainable Intensification (grant AID-OAA-L-14-00006). The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the US Agency for International Development or the authors’ organizations.

The authors gratefully acknowledge all data contributors, that is, the authors and their originating laboratories responsible for obtaining the specimens and their submitting laboratories for generating the genetic sequence and metadata and sharing via the Global Initiative on Sharing All Influenza Data initiative on which this research is based.

Data Availability
The data sets analyzed during this study are available in the Our World in Data repository [46] and the Global Initiative on Sharing All Influenza Data repository [49].

Conflicts of Interest
None declared.

References


Abbreviations

EU: European Union
GISAID: Global Initiative on Sharing All Influenza Data
NGEU: Next Generation European Union
WHO: World Health Organization
Discovering Subgroups of Children With High Mortality in Urban Guinea-Bissau: Exploratory and Validation Cohort Study

Andreas Rieckmann1, PhD; Sebastian Nielsen2,3, PhD; Piotr Dworzynski4, PhD; Heresh Amini5,6, PhD; Søren Wengel Mogensen7, PhD; Isaquel Bartolomeu Silva2,3, MSc; Angela Y Chang8,9, ScD; Onyebuchi A Arah10,11,12, MD, PhD; Wojciech Samek13,14,15, Dr rer nat; Naja Hulvej Rod1, PhD, DMSc; Claus Thorn Ekstrøm16, PhD; Christine Stabell Benn2,3,8, MD, PhD, DMSc; Peter Aaby2,3, DMSc; Ane Berent Fisker2,3, MD, PhD

1Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
2Bandim Health Project, INDEPTH Network, Bissau, Guinea-Bissau
3Bandim Health Project, Research unit Odense Patient Data Explorative Network (OPEN), Department of Clinical Research, Odense University Hospital and University of Southern Denmark, Odense, Denmark
4Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark
5Department of Environmental Medicine and Climate Science, Icahn School of Medicine at Mount Sinai, New York, NY, United States
6Institute for Climate Change, Environmental Health, and Exposomics, Icahn School of Medicine at Mount Sinai, New York, NY, United States
7Department of Automatic Control, Lund University, Lund, Sweden
8Danish Institute for Advanced Study, University of Southern Denmark, Odense, Denmark
9The Interdisciplinary Centre on Population Dynamics, University of Southern Denmark, Odense, Denmark
10Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States
11Department of Statistics and Data Science, College of Letters and Science, University of California, Los Angeles, Los Angeles, CA, United States
12Research Unit for Epidemiology, Department of Public Health, University of Aarhus, Aarhus, Denmark
13Department of Artificial Intelligence, Fraunhofer Heinrich Hertz Institute, Berlin, Germany
14Department of Electrical Engineering and Computer Science, Technical University of Berlin, Berlin, Germany
15Berlin Institute for the Foundations of Learning and Data, Berlin, Germany
16Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Corresponding Author: Andreas Rieckmann, PhD
Section of Epidemiology
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
Copenhagen, 1353
Denmark
Phone: 45 35326765
Email: aric@sund.ku.dk

Abstract

Background: The decline in global child mortality is an important public health achievement, yet child mortality remains disproportionately high in many low-income countries like Guinea-Bissau. The persisting high mortality rates necessitate targeted research to identify vulnerable subgroups of children and formulate effective interventions.

Objective: This study aimed to discover subgroups of children at an elevated risk of mortality in the urban setting of Bissau, Guinea-Bissau, West Africa. By identifying these groups, we intend to provide a foundation for developing targeted health interventions and inform public health policy.

Methods: We used data from the health and demographic surveillance site, Bandim Health Project, covering 2003 to 2019. We identified baseline variables recorded before children reached the age of 6 weeks. The focus was on determining factors consistently linked with increased mortality up to the age of 3 years. Our multifaceted methodological approach incorporated spatial analysis for visualizing geographical variations in mortality risk, causally adjusted regression analysis to single out specific risk factors, and machine learning techniques for identifying clusters of multifactorial risk factors. To ensure robustness and validity, we
divided the data set temporally, assessing the persistence of identified subgroups over different periods. The reassessment of mortality risk used the targeted maximum likelihood estimation (TMLE) method to achieve more robust causal modeling.

**Results:** We analyzed data from 21,005 children. The mortality risk (6 weeks to 3 years of age) was 5.2% (95% CI 4.8%-5.6%) for children born between 2003 and 2011, and 2.9% (95% CI 2.5%-3.3%) for children born between 2012 and 2016. Our findings revealed 3 distinct high-risk subgroups with notably higher mortality rates, children residing in a specific urban area (adjusted mortality risk difference of 3.4%, 95% CI 0.3%-6.5%), children born to mothers with no prenatal consultations (adjusted mortality risk difference of 5.8%, 95% CI 2.6%-8.9%), and children from polygamous families born during the dry season (adjusted mortality risk difference of 1.7%, 95% CI 0.4%-2.9%). These subgroups, though small, showed a consistent pattern of higher mortality risk over time. Common social and economic factors were linked to a larger share of the total child deaths.

**Conclusions:** The study’s results underscore the need for targeted interventions to address the specific risks faced by these identified high-risk subgroups. These interventions should be designed to work to complement broader public health strategies, creating a comprehensive approach to reducing child mortality. We suggest future research that focuses on developing, testing, and comparing targeted intervention strategies unraveling the proposed hypotheses found in this study. The ultimate aim is to optimize health outcomes for all children in high-mortality settings, leveraging a strategic mix of targeted and general health interventions to address the varied needs of different child subgroups.

(KJMIR Public Health Surveill 2024;10:e48060) doi:10.2196/48060

**KEYWORDS**
child mortality; causal discovery; Guinea-Bissau; inductive-deductive; machine learning; targeted preventive and risk-mitigating interventions

**Introduction**

Child mortality in Guinea-Bissau has decreased significantly over the past 40 years but is still unacceptably high (1 in 13 children dying before the age of 5 years in 2021 [1]). Thus, there is a constant need to design relevant interventions to reduce mortality [2,3]. In particular, identifying subgroups of children at high risk of dying may inform targeted preventive or risk-mitigating interventions to supplement population-wide approaches [4,5].

To identify actionable points for interventions to prevent or mitigate risk, we want to document the fuller causal structure. This spans distal causes such as social and economic conditions, legal rights, and welfare policies to immediate causes such as congenital malformations or infectious agents [6]. However, obtaining high-quality data on these factors can be challenging, particularly in low-income countries. One potential data source is Health and Demographic Surveillance Systems (HDSS), which collect individual-level data on demographics and health for a portion of the population [7].

In this exploratory study, we used HDSS data from urban Bissau, the capital of Guinea-Bissau, to identify subgroups of children at high risk of dying before 3 years of age. We analyzed data from 2003 to 2019, where the birth years from 2003 to 2011 were used to identify risk factors and high-risk groups, which we then tested in the birth years from 2012 to 2016. This allowed us to focus on factors consistently associated with high mortality over time. To do this, we used 3 different types of analyses, that are, spatial analysis to map child mortality in specific areas, regression analysis to identify single risk factors associated with high mortality, and a machine learning model to identify multifactorial risk groups. By integrating these approaches, we aimed to discover subgroups of children with high mortality without being limited to prior hypotheses [8]. Such discoveries are necessary for developing new hypotheses and identifying interventions to reduce child mortality.

**Methods**

**Study Population and Follow-Up**

The study population included children living in Bissau, the capital city of Guinea-Bissau. All the children were part of the HDSS Bandim Health Project [9] and were seen by data collectors within the first 6 weeks of life. Children under 3 years of age are routinely visited every 3-4 months to collect vital and health information. Many recorded child deaths in this population are due to infectious diseases such as respiratory infections, malaria, and diarrhea [10,11].

Follow-up for this study began at 6 weeks of age to ensure that a sufficient proportion of children had their baseline information recorded. Children, who died before 6 weeks of age or did not have complete baseline information, were excluded from the study (30,441/51,446, see flowchart in Figure S1 in Multimedia Appendix 1). To account for the potential selection bias caused by migrating children, inverse probability of censoring weights (IPCW) was used in all analyses and presented results (Multimedia Appendix 2) [12].

**Baseline Information**

To identify relevant factors for child mortality, available baseline information was divided into environmental, household, and individual and birth domains. Figure 1 depicts the assumed causal structure [12] or the data-generating process linking these domains. Operational definitions of the variables and a visualization of their pairwise associations can be found in Table S1 and Figure S2 in Multimedia Appendix 1, respectively.
Analyses

A temporal split of the data, rather than a random split, was used in this study. This allowed us to determine whether identified subgroups consistently had a higher mortality risk across different time periods and thus, may be relevant subgroups for future intervention. The sample size allowed us to divide the data into 2 cohorts. The temporal validation cohort also allowed us to test the robustness of our findings, as there have been various changes over time that may have affected child mortality, such as a significant decline in respiratory infections [11,13].

To describe the temporal trends in child mortality, the Kaplan-Meier estimator was used to calculate overall risks and risks split by age, accounting for censoring during follow-up. All subsequent analyses excluded censored children and used IPCW to adjust for selection bias.

We conducted the following 3 analyses to investigate the association between the baseline information collected before 6 weeks of age and mortality during the entire follow-up period (from 6 weeks to 3 years). All association measures were reported as mortality risk difference (MRD), with 0 indicating no difference between the compared subgroups. The MRD was expressed as a percentage (ie, the difference in deaths per 100 children). All estimates in the results section are adjusted mortality risk differences (aMRD).

Spatial Analysis

We examined whether certain residential areas had a higher mortality risk than others by mapping the children’s households at baseline and moving a sliding window (250 m × 250 m) 10 m at a time to visualize the mortality risk across the study area. Estimates were only presented when at least 100 children were included within the sliding window to avoid small cell sizes. The estimates were adjusted for the linear effect of birth year, as child mortality has approximately decreased linearly by birth year (Figure S3 in Multimedia Appendix 1).

Single Risk Factors

We used generalized linear regressions to investigate the associations between single factors and a higher risk of child mortality. Adjustments were made according to the assumed causal structure (depicted in Figure 1) by blocking the common causes in higher-order domains using the backdoor criterion [12].

Multifactorial Risk Groups

We applied the Causes of Outcome Learning approach [14,15] to identify vulnerable subgroups with a combination of baseline information that was associated with a higher risk of child mortality. This causal inference-inspired machine learning approach has been optimized to prevent causal biases such as confounding by calendar time and collider bias, which could occur in other supervised clustering approaches. Details of the implementation of the Causes of Outcome Learning approach can be found in Multimedia Appendix 3. Since this approach is optimized for interactions in subpopulations, it is expected to find other patterns than the first-order linear regression which averages across the entire population.

Summarization and Causal Modeling

To summarize the findings from the 3 analyses, key statistics such as prevalence, crude risks, and identification of synergistic associations [16] (where the risk from simultaneous exposure to multiple factors is greater than the sum of the individual risks) were calculated. Adjusted risk differences were determined using causal modeling (targeted maximum likelihood estimation [TMLE] [17]) for the defined subgroups compared to all other children. The probability of the estimates from the hypothesis-generating and temporal validation cohorts being similar was also calculated. In addition, a combined estimate for both cohorts was obtained to estimate the population attributable fraction (PAF) [18], which represents the fraction of all mortality that would be prevented if the causal exposure of interest was removed. The analyses were conducted using R (version 4.2; R Core Team), and some sentences were revised using ChatGPT (OpenAI) to improve clarity.

Ethical Considerations

The study does not include biologically, physically invasive, or potentially dangerous procedures. The HDSS collection of data is at the request of the Ministry of Health, Guinea-Bissau.
**Results**

**Overview**

A total of 51,446 children were registered between 2003 and 2019, with 30,441 being excluded from our analysis due to registration after 6 weeks of age, lack of follow-up information, death by 6 weeks of age, missing baseline information, or emigration during follow-up (see flowchart in Figure S1 in Multimedia Appendix 1). The study sample included 21,005 children, which was weighted to an analytical sample of 27,998 children using IPCW to account for nonrandom emigration. The hypothesis-generating and temporal validation cohorts were based on weighted samples of 19,311 and 8687 children, respectively. The weights were not extreme (Multimedia Appendix 2). The mortality risk during the follow-up period (from 6 weeks to 3 years of age) was 5.2% (95% CI 4.8%-5.6%) in the hypothesis-generating cohort and 2.9% (95% CI 2.5%-3.3%) in the temporal validation cohort.

**Spatial Analysis**

We explored the results from the spatial analysis of the hypothesis-generating cohort, which gave rise to defining 4 areas; A, B, C, and D where the child mortality rate was considerably high. We marked these areas with circles on top of the spatial results in Figure 2. By comparing children living in the residential areas marked by circles A, B, C, and D (constituting between 1% [n=253] and 3% [n=533] of children) to those living outside these areas (Figure 2), the aMRD was 4.5% (95% CI –0.6% to 9.6%), 1.9% (95% CI –1.2% to 5.0%), 3.3% (95% CI –0.3% to 6.9%), and 4.0% (95% CI 0.1%-8.0%), respectively. When the 4 suggested high-risk residential areas were assessed in the temporal validation cohort, only area D still tended to exhibit higher mortality though the estimate was associated with more uncertainty (aMRD of 2.0%, 95% CI –2.8% to 6.7%) (Table 1 and Figure S4 in Multimedia Appendix 1). The combined estimate for both cohorts for area D was an aMRD of 3.4% (95% CI 0.3%-6.5%). If causal, the excess risk translates to a PAF of 1.1% of all deaths.

**Figure 2.** Spatial analysis mortality risk among the hypothesis-generating cohort adjusting for a linear effect of calendar time. Mortality risk (deaths per 100 children) in 250 m x 250 m squares by a resolution of 10 m for the birth years 2003-2011. Results are only shown if at least 100 children were under observation. A, B, C, and D indicate areas with a high child mortality risk. E is an uninhabited area, which is flooded during the rainy season.
Table 1. Summary and validations of findings relevant for hypotheses-generations for targeted interventions.

<table>
<thead>
<tr>
<th>Spatial analysis</th>
<th>Hypothesis-generating cohort</th>
<th>Temporal validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence (n=19,311), n (%)</td>
<td>Crude mortality risks (95% CI) within subgroup versus the rest</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>1.3 (253)</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2.0 (381)</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2.8 (533)</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>1.6 (314)</td>
</tr>
<tr>
<td></td>
<td>Single risk factors</td>
<td></td>
</tr>
<tr>
<td>Below 7 years of maternal education</td>
<td>60.8 (11,739)</td>
<td>5.9 (5.5 to 6.4)           vs 4.1 (3.5 to 4.6)</td>
</tr>
<tr>
<td>More children in the household younger than 3 years</td>
<td>83.1 (16,054)</td>
<td>5.5 (5.1 to 5.9)           vs 3.8 (2.9 to 4.6)</td>
</tr>
<tr>
<td>Mothers not under HDSS surveillance</td>
<td>3.6 (690)</td>
<td>19.1 (17.1 to 21.0)        vs 4.7 (4.3 to 5.0)</td>
</tr>
<tr>
<td>No prenatal consultations</td>
<td>4.3 (834)</td>
<td>9.5 (7.7 to 11.3)          vs 5.0 (4.6 to 5.4)</td>
</tr>
</tbody>
</table>

Note: aMRD = average marginal risk difference; TMLE = targeted maximum likelihood estimation; CI = confidence interval; N/A = not applicable; f = full hypothesis-generating cohort; g = temporal validation cohort.
Multifactorial risk groups

<table>
<thead>
<tr>
<th>Being a twin and born in the rainy season</th>
<th>Prevalence (n=19,311), n (%)</th>
<th>Crude mortality risks (95% CI) within subgroup versus the rest</th>
<th>Synergistic associations, 95% CI</th>
<th>TMLE aMRD (deaths per 100 children) in subgroup versus the rest, 95% CI</th>
<th>Prevalence (n=8687), n (%)</th>
<th>Crude mortality risks (95% CI) within subgroup versus the rest</th>
<th>Synergistic associations, 95% CI</th>
<th>TMLE aMRD (deaths per 100 children) in subgroup versus the rest, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children of polygamous families born in the dry season</td>
<td>9.2 (1770)</td>
<td>7.2 (6.0 to 8.4) vs 5.0 (4.6 to 5.4)</td>
<td>1.9 (0.0 to 3.8)</td>
<td>1.8 (0.2 to 3.3)</td>
<td>8.3 (720)</td>
<td>5.7 (4.3 to 7.2) vs 2.7 (2.3 to 3.1)</td>
<td>2.0 (–0.2 to 4.1)</td>
<td>1.7 (–0.4 to 3.9)</td>
</tr>
<tr>
<td>No prenatal consultations, boys, and other ethnicity</td>
<td>0.3 (61)</td>
<td>Very few children</td>
<td>Very few children</td>
<td>Very few children</td>
<td>0.2 (20)</td>
<td>Very few children</td>
<td>Very few children</td>
<td>N/A</td>
</tr>
<tr>
<td>Mothers not under HDSS surveillance, Fula, or Mandinga ethnicity, born in the dry season</td>
<td>0.6 (124)</td>
<td>23.2 (18.6 to 27.9) vs 5.1 (4.7 to 5.4)</td>
<td>5.5 (0.5 to 10.5)</td>
<td>1.0 (–16.1 to 18.1)</td>
<td>0.0 (0)</td>
<td>Very few children</td>
<td>Very few children</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**a**aMRD: adjusted mortality risk difference.

**b**TMLE: targeted maximum likelihood estimation.

**c**The additional risk above the linear effect of the single factors is presented (ie, the parameter for the subgroup parameter while adjusting for each of the variables used to create the subgroup definition). Also adjusted for calendar time but no potential confounders.

**d**The additional risk-adjusted using targeted maximum likelihood estimation with using linear models.

**e**N/A: not applicable.

**f**Adjustment for calendar time.

**g**These findings were considered consistent across the cohorts by the authors.

**h**Adjustment for calendar time and environmental factors.

**i**HDSS: Health and Demographic Surveillance Systems.

**j**Adjustment for calendar time and environmental and household factors.
**Single Risk Factors**

In the hypothesis-generating cohort, children of mothers with less education than 7 years compared to those with 7 years or more of education were common (n=11,739, 60.8%) and had an aMRD of 1.6% (95% CI 0.9%-2.4%; Table 2). Crowding (ie, having multiple children in the household under 3 years of age) was common (n=16,054, 83%) and associated with an aMRD of 1.6% (95% CI 0.6%-2.6%) compared to being the sole child (Table 2). Having functioning electricity, a television, and a toilet inside the house indicates higher wealth, which was associated with lower child mortality. Across both cohorts with very similar estimates, both less maternal schooling and more crowding were associated with an aMRD of approximately 1.5% (Table 1, the column “aMRD visualizations” shows virtually the same estimates). If causal 19.4% of all deaths could be attributed to low-maternal education and 30.0% to crowding.

The most pronounced environmental factor was living within 50 m of a major road, associated with an aMRD of 2.1% (95% CI 0.6%-3.6%) compared with children living further away.

Children of mothers lost to follow-up were at a marked increased mortality risk. Still, they constituted a very small number of children in the temporal validation cohort (additional explanation in Table S1 in Multimedia Appendix 1).

Being a twin was consistently associated with higher mortality, with an aMRD of 4.3% (95% CI 1.0%-7.7%) and 3.0% (95% CI –0.5% to 6.5%) in the hypothesis-generating and temporal validation cohorts, respectively (Table 1).

No prenatal consultation was recorded for 4% (n=834) of the children in the hypothesis-generating cohort and was associated with an aMRD of 6.4% (95% CI 2.5%-10.2%; Table 1). In the temporal validation data set, this was associated with an aMRD of 2.8% (95% CI –0.9% to 6.6%).
Table 2. The association between single risk factors and child mortality below 3 years of age. Hypothesis-generating cohort (Bandim Health Project Health and Demographic Surveillance Systems data from the birth years from 2003 to 2011).

<table>
<thead>
<tr>
<th>Baseline information and category</th>
<th>Prevalence of children (n=19,311), n (%)</th>
<th>Deaths per 100 children, n</th>
<th>Unadjusted mortality risk difference (95% CI; additional deaths per 100 children)</th>
<th>Adjusted mortality risk difference* (95% CI; additional deaths per 100 children)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Environmental</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vegetation near the household at birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The 90% least vegetation</td>
<td>17,890 (93b)</td>
<td>5.2</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>The top 10% vegetation</td>
<td>1421 (7b)</td>
<td>5.1</td>
<td>−0.1 (−1.5 to 1.2)</td>
<td>−0.1 (−1.5 to 1.3)</td>
</tr>
<tr>
<td><strong>Distance to a major road</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within 50 m</td>
<td>1178 (6)</td>
<td>3.2</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>More than 50 m</td>
<td>18,133 (94)</td>
<td>5.3</td>
<td>2.1 (0.6 to 3.6)</td>
<td>2.1 (0.6 to 3.6)</td>
</tr>
<tr>
<td><strong>Population density 500 m × 500 m</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90% lowest density</td>
<td>17,659 (91b)</td>
<td>5.2</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>10% highest density</td>
<td>1652 (9b)</td>
<td>5.6</td>
<td>0.5 (−0.8 to 1.8)</td>
<td>0.4 (−0.9 to 1.7)</td>
</tr>
<tr>
<td><strong>Vaccination against tuberculosis in the local area</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below 80%</td>
<td>551 (3)</td>
<td>4.6</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>80% or above</td>
<td>18,760 (97)</td>
<td>5.2</td>
<td>0.6 (−1.6 to 2.8)</td>
<td>0.4 (−1.8 to 2.6)</td>
</tr>
<tr>
<td><strong>Distance to a health center</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤1 km</td>
<td>17,849 (92)</td>
<td>5.1</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>&gt;1 km</td>
<td>1462 (8)</td>
<td>6.0</td>
<td>0.9 (−0.5 to 2.3)</td>
<td>1.0 (−0.4 to 2.3)</td>
</tr>
<tr>
<td><strong>Household</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother lost to follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18,621 (96)</td>
<td>4.7</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>No</td>
<td>690 (4)</td>
<td>19.1</td>
<td>14.4 (12.5 to 16.3)</td>
<td>14.4 (12.5 to 16.4)</td>
</tr>
<tr>
<td>Whether the mother lives with the father</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>12,255 (63)</td>
<td>4.8</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>No</td>
<td>7056 (37)</td>
<td>5.8</td>
<td>1.0 (0.2 to 1.7)</td>
<td>1.0 (0.3 to 1.8)</td>
</tr>
<tr>
<td><strong>Roof</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zinc</td>
<td>18,581 (96)</td>
<td>5.1</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>Other</td>
<td>730 (4)</td>
<td>6.3</td>
<td>1.1 (−0.8 to 3.0)</td>
<td>0.9 (−1.0 to 2.8)</td>
</tr>
<tr>
<td><strong>Electricity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5995 (31)</td>
<td>4.5</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>No</td>
<td>13,317 (69)</td>
<td>5.5</td>
<td>1.0 (0.2 to 1.7)</td>
<td>0.9 (0.1 to 1.7)</td>
</tr>
<tr>
<td><strong>Television</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>6338 (33)</td>
<td>4.4</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>No</td>
<td>12,973 (67)</td>
<td>5.6</td>
<td>1.2 (0.5 to 2.0)</td>
<td>1.0 (0.3 to 1.8)</td>
</tr>
<tr>
<td><strong>Toilet</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inside the house</td>
<td>2916 (15)</td>
<td>4.3</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>Other</td>
<td>16,395 (85)</td>
<td>5.3</td>
<td>1.0 (0.0 to 2.0)</td>
<td>0.8 (−0.2 to 1.9)</td>
</tr>
<tr>
<td><strong>Maternal schooling (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥7</td>
<td>7572 (39)</td>
<td>4.1</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>&lt;7</td>
<td>11,739 (61)</td>
<td>5.9</td>
<td>1.9 (1.1 to 2.6)</td>
<td>1.6 (0.9 to 2.4)</td>
</tr>
<tr>
<td><strong>Polygamous families</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline information and category</td>
<td>Prevalence of children (n=19,311), n (%)</td>
<td>Deaths per 100 children, n</td>
<td>Unadjusted mortality risk difference (95% CI; additional deaths per 100 children)</td>
<td>Adjusted mortality risk difference (95% CI; additional deaths per 100 children)</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------------------------------------</td>
<td>--------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>No</td>
<td>15,803 (82)</td>
<td>5.0</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>Yes</td>
<td>3508 (18)</td>
<td>6.2</td>
<td>1.2 (0.3 to 2.1)</td>
<td>1.1 (0.2 to 2.1)</td>
</tr>
<tr>
<td><strong>Mother works outside of the home</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>555 (3)</td>
<td>4.3</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>No</td>
<td>18,756 (97)</td>
<td>5.2</td>
<td>0.9 (−1.3 to 3.1)</td>
<td>0.9 (−1.2 to 3.1)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balanta</td>
<td>1637 (8)</td>
<td>4.5</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>Fula or Madinga</td>
<td>5359 (28)</td>
<td>4.9</td>
<td>0.3 (−1.1 to 1.8)</td>
<td>0.4 (−1.0 to 1.8)</td>
</tr>
<tr>
<td>Manjaco or Mancanha</td>
<td>3758 (19)</td>
<td>5.3</td>
<td>0.7 (−0.8 to 2.2)</td>
<td>0.8 (−0.7 to 2.3)</td>
</tr>
<tr>
<td>Pepel</td>
<td>3154 (16)</td>
<td>5.3</td>
<td>0.8 (−0.8 to 2.3)</td>
<td>0.9 (−0.6 to 2.4)</td>
</tr>
<tr>
<td>Other</td>
<td>5404 (28)</td>
<td>5.6</td>
<td>1.0 (−0.4 to 2.4)</td>
<td>1.1 (−0.4 to 2.5)</td>
</tr>
<tr>
<td><strong>Other children below 3 years in the household</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3257 (17)</td>
<td>3.8</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>Yes</td>
<td>16,054 (83)</td>
<td>5.5</td>
<td>1.7 (0.8 to 2.7)</td>
<td>1.6 (0.7 to 2.6)</td>
</tr>
<tr>
<td><strong>Information related to delivery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boy</td>
<td>9880 (51)</td>
<td>5.3</td>
<td>0 (ref)</td>
<td>N/A^d</td>
</tr>
<tr>
<td>Girl</td>
<td>9431 (49)</td>
<td>5.1</td>
<td>−0.3 (−1.0 to 0.5)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Twin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>18,696 (97)</td>
<td>5.1</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>Yes</td>
<td>615 (3)</td>
<td>9.4</td>
<td>4.3 (2.2 to 6.4)</td>
<td>4.3 (2.2 to 6.3)</td>
</tr>
<tr>
<td><strong>Birth season</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dry</td>
<td>9836 (51)</td>
<td>5.4</td>
<td>0 (ref)</td>
<td>N/A</td>
</tr>
<tr>
<td>Rainy</td>
<td>9475 (49)</td>
<td>4.9</td>
<td>−0.5 (−1.2 to 0.2)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Place of birth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital or health center</td>
<td>13,786 (71)</td>
<td>5.0</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>At home</td>
<td>5525 (29)</td>
<td>5.6</td>
<td>0.6 (−0.2 to 1.4)</td>
<td>0.1 (−0.8 to 0.9)</td>
</tr>
<tr>
<td><strong>Maternal age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;25</td>
<td>9780 (51)</td>
<td>4.9</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>≤25</td>
<td>9531 (49)</td>
<td>5.5</td>
<td>0.6 (−0.1 to 1.4)</td>
<td>0.4 (−0.4 to 1.2)</td>
</tr>
<tr>
<td><strong>Firstborn</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not firstborn</td>
<td>13,389 (69)</td>
<td>5.0</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>Firstborn</td>
<td>5922 (31)</td>
<td>5.6</td>
<td>0.5 (−0.2 to 1.3)</td>
<td>0.6 (−0.3 to 1.4)</td>
</tr>
<tr>
<td><strong>Born by cesarean section</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>950 (5)</td>
<td>4.0</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
<tr>
<td>No</td>
<td>18,361 (95)</td>
<td>5.3</td>
<td>1.2 (−0.5 to 2.9)</td>
<td>1.2 (−0.5 to 2.9)</td>
</tr>
<tr>
<td><strong>Prenatal consultations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18,478 (96)</td>
<td>5.0</td>
<td>0 (ref)</td>
<td>0 (ref)</td>
</tr>
</tbody>
</table>
Adjusted mortality risk difference

Unadjusted mortality risk difference (95% CI; additional deaths per 100 children)

Deaths per 100 children, n

Prevalence of children (n=19,311), n (%)

Baseline information and category

<table>
<thead>
<tr>
<th>Category</th>
<th>Prevalence of children</th>
<th>Deaths per 100 children, n</th>
<th>Unadjusted mortality risk difference (95% CI; additional deaths per 100 children)</th>
<th>Adjusted mortality risk difference (95% CI; additional deaths per 100 children)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>834 (4)</td>
<td>9.5</td>
<td>4.5 (2.7 to 6.3)</td>
<td>3.9 (2.1 to 5.7)</td>
</tr>
</tbody>
</table>

*a* For the environmental category, adjusted risk differences are adjusted for a calendar effect, risk differences in the household category are additionally adjusted for the environmental variables, and risk differences in the information related to the delivery category are further adjusted for the household variables.

*b* The prevalence differs from 10% to 90% because the cutoff is based on data from both cohorts (2003-2016).

*c* Reference group.

*d* N/A: not applicable, since the effect of sex and birth season is not expected to be confounded and thus does not need adjustment.

**Multifactorial Risk Groups**

In the hypothesis-generating cohort, twins born in the rainy season had higher mortality risk compared with those born in the dry season (aMRD of 7.0%, 95% CI 2.0%-11.9%; Figure 3, group 4, and Table 1), but this association was not found in the temporal validation data (aMRD 1.3%, 95% CI –3.0% to 5.6%; Table 1). Children of polygamous families born in the dry season had an aMRD of 1.8% (95% CI 0.2%-3.3%) compared to all other children (Figure 3, group 5, and Table 1). This subgroup constituted 9% (n=1770) of the children in the hypothesis-generating cohort, and the finding was consistent in the temporal validation cohort (an aMRD of 1.7%, 95% CI –0.4% to 3.9%, covering 8% [n=720] of all children; Table 1, see the column with aMRD visualization for consistency). If these associations are causal, 3.3% of all deaths could be attributed to this combination. A supplementary analysis of both cohorts (birth years 2003-2016) was conducted to understand the phenomenon better. The results suggested that (1) the finding was not artificially introduced by the IPCW approach; (2) the increased risk was highest in the first half-year of follow-up (6 weeks to 7 months of age) but continued throughout the entire follow-up period (up to 3 years of age); (3) the association varied across birth years without any trend; (4) the association was strongest among the Manjaco and Mancanha ethnic groups; (5) the association was most pronounced in the eastern part of the HDSS area; and (6) the association was not confounded by crowding, but was driven by the strata of children living in a household with other children under 3 years of age (Multimedia Appendix 4).
Figure 3. Multifactorial risk groups results from the Causes of Outcome Learning approach. (A) The prevalence and mean risk by the identified subgroups based on the Causes of Outcome Learning approach. Group 8 had a higher mortality risk than could be shown in the plot. (B) The mean risk contributions by each subgroup are visualized, where the color white indicates no risk contribution and dark red indicates the highest risk contribution. We identify 4 subgroups with a prevalence above 0.5% (n=97): (1) twins born in the rainy season (group 4); (2) children of polygamous families born in the dry (group 5); (3) no consultations, boys, and other ethnicity (group 6); (4) mothers not under Health and Demographic Surveillance Systems surveillance, Fula, or Mandinga ethnicity, born in the dry season (group 7). For more details, see Multimedia Appendix 3. BCG: Bacillus Calmette-Guérin vaccination.
Discussion

Principal Findings
In this study, we aimed to discover subgroups of children with high mortality in urban Guinea-Bissau. We used complementary analyses and split the data into a hypothesis-generating cohort and a temporal validation cohort. Of children with high mortality who may be targeted for interventions, we identified (1) a residential area (area D), (2) children of mothers who did not attend prenatal consultations, and (3) children born in polygamous families during the dry season had excess mortality risk throughout the study period. Of population-wide findings, maternal education and household crowding were important factors.

Limitations
Excluding children without complete information (Figure S1 in Multimedia Appendix 1), and conditioning on children being alive at 6 weeks of age [19] may have limited the generalizability of our findings. The HDSS is a valuable source of information but is focused on specific key health indicators as data collection in Guinea-Bissau is resource-demanding. Thus, we lacked information about other relevant baseline characteristics such as vector-borne diseases, health care and systems, weather, pollution, water and sanitation, community relations (social capital), and household and macroeconomic conditions. We did not include factors that varied during follow-up, such as season, vaccinations, vitamins, and other health campaigns. This may be important as child mortality is considerably higher in the rainy season [19,20]. Live and nonlive vaccines have been shown to affect child mortality more generally than their effects on the targeted diseases explained [21].

We acknowledge that data gathered from the real world may have some natural limitations in terms of completeness and accuracy, which could potentially affect the reliability of the identified risk factors. Efforts to account for missing or incomplete data were made where feasible. Furthermore, as this study concentrates on urban Guinea-Bissau, its findings may not readily apply to different socioeconomic and cultural contexts. Further research in varied settings is necessary to validate and understand the transferability of these factors.

To acknowledge the challenges in establishing causality, we integrated multiple methods and used an inductive-deductive research methodology (ie, take the learnings from the hypothesis-generating cohort to be tested in the temporal validation cohort) [22]. This approach guided us to propose future research directions to validate and understand the mechanisms driving the observed phenomena. While total effects can be diluted in the Causes of Outcome Learning approach, particularly when including individual and birth-related factors in the model (Figure 3), our methodology is strengthened by using a causal structure (Table 1) for adjustments and the TMLE approach. This enhances the validity of our findings and contributes toward a more robust inference of causality by better adjustment and more robust model specification [17]. It could be explored if other novel machine learning methods could supplement the findings [23].

Interpretation
The temporal split allowed us to investigate consistency across 2 time periods. While a lack of consistency can be due to chance, it may also reflect changes in the causal structure over time. We found that children of mothers not under HDSS surveillance were strongly associated with child mortality in the hypothesis-generating cohort. However, in the temporal validation cohort, close to no children had mothers who were not under HDSS surveillance. Changes in data collection methods might explain this discrepancy; after 2013, mothers of children in new families were registered by the same data collector as the child, whereas before 2013, the mother’s registration was handled by a separate team. The increased risk among twins born in the rainy season in the hypothesis-generating cohort may have occurred by chance. Still, it could also indicate better health care for high-risk children in the validation cohort.

Local Environmental Factors
Previous spatial studies have shown large differences in disparities within and between countries [24,25], and temporal persistence at local levels [26]. The population movement in Bissau may have made it more difficult to identify high-risk residential areas. Area D contains a busy market called Caracol which is known for traditional medicine and care. High population density and possibly high infectious load, may offer 1 explanation for the high risk in residential area D. The proportion of mothers with less than 7 years of schooling was similar in area D as outside of it (Figure S5 in Multimedia Appendix 1).

To further understand and address the high mortality rate in this residential area, several future studies could be conducted, such as (1) qualitative study following families in this area may add insight and create new hypotheses; (2) network analysis to reveal contact patterns and exposed jobs most relevant in this area; and (3) spatial analysis of distance to specific proximate places (eg, places for traditional medicine and care), infrastructures (eg, wells), or potentially hazardous areas (eg, waste collection areas).

Lack of Prenatal Consultations
Prenatal consultations are designed to prevent early child mortality and may directly affect maternal behavior. The association between lack of a prenatal consultation and mortality is reflected in other studies [27], but we cannot exclude that some of the association was confounded by social and economic factors, as well as health care-seeking behavior. This may be especially important as we are considering postneonatal mortality. Various mechanisms may contribute to postneonatal mortality, such as out-of-pocket fees associated with increased child mortality in sub-Saharan Africa [28]. In Guinea-Bissau, the expansion of free antenatal care was, however, not associated with reduced perinatal mortality [29], and thus some of the observed associations of prenatal consultations may reflect confounding.

To further understand and address the lack of prenatal consultations and its impact on child mortality, a number of future studies could be conducted, such as (1) studies examining...
various characteristics of mothers not participating in prenatal consultation and their outcomes to understand further how this subgroup is associated with mortality and morbidity, (2) assess the effectiveness of interventions such as active home visits with prenatal consultations in reducing child mortality, and (3) explore if health care decisions during prenatal consultations can be assisted by artificial intelligence–based assessment systems [30].

**Family Type and Birth Season**

Connecting children of polygamous families born in the dry season, an HDSS-based study from the Gambia identified that children born in the harvest season (January–June, approximately equivalent to the dry season in Guinea-Bissau which is December–May) were at increased postneonatal mortality risk [31] and a study from Ghana found that children from polygamous families had higher child mortality than those of monogamous families [32]. We could not identify other studies assessing the combination of family type and birth season. Within our study, we further observed that the finding was not indicated to be confounded by crowding, though residual confounding may persist. However, we found that the pattern was only present for children living in households with other children under 3 years of age. Some mothers travel to rural villages to harvest cashew nuts in the late dry season and return in the rainy season. One explanation may lie in the divided attention between labor in the cashew plantations and care for other children (potentially in a different environment). With reduced or limited breastfeeding during the cashew harvest, children may lose maternal antibodies and thus become more susceptible to infections. How these mechanisms interact with family structure is still to be understood.

To further understand and address the association between birth season and family type on child mortality, several future studies could be conducted, such as (1) investigation into accidents as causes of death may reveal if the combination of shared child attendance and busy months in relation to the harvest increases the risk of domestic accidents, (2) interviews with these families may give insight into the observed phenomenon, and (3) triangulating the findings with other health-related behaviors such as vaccination uptake may help uncover mechanisms.

**Resource Prioritization**

As repeatedly described in the literature, social and economic factors affecting a wide part of the population strongly predict mortality [33]. In our data, social and economic factors may account for 20%-30% of all deaths in children aged 6 weeks to 3 years. In contrast, the 3 subgroups of children with high mortality identified in our study may represent a smaller fraction of the overall mortality burden (less than 5%), but they are characterized by significantly higher absolute mortality risks. This distinction raises important questions about the feasibility and potential impact of targeted interventions for these subgroups as compared to more widespread, universal public health strategies. While recognizing the challenges in reaching these smaller subgroups, targeted interventions could be crucial in addressing their disproportionately high mortality risks. Therefore, it is imperative to consider both cost-effectiveness and equity in designing these interventions, ensuring they complement broader public health measures to provide comprehensive and effective child health care.

**Demonstration of a Novel Approach for Targeted Public Health Research**

This study not only provides insights into child mortality in urban Guinea-Bissau but also demonstrates the practical application of the Causes of Outcome Learning approach [14] on real-world data. Our findings illustrate how this approach effectively deciphers complex patterns and suggests potential synergistic causes in public health data, revealing phenomena that would be overlooked by traditional analytical methods. Future research should focus on identifying when the Causes of Outcome Learning approach is most effective and on refining the methodology to improve its accuracy and adaptability for a variety of public health research questions and study designs.

**Conclusions**

Reaching the Sustainable Development Goal of reducing under-5 child mortality to below 1 in 25 children by 2030 will require a range of interventions. By using several different and complementary approaches, we were able to identify subgroups of children at a high mortality risk that would not be evident otherwise. These high-risk children live in a specific area near a marked area known for traditional medicine and care, have mothers who did not attend prenatal consultations, and were born in the dry season and in polygamous families. We have suggested several future studies that may help explore these hypotheses. Potential targeted interventions should be evaluated in comparison with the impact of population-wide structural interventions both from cost-effectivity aspects and equity aspects and tested under proper evaluation schemes [34] to reduce child mortality.

**Acknowledgments**

The study would not have been possible without the dedicated work of the many data collectors, supervisors, and mothers of children under surveillance who were willing to provide answers to the questions. AR was supported by an International Postdoctoral Grant (9034-00006B) from the Independent Research Fund Denmark. HA was supported by Novo Nordisk Foundation Challenge Programme (NNF17OC0027812) and by National Institutes of Health (UL1TR004419 and P30ES023515). SWM was supported by a DFF-International Postdoctoral Grant (0164-00023B) from the Independent Research Fund Denmark. ABF was supported by an Ascending investigator grant from Lundbeck Foundation (R313-2019-635) and a Sapere Aude grant from Independent Research Fund Denmark (9060-00018B). Many funders supported the data collection at Bandim Health Project over the years (full list available [35]).
Authors' Contributions
The study was based on secondary analyses of data collected at the Bandim Health Project under the supervision of SWM, IBS, PA, and ABF. All authors contributed to the study concept and design. AR, SN, and ABF extracted and verified the data. AR is responsible for all the data analyses. AR, PD, WS, NHR, and CTE contributed to the implementation of the Causes of Outcome Learning approach. All authors contributed to the interpretation of the results. AR, SN, and ABF contributed to drafting the study. All authors critically revised the study for important intellectual content.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The variable operationalisation, flowchart, pairwise associations between basic HDSS information, temporal overview, geo-spatial patterns of the temporal validation data adjusting for a linear effect of calendar time, and geo-spatial patterns of the prevalence of mothers with fewer than 7 years of schooling.

Multimedia Appendix 2
The inverse probability of censoring weights.

Multimedia Appendix 3
The Causes of Outcome Learning method.

Multimedia Appendix 4
The analysis regarding the birth season and polygamous families.

References
9. Bandim Health Project: a health and demographic surveillance system site situated in Guinea-Bissau, West Africa. Bandim Health Project. URL: https://www.bandim.org/ [accessed 2023-12-14]


Abbreviations

- aMRD: adjusted mortality risk differences
HDSS: Health and Demographic Surveillance Systems
IPCW: inverse probability of censoring weights
MRD: mortality risk difference
PAF: population attributable fraction
TMLE: targeted maximum likelihood estimation
Data-Driven Identification of Potentially Successful Intervention Implementations Using 5 Years of Opioid Prescribing Data: Retrospective Database Study

Lisa EM Hopcroft¹, PhD; Helen J Curtis¹, DPhil; Richard Croker¹, MSc; Felix Pretis², DPhil; Peter Inglesby¹, MPhil; David Evans¹, MPhil; Sebastian Bacon¹, BA; Ben Goldacre¹, MRCPsy; Alex J Walker¹, PhD; Brian MacKenna¹, MPharm

¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
²Department of Economics, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Alex J Walker, PhD
Nuffield Department of Primary Care Health Sciences
Radcliffe Primary Care Building
Observatory Quarter
Oxford, OX2 6GG
United Kingdom
Phone: 44 01865289313
Email: alex.walker@phc.ox.ac.uk

Abstract

Background: We have previously demonstrated that opioid prescribing increased by 127% between 1998 and 2016. New policies aimed at tackling this increasing trend have been recommended by public health bodies, and there is some evidence that progress is being made.

Objective: We sought to extend our previous work and develop a data-driven approach to identify general practices and clinical commissioning groups (CCGs) whose prescribing data suggest that interventions to reduce the prescribing of opioids may have been successfully implemented.

Methods: We analyzed 5 years of prescribing data (December 2014 to November 2019) for 3 opioid prescribing measures—total opioid prescribing as oral morphine equivalent per 1000 registered population, the number of high-dose opioids prescribed per 1000 registered population, and the number of high-dose opioids as a percentage of total opioids prescribed. Using a data-driven approach, we applied a modified version of our change detection Python library to identify reductions in these measures over time, which may be consistent with the successful implementation of an intervention to reduce opioid prescribing. This analysis was carried out for general practices and CCGs, and organizations were ranked according to the change in prescribing rate.

Results: We identified a reduction in total opioid prescribing in 94 (49.2%) out of 191 CCGs, with a median reduction of 15.1 (IQR 11.8-18.7; range 9.0-32.8) in total oral morphine equivalence per 1000 patients. We present data for the 3 CCGs and practices demonstrating the biggest reduction in opioid prescribing for each of the 3 opioid prescribing measures. We observed a 40% proportional drop (8.9% absolute reduction) in the regular prescribing of high-dose opioids (measured as a percentage of regular opioids) in the highest-ranked CCG (North Tyneside); a 99% drop in this same measure was found in several practices (44%-95% absolute reduction). Decile plots demonstrate that CCGs exhibiting large reductions in opioid prescribing do so via slow and gradual reductions over a long period of time (typically over a period of 2 years); in contrast, practices exhibiting large reductions do so rapidly over a much shorter period of time.

Conclusions: By applying 1 of our existing analysis tools to a national data set, we were able to identify rapid and maintained changes in opioid prescribing within practices and CCGs and rank organizations by the magnitude of reduction. Highly ranked organizations are candidates for further qualitative research into intervention design and implementation.

(JMIR Public Health Surveill 2024;10:e51323) doi:10.2196/51323)
Introduction

The prescription of opioids is common and appropriate in the management of acute pain, but their efficacy with regards to chronic pain is not supported by empirical evidence [1], and there is a global problem with opioid overuse [2]. Long-term use of opioids has been shown to be associated with the accumulating risk of dependence and overdose [3]. The continually rising rates of opioid prescription, particularly in England and Wales [4-6], prompted the publication of new guidance in 2010 [7] advocating for a cautious approach in the long-term prescribing of opioids [8], and opioids have been a specific priority for governmental advisory groups [9]. In 2019, Public Health England (PHE) published the *Prescribed Medicines Review*, which aimed to “identify the scale, distribution and causes of prescription drug dependence, and what might be done to address it” [10]. This review included data from the National Health Service Business Services Authority (NHSBSA) primary care prescription data set, which suggested that some progress had been made in reducing opioid prescribing, with a small but consistent fall in rates between 2015 and 2018. However, there was also evidence that opioid prescribing remains a persistent public health problem in England, with higher rates of prescription in areas of higher deprivation and evidence that long-term prescribing was associated with opioid overdose and dependence. The first recommendation of this report was “increasing the availability and use of data on the prescribing of medicines that can cause dependence or withdrawal to support greater transparency and accountability and help ensure practice is consistent and in line with guidance” [10].

Our group produces OpenPrescribing [11], which allows open access to the same NHSBSA primary care prescription data set used in the PHE review. It is a free and widely used tool with 20,000 unique users per month, where anyone can explore the prescriptions dispensed at any practice in England and monitor prescribing patterns down to the level of individual brands, formulations, and doses.

In OpenPrescribing, we perform automated analyses to generate monthly reports covering 80 measures of prescribing safety, effectiveness, and cost. Our analyses included all general practices (GPs) and their regional organizations, which were known as clinical commissioning groups (CCGs) at the time of this study. Several measures have been developed to capture trends and variations in opioid prescribing [12]. This window into national opioid prescribing data presents an opportunity to identify changes—both increases and decreases—in prescribing that could inform National Health Service (NHS) decision-making and policy.

It is our experience that the best practice is typically defined by organizations identifying themselves as having improved, following the implementation and internal assessment of interventions. We are seeking to pursue an alternative, data-driven, and unbiased approach that instead exploits the national prescribing data set to identify prescribing patterns that may be representative of best practice (i.e., where we can identify a significant reduction in opioid prescribing).

We set out to apply our change detection algorithm [13] to identify patterns indicative of maintained and significant reduction that may help identify best practices with regard to opioid prescribing policy.

Methods

Study Design

We conducted a retrospective database study using GP primary care electronic health record data from all GPs in England.

Data Source

We extracted data from the OpenPrescribing database. This imports openly accessible prescribing data from the large monthly files published by the NHSBSA, which contain data on cost and items prescribed for each month, for every typical GP and CCG in England since mid-2010 [14]. We extracted data up to November 2019. We note that CCGs were replaced by integrated care boards as of July 1, 2022. We have retained results by CCGs as this was an active administrative unit of the NHS in England during the study period. The monthly prescribing data sets contain 1 row for each different medication and dose, in each prescribing organization in NHS primary care in England, describing the number of items (i.e., prescriptions issued) and the total cost. These data are sourced from community pharmacy claims data and, therefore, contain all items that were dispensed. We extracted all available data for typical GPs, excluding other organizations such as prisons and hospitals, according to the NHS Digital data set of practice characteristics [15]. The numbers of patients registered at each practice were obtained from NHS Digital [15].

Study Measures

A total of 3 measures were used in this study to capture various aspects of opioid prescribing. The first (“total oral morphine equivalence per 1000 patients”) expresses the oral morphine equivalence (OME) of all opioid prescriptions per 1000 patients [16]. The second and third look to capture information about regularly prescribed opioids—those used on a regular basis to control pain rather than preparations used for breakthrough pain or opioid injections. Of the regularly prescribed opioids, high-dose opioids were defined as those with ≥120 mg OME per day [8]. The “high dose opioids as percentage regular opioids” measure captures the number of prescriptions of these high-dose, regularly prescribed opioids as a percentage of all long-acting opioids [17]; the “high dose opioid items per 1000 patients” measure captures the same number of high-dose, long-acting opioids but expresses this per 1000 patients [18]. For all measures, higher values represent higher rates of opioid prescription.
In England, an individual will be registered at 1 GP or practice; and each practice belonged, at the time of analysis, to a regional CCG that can influence their prescribing. These organizations and their membership can change over time (e.g., a practice may be reassigned to a different CCG, a CCG may be renamed or replaced, or a practice may close). In our results, we report results for any practice or CCG that existed during the study period, acknowledging that some of these no longer exist. CCGs have now also been replaced with subintegrated care board locations, but some still retain their previous CCG code.

Practices may act independently to change prescribing or participate in an action coordinated by their CCG. We, therefore, conducted analysis at both organizational levels. Monthly values for each measure were calculated for every practice and CCG between December 2014 and November 2019 (this study period was chosen so as to assess prescribing rates over a reasonable period of time, without being affected by the COVID-19 pandemic). The monthly data were summarized as deciles and presented as decile charts across all practices or CCGs each month.

**Statistical Methods**

For this study, we used our innovative change detection Python library (available via the Python Package Index) [19], which is an automated method of detecting change in time-series data. This algorithm was originally developed to determine how clinicians vary in their response to new guidance on existing or new interventions. By measuring the timing and magnitude of change in the relevant organizations, it is able to identify both steep, sudden changes and more gradual, smooth transitions over multiple months. The full methods are described elsewhere [13] and the code is available for anyone to use as a single command with our open Python library [19].

Data for each of the 3 measures were analyzed for all 191 CCGs and 7458 practices. The time series for each organization was analyzed using our change detection algorithm (using the default parameters) to identify the location and magnitude of significant reductions in the measure (substantial increases were filtered out as they are not relevant to the research question). These results were then filtered to remove (1) a total of 678 closed or dormant practices and (2) a further 237 practices with a list size of less than 2000 (this latter group was excluded to avoid analyses of time series with a high level of noise due to low prescribing volume); this process left 6543 practices to be subject to further analysis. We filtered out practices where more than half of the monthly denominator values were 0 (either no registered population or no total opioid prescribing as per measure definitions) across the study period. Among the organizations where our code detected a substantial reduction, for each measure, we selected those whose starting level immediately before the reduction was in the top 20% of all practices (top 150) or CCGs (top 38); this was to remove any organizations with consistently low prescribing from our results. For each measure, we then ranked practices and CCGs by the total measured change (the percentage reduction between the predrop value and the end-drop value) to identify which organizations exhibited the most substantial reductions.

The decile plots provided show an individual organization’s prescribing rates across the period (thick red line), in the context of all peer organizations (summarized using deciles, as blue lines).

**Software and Reproducibility**

Data management and analysis were carried out using Python (version 3.8; Python Software Foundation) and Google BigQuery. Our change detection library [19] is a Python wrapper for the GETS R package [20]. All our methods and underlying code are openly available on GitHub [21]. The full results, summary statistics of changes detected, and top 10 CCGs and practices can be seen in the notebooks folder, in the files ccg-opioids-change-detection-analysis.ipynb and practice-opioids-change-detection-analysis.ipynb. All organizations that existed in the study period (including those that have since closed or been replaced) are included in these reports.

**Ethical Considerations**

This study uses open, publicly available, and anonymized data. This analysis did not need a review from an institutional review board because it used previously collected, fully anonymized data [22]. Informed consent and compensation were similarly not required and would not be possible.

**Results**

**Overview**

We identified substantial reductions in at least 49% of all CCGs (94/191, 49.2%) and practices (4100/7460, 55%) for all measures; summary statistics for these reductions are provided in Table 1. Note that these data describe all substantial reductions detected, that is, before filtering for a top 20% (top 38 CCGs or top 150 practices) starting value. For both CCGs and practices, reductions are on average greater for both high-dose opioid prescribing measures as compared to those observed for the total OME measure, although the IQR values demonstrate that there is also more variability in the high-dose opioid prescribing measures. Reductions appear more modest among CCGs than practices (with lower medians and lower maximum values), but these reductions may be more consistent (with lower variability and greater minimum values observed in CCGs as compared to practices). There is at least 1 practice in each measure where the reduction is almost 99% to 100% and at least 1 practice where the reduction detected is very close to 0.
Table 1. Summary of all opioid reductions identified across clinical commissioning groups (n=191) and practices (n=7458) in England between December 2014 and November 2019 using the automated change detection algorithm.

<table>
<thead>
<tr>
<th>Organization and measure</th>
<th>Count, n</th>
<th>Reduction (%), median (IQR)</th>
<th>Reduction (%), range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical commissioning groups (n=191)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total OME\textsuperscript{b} per 1000 patients</td>
<td>94</td>
<td>15.1 (11.8-18.7)</td>
<td>9.0-32.8</td>
</tr>
<tr>
<td>High-dose opioids as percentage regular opioids</td>
<td>168</td>
<td>19.0 (13.7-25.8)</td>
<td>3.6-41.5</td>
</tr>
<tr>
<td>High-dose opioids per 1000 patients</td>
<td>115</td>
<td>22.2 (17.2-30.0)</td>
<td>1.0-45.4</td>
</tr>
<tr>
<td><strong>Practices (n=7460)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total OME per 1000 patients</td>
<td>4100</td>
<td>28.2 (19.8-39.7)</td>
<td>0.1-99.1</td>
</tr>
<tr>
<td>High-dose opioids as percentage regular opioids (%)</td>
<td>4632</td>
<td>47.7 (33.0-65.9)</td>
<td>0.0-100.0</td>
</tr>
<tr>
<td>High-dose opioids per 1000 patients</td>
<td>4334</td>
<td>56.0 (37.7-73.4)</td>
<td>0.0-100.0</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Count indicates the number of organizations (clinical commissioning groups or practices) in which a reduction was identified. Median, IQR, and range summarize the size of the reductions identified in those organizations (expressed as % reduction from the predrop value to the end-drop value).

\textsuperscript{b}OME: oral morphine equivalence.

**Changes for CCGS**

Table 2 illustrates the CCGs that exhibited the biggest reduction in each of the 3 OpenPrescribing measures over the study period, detailing the proportion of change and the month in which the change started. Note that these CCGs meet the criteria for identification, that is, their prescribing rate immediately before the reduction was in the top 38 (20%) CCGs.

The total OME measure shows a gradual reduction over time in all 3 CCGs, with the algorithm identifying a reduction of up to 31%. The results for the 2 regular high-dose opioid measures also exhibit a gradual reduction over time but capture greater reductions in regular high-dose opioid prescription, with 40% and 39% reductions identified as a proportion of all regular opioids and per 1000 patients respectively.
Table 2. Automatically detected changes across 3 measures of opioid prescribing in a retrospective prescribing database study of CCGs in England.

<table>
<thead>
<tr>
<th>Measure and rank</th>
<th>CCG</th>
<th>Absolute change detected (difference)</th>
<th>Proportional change (%)</th>
<th>Month when change was detected</th>
<th>Decile chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total OME(^e) per 1000 patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Vale Royal</td>
<td>15,461</td>
<td>31</td>
<td>November 2015</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Great Yarmouth and Waveney</td>
<td>19,182</td>
<td>26</td>
<td>February 2017</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Heywood, Middleton and Rochdale</td>
<td>15,393</td>
<td>26</td>
<td>August 2017</td>
<td></td>
</tr>
<tr>
<td>High-dose opioids as percentage regular opioids(^f)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>North Tyneside</td>
<td>8.9</td>
<td>40</td>
<td>September 2018</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Great Yarmouth and Waveney</td>
<td>8.7</td>
<td>33</td>
<td>May 2018</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Heywood, Middleton and Rochdale</td>
<td>8.9</td>
<td>33</td>
<td>September 2018</td>
<td></td>
</tr>
<tr>
<td>High-dose opioid items per 1000 patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Great Yarmouth and Waveney</td>
<td>2.0</td>
<td>39</td>
<td>August 2017</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Hastings and Rother</td>
<td>1.3</td>
<td>39</td>
<td>February 2018</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Heywood, Middleton and Rochdale</td>
<td>1.6</td>
<td>38</td>
<td>August 2017</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)CCG: clinical commissioning group.

\(^b\)Ranked top 3 CCGs exhibiting a reduction in each of the OpenPrescribing opioid measures (December 2014 to November 2019). The decile chart shows the prescription rate for the CCG as a thick red line; prescribing rates for all other CCGs are summarized using deciles (dotted blue lines) with the median highlighted (thick dashed blue line). Note, y-axis scales differ.

\(^c\)The absolute change is the difference between the starting value and final value during the detected change period.

\(^d\)The relative change gives the difference as a percentage of the starting value.

\(^e\)OME: oral morphine equivalence.

\(^f\)This measure is calculated as a percentage, so the absolute change represents the percentage points difference.

Changes for Practices

Table 3 illustrates the practices that exhibited the biggest change in each of the 3 OpenPrescribing measures over the study period, detailing the proportion of change and the month in which the change started. Note that these practices meet the criteria for identification as described in the “Statistical Methods” section, that is, their prescribing rate immediately before the reduction was in the top 150 (20%) practices.

The practice time series (Table 3) are noticeably different from those of the CCGs (Table 2), the magnitude of the measured changes is larger, and the slope of the time series is much steeper for practices. In the case of the regular high-dose opioids as a percentage of all opioids, all 3 practices are seen to completely eliminate all regular high-dose opioids for several months; similarly, very low values are observed for the top 3 practices with regards to reductions in high-dose opioid items per 1000 patients.
Table 3. Automatically detected changes across 3 measures of opioid prescribing in a retrospective prescribing database study of practices in England.a

<table>
<thead>
<tr>
<th>Measure and rank</th>
<th>Practice</th>
<th>Absolute change detected (difference)b</th>
<th>Proportional change (%)c</th>
<th>Month when change was detected</th>
<th>Chart</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total OMEd per 1000 patients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Practice A (Manchester CCG)</td>
<td>76,161</td>
<td>74</td>
<td>June 2018</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Practice B (Manchester CCG)</td>
<td>38,574</td>
<td>62</td>
<td>September 2018</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Practice C (West Cheshire CCG)</td>
<td>88,109</td>
<td>61</td>
<td>February 2017</td>
<td></td>
</tr>
<tr>
<td><strong>High-dose opioids as percentage regular opioidsf</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Practice D (City and Hackney CCG)</td>
<td>44</td>
<td>99</td>
<td>August 2018</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Practice E (Harrow CCG)</td>
<td>52</td>
<td>99</td>
<td>May 2017</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Practice F (Ealing CCG)</td>
<td>95</td>
<td>99</td>
<td>March 2016</td>
<td></td>
</tr>
<tr>
<td><strong>High-dose opioid items per 1000 patients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Practice G (Portsmouth CCG)</td>
<td>3.6</td>
<td>97</td>
<td>August 2018</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Practice H (Coventry and Rugby CCG)</td>
<td>4.8</td>
<td>97</td>
<td>February 2018</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Practice I (Salford CCG)</td>
<td>4.8</td>
<td>95</td>
<td>February 2018</td>
<td></td>
</tr>
</tbody>
</table>

aRanked top 3 practices exhibiting a reduction in each of the OpenPrescribing opioid measures (December 2014 to November 2019). The decile chart shows the prescription rate for the practice as a thick red line; prescribing rates for all other practices are summarized using deciles (dotted blue lines) with the median highlighted (thick dashed blue line). Note, y-axis scales differ.
bThe absolute change is the difference between the starting value and final value during the detected change period.
cThe relative change gives the difference as a percentage of the starting value.
dOME: oral morphine equivalence.
eCCG: clinical commissioning group.
fThis measure is calculated as a percentage, so the absolute change represents the percentage points difference.

Discussion

Main Findings

We have identified significant reductions in 3 measures of opioid prescribing using a data-driven approach in over 7000 practices across 191 CCGs in England (Table 1). These organizations have then been ranked by the magnitude of reduction to identify where the largest reductions have been realized. The top-ranked CCGs exhibit a slow and gradual reduction in opioid use (Tables 1 and 2); by contrast, the top-ranked practices exhibit rapid and sudden reductions over a few months (Tables 1 and 3). Opioid prescribing and treatment of pain more broadly can be complex, but our findings illustrate that some CCGs and practices appear to have significantly reduced their prescribing of opioids over the study period, more so than many of their peers.

Findings in Context

The PHE review identified evidence of tentative progress in reducing opioid prescribing between 2015 or 2016 and 2017 or 2018 [10]. Our analysis includes and extends this time period and finds evidence that some organizations may be driving this tentative progress more than others (eg, the CCGs reported in Table 2).

We do have evidence that 1 of the organizations that has emerged as a potential candidate by our methodology is a genuine example of improved performance. Between 2017 and 2019, Great Yarmouth and Waveney designed and implemented an extensive program of opioid reduction interventions, including target trajectories for improvement; incentive schemes for clinicians; dialogue with practice pharmacists, patient groups, and relevant clinical groups (eg, prescribing leads and pain management teams); new patient information materials; collecting case studies for discussion; and associated press and social media to raise awareness. While this CCG still exhibits high levels of opioid prescribing, rates have reduced significantly, with the organization being recognized for this progress nationally [23]. Our methodology ranked Great Yarmouth and Waveney as first (reduction of 39% [absolute reduction 2.0] starting in August 2017) for high-dose opioid prescribing per 1000 patients and second (reduction of 33%...
implement this best practice can systematically identify organizations that may have already implemented effective interventions. However, it is critical that policy makers undertake further investigations for reasons outlined in the limitations. This tool offers an opportunity to reduce the resources needed to identify best practices. Similarly, local medicines optimization teams may wish to use data and tools like this to identify peers across the country who have already delivered successful interventions to inform local initiatives. Further quantitative research is also possible from the data set, for example, drilling down by opioid type and monitoring the impact of any local or national interventions or policies.

Strengths and Limitations

The national prescribing data used for this analysis, being collected for reimbursement purposes, are highly complete and accurate. We have taken into account most (6543/7458, 87.7%) typical primary care practices in England, thereby minimizing the risk of biased sampling. Executing this analysis in an existing, open platform such as OpenPrescribing ensures accountability and transparency—both identified as priorities in the PHE report [10]; by default, all code in this study, from data curation to completed output, is shared openly on GitHub and the Python Package Index. Furthermore, there exists a robust and tested framework with which relevant new measures can be introduced or existing measures can be amended as required in order to respond to any evolving change in tackling opioid dependency and abuse. Our use of OME conversion permits the reporting of trends for opioid medicines overall while accounting for variation in strength.

We also note some limitations. First, the prescribing data set does not include secondary care prescriptions as this was unavailable at the time of the study [32], and as such, the opioid measures implemented here may underestimate the extent of opioid prescribing nationally, although financial data would indicate that the vast majority of analgesics (British National Formulary (BNF) section 4.7, which includes the BNF subsection 4.7.2 opioid analgesics) are prescribed in primary care [33]. Second, we acknowledge that practice-level time-series data, in particular, could be significantly impacted by local circumstances, including low patient numbers, a change in patient population, a change to prescription frequency (eg, from weekly to monthly scripts), or a shift in responsibility of opioid prescribing (eg, from primary to secondary care) and, therefore, that an apparent reduction in any opioid measure may not be due to a successful intervention. For example, practice G (Table 3) rapidly increased their high-dose opioid items per 1000 patients in 2016 followed by a similar rapid reduction 2 years later; this could be due to a change to daily prescribing as can be clinically justified for some patients. While we acknowledge these limitations, it is important to note that the intention of this methodology was always to rank or prioritize organizations for further investigation, rather than definitively ascribe reductions in opioid prescribing to successful interventions.

Conclusions

We have demonstrated that data-driven approaches to detect substantial changes in time-series data have potential value in the context of opioid prescribing. We have been able to rank organizations with regards to the extent of opioid prescribing reduction; organizations occupying the top of that list show large drops that warrant further qualitative investigation and could be indicative of success in tackling an important public health concern.

Should this further qualitative research reveal that reductions have been driven by well-designed and well-implemented interventions, methods of best practice will have been identified using an unbiased, evidence-based approach. The organizations found to be implementing this best practice may have valuable insights, approaches, and policies to share regarding how positive change can be achieved elsewhere. It also demonstrates, particularly in the most robust and gradual change observed among CCGs, that positive change is possible and, therefore, that continued and wider success in reducing opioid prescribing is dependent, at least in part, on closing the implementation gap.
Acknowledgments
We are grateful to wider National Health Service (NHS) colleagues for discussions that have informed our work on this topic. This project is funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RIPPB) Programme (PB-PG-0418-20036). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. Funders had no role in the study design, collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Data Availability
The data sets generated and analyzed during this study are available in the “opioids-change-detection-notebook” repository [21]. All our methods and code can also be found in this repository [21]. The underlying data are available via the National Health Service Business Services Authority (NHSSBA) [14].

Authors’ Contributions
AJW, HJC, BG, and BM contributed to the conceptualization. PI, DE, and SB contributed to the data curation. LEMH, AJW, and BM performed the formal analysis. AJW, HJC, and BG contributed to the funding acquisition. LEMH, AJW, HJC, RC, and BM performed the investigation. LEMH, AJW, HJC, RC, and BM contributed to the methodology. PI, DE, and SB contributed to the resources. FP, PI, DE, and SB contributed to the software. BG performed the supervision. LEMH, AJW, HJC, and BM did the visualization. LEMH, AJW, HJC, and BM contributed to writing—original draft. All authors contributed to writing—review and editing.

Conflicts of Interest
All authors have completed the International Committee of Medical Journal Editors (ICMJE) uniform disclosure form and declare the following: BG has received research funding from the Bennett Foundation, the Laura and John Arnold Foundation, the National Health Service (NHS) National Institute for Health Research (NIHR), the NIHR School of Primary Care Research, NHS England, the NIHR Oxford Biomedical Research Centre, the Mohn-Westlake Foundation, NIHR Applied Research Collaboration Oxford and Thames Valley, the Wellcome Trust, the Good Thinking Foundation, Health Data Research UK, the Health Foundation, the World Health Organization, UK Research and Innovation Medical Research Council (UKRI MRC), Asthma UK, the British Lung Foundation, and the Longitudinal Health and Wellbeing strand of the National Core Studies programme; he has previously been a nonexecutive director at NHS Digital; he also receives personal income from speaking and writing for lay audiences on the misuse of science. BMK is also employed by NHS England working on medicines policy and clinical lead for primary care medicines data.

References
8. Opioids aware. Faculty of Pain Medicine of the Royal College of Anaesthetists. URL: https://fpm.ac.uk/opioids-aware [accessed 2022-02-18]
Abbreviations

BNF: British National Formulary
CCG: clinical commissioning group
GP: general practice
NHS: National Health Service
NHSBSA: National Health Service Business Services Authority
OME: oral morphine equivalence
PHE: Public Health England

Edited by A Mavragani, T Sanchez; submitted 31.07.23; peer-reviewed by C Tolley, S Leitch; comments to author 07.11.23; revised version received 23.11.23; accepted 12.02.24; published 05.06.24.

Please cite as:
Data-Driven Identification of Potentially Successful Intervention Implementations Using 5 Years of Opioid Prescribing Data: Retrospective Database Study
J Miranda, Public Health Surveill 2024;10:e51323
URL: https://publichealth.jmir.org/2024/1/e51323
doi: 10.2196/51323
PMID: 38838327

©Lisa EM Hopcroft, Helen J Curtis, Richard Croker, Felix Pretis, Peter Inglesby, David Evans, Sebastian Bacon, Ben Goldacre, Alex J Walker, Brian MacKenna. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 05.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: Adverse events associated with vaccination have been evaluated by epidemiological studies and more recently have gained additional attention with the emergency use authorization of several COVID-19 vaccines. As part of its responsibility to conduct postmarket surveillance, the US Food and Drug Administration continues to monitor several adverse events of special interest (AESIs) to ensure vaccine safety, including for COVID-19.

Objective: This study is part of the Biologics Effectiveness and Safety Initiative, which aims to improve the Food and Drug Administration’s postmarket surveillance capabilities while minimizing public burden. This study aimed to enhance active surveillance efforts through a rules-based, computable phenotype algorithm to identify 5 AESIs being monitored by the Center for Disease Control and Prevention for COVID-19 or other vaccines: anaphylaxis, Guillain-Barré syndrome, myocarditis/pericarditis, thrombosis with thrombocytopenia syndrome, and febrile seizure. This study examined whether these phenotypes have sufficiently high positive predictive value (PPV) to ensure that the cases selected for surveillance are reasonably likely to be a postbiologic adverse event. This allows patient privacy, and security concerns for the data sharing of patients who had nonadverse events can be properly accounted for when evaluating the cost-benefit aspect of our approach.

Methods: AESI phenotype algorithms were developed to apply to electronic health record data at health provider organizations across the country by querying for standard and interoperable codes. The codes queried in the rules represent symptoms, diagnoses, or treatments of the AESI sourced from published case definitions and input from clinicians. To validate the performance of the algorithms, we applied them to electronic health record data from a US academic health system and provided a sample of cases for clinicians to evaluate. Performance was assessed using PPV.

Results: With a PPV of 93.3%, our anaphylaxis algorithm performed the best. The PPVs for our febrile seizure, myocarditis/pericarditis, thrombocytopenia syndrome, and Guillain-Barré syndrome algorithms were 89%, 83.5%, 70.2%, and 47.2%, respectively.

Conclusions: Given our algorithm design and performance, our results support continued research into using interoperable algorithms for widespread AESI postmarket detection.

(JMIR Public Health Surveill 2024;10:e49811) doi:10.2196/49811
KEYWORDS
adverse event; vaccine safety; computable phenotype; postmarket surveillance system; real-world data; validation study; Food and Drug Administration; electronic health records; COVID-19 vaccine

Introduction

Background

The US Food and Drug Administration (FDA) Center for Biologics Evaluation and Research (CBER) is responsible for ensuring the safety, purity, potency, and effectiveness of biological products. This includes vaccines; allergens; blood and blood products; and cell, tissue, and gene therapies for the prevention, diagnosis, and treatment of human diseases, conditions, or injuries [1]. The FDA’s history of safety surveillance for vaccines includes the creation and monitoring of the Vaccine Adverse Event Reporting System (VAERS). VAERS, jointly administered by the FDA and the Centers for Disease Control and Prevention (CDC), accepts spontaneous reports of suspected vaccine adverse events (AEs) after administration of any vaccine licensed in the United States.

VAERS has been successfully used as an early warning system to identify rare AEs; however, it has limitations. VAERS is a passive surveillance system that relies on individuals, patients, and clinical staff to send in reports, as opposed to automatically collecting them based on clinical data. This can lead to undercounting AEs. In addition, a causal relationship cannot be established using information from VAERS reports alone [2]. Because of VAERS’s limitations, more robust data systems are needed to enhance AE detection. These systems would be especially important for detecting the most severe AEs that require medical attention so that the FDA and CDC can offer guidance for these potentially life-threatening events and ensure that product labeling reflects known risks.

To address this gap, CBER established the Biologics Effectiveness and Safety Initiative (BEST) Initiative in 2017 to build data assets, analytics, and infrastructure for an active, large-scale, efficient postmarket surveillance system that can evaluate the safety and effectiveness of biologic products and develop innovative methods [3]. The BEST system is a collection of real-world data (RWD) sources: data related to patient health status and the delivery of health care that are routinely collected from several sources, such as electronic health record (EHR) or claims data [4]. EHR databases, specifically, are a rich source of information. They include data such as clinical notes, which can help address the limitations of VAERS. They also include entire populations of patients to identify whether cases are underreported. In addition, they may include patients’ entire clinical history, which can help establish a causal relationship for an AE. BEST has reached agreements with a limited number of foundational data partners. Access to these data partnerships does not fully address the possible undercounting of AEs of special interest (AESIs). However, these partnerships allow accelerated development and testing of AESI detection algorithms.

BEST is currently researching a system of distributed computable phenotype algorithms that could be applied at scale to many or all EHR systems across the United States to semiautomatically detect and report potential AESIs from RWD. Such a system could increase the speed and scope of AE surveillance beyond what is currently available to public health agencies through data partner agreements. To be candidate phenotypes for distributed surveillance use, the phenotypes need to identify probable AEs and avoid false detections. This reflects the need to balance the correct detection of AESIs with the protection of privacy and the reduction of burden on health provider systems. For the wider population of health providers to consider deploying such detection algorithms, these phenotype algorithms need to have reasonably high performance (measured by positive predictive value [PPV]) to ensure that the cases identified as AEs are likely to be verifiable cases with the outcome of interest. Toward this goal, the computable phenotypes in this study focus on existing EHR data reflecting a detected AE, which are reportable events for public health purposes. The algorithmic identification of undetected AEs or AEs that were not coded properly is beyond the scope of this study. Such research must include data from patients who had no AEs to fully evaluate the performance of a computable phenotype algorithm. Although scientifically desirable in the long term, the inclusion of non-AE cases falls outside of initial goals for a distributed surveillance system, which is assessing performance (measured by PPV) of the phenotypes for wide-scale surveillance purposes. The goal of distributing the phenotypes also poses limitations on designing the algorithms. Specifically, the components and complexity of the underlying algorithms need to take into account the current EHR standards and technology because they must be deployable and executable across EHR databases without imposing large overhead on health provider systems. If the phenotypes have sufficient PPV and are sufficiently easy to implement at health provider sites, the FDA could share the phenotypes to detect AESIs following vaccination in EHRs across the country, which could then be reported to the FDA for further review. The ability to detect AESIs using RWD could create an active surveillance system that enhances overall vaccine safety and helps make recommendations to minimize risks for postvaccination AESIs. The implementation of algorithmic detection and automated reporting of AESIs found in RWD has been shown to increase the odds of submitting a VAERS report by >30 times the preimplementation rate [5].

Objective

Although there is a history of studies around postvaccination AESIs, including those for influenza [6-8] and COVID-19 vaccines [9-13], there has been an increased interest in the analysis of vaccine safety and surveillance since the emergency use authorization (EUA) of 3 COVID-19 vaccines in the United States (Pfizer-BioNTech, Moderna, and Novavax) and their subsequent boosters (eg, bivalent boosters). The FDA hopes to contribute to this research through the development and performance validation of phenotypes for 5 postvaccination AESIs to identify potential vaccine safety events within EHR databases for this study. The 5 AESIs chosen include
myocarditis/pericarditis, anaphylaxis, Guillain-Barré syndrome (GBS), intracranial or intra-abdominal thrombosis with thrombocytopenia syndrome (TTS), and febrile seizure. These AESIs were chosen because they are documented priorities of the CDC’s vaccine surveillance [14] for COVID-19 vaccine safety. In addition, several of these AESIs (anaphylaxis, GBS, and febrile seizure) are found following exposure to other vaccines, such as influenza; shingles; pneumococcal conjugate; and measles, mumps, and rubella. This study describes the methods to develop and validate these 5 computable phenotype algorithms on an EHR database and the validation results. It is part of the FDA’s efforts to improve postmarket surveillance and is valuable for public awareness, safety, and transparency.

Methods

Ethical Considerations

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. This study was part of the Sentinel activities conducted by the FDA as part of its postmarket surveillance duties. The Office of Human Research Protection (OHRP) in Health and Human Services (HHS) determined that the studies done under the Sentinel programs are not subject to regulation (45 CFR part 46) administered by OHRP. Written informed consent to participate in this study was not required from the participants or the participants’ legal guardians or next of kin in accordance with the national legislation and the institutional requirements.

Computable Phenotype Development

Overview

In total, 5 AESIs were selected to develop computable phenotypes for our validation study. The study’s main focus was detecting COVID-19 vaccine–related AESIs; therefore, we selected AESIs that the CDC specifically identified for monitoring after COVID-19 vaccination [14] or AESIs that have been reported for some subpopulations [15]. Given the uncertainty about the future use of COVID-19 seasonal boosters, the FDA also wanted to ensure that the AESIs selected had broad applicability to the safety surveillance of other widely used vaccines such as influenza; shingles; pneumococcal conjugate; diphtheria-tetanus-pertussis; and measles, mumps, and rubella. Three of our 5 selections met those criteria given in the CDC’s documented monitoring of anaphylaxis [16], GBS [17], and febrile seizures [18] for at least one of the vaccines listed.

The phenotype algorithms were designed to be relatively simple and interoperable so that any new health care organization’s IT department could translate and run them on their EHR database. They were built to query only structured data for interoperable, standard codes, such as Logical Observation Identifiers Names and Codes, Systematized Nomenclature of Medicine Clinical Terms, and RxNorm, so that the algorithm can be generalized or translated across different EHR systems. Historically, this has been a challenge for developing algorithms, since EHR databases often contain their own local code systems specific to the EHR vendor. For example, for this effort, we worked with the study partner to map Cerner Multum medication and observation codes to standard RxNorm and Logical Observation Identifiers Names and Codes, respectively.

Recent regulation now requires each EHR database to have an application programming interface (API) endpoint that translates any EHR data and many of the EHR’s proprietary codes to the United States Core Data for Interoperability (USCDI) implementation of the Fast Healthcare Interoperable Resources (FHIR) specification [19]. This specification requires the use of interoperable, published code lists [20] (Table 1). These code systems cover almost all clinical events for the detection of AEs, such as medical diagnoses, medication prescriptions, laboratory tests or vital signs taken, and procedures performed. These APIs currently focus on supporting use cases where a single patient’s data are queried as opposed to aggregate searches across patients; therefore, we were unable to use them to identify the cohort that our phenotype would select. We were, however, able to use the FHIR API endpoints to pull data for each patient in our validation samples so that the participating clinicians could have data with the standard, interoperable code sets for their review.
Table 1. AESI<sup>a</sup> case definitions and descriptions.

<table>
<thead>
<tr>
<th>AESI</th>
<th>Description</th>
<th>Case definition reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AESI: adverse event of special interest.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
diagnosis care settings values were grouped into inpatient, outpatient, or emergency care setting types. Care setting was used to filter out diagnosis codes made during encounters with care settings unlikely to have the specific AESI diagnosis in the phenotype. The included care settings are defined by case definition and clinician input.

**Clean Window**

Next, a clean window (ie, a period before the coded diagnosis identified in step 1) is checked to ensure that the target diagnosis is the first known diagnosis of its type. This prevents the inclusion of historical or ongoing conditions. For all algorithms in this paper, the clean window is defined by all historical patient data in our data set. To make sure that all patient cases had at least a 1-year clean window, we pulled an additional historical year of data from our data partner before the study period. Cases where there were multiple occurrences of an AESI diagnosis suggested possible evidence of a chronic condition unrelated to vaccine exposure and thus were excluded.

**Condition Window**

Finally, the algorithm searches for sufficient supportive evidence within a condition window. The condition window is defined around the AESI diagnosis date and includes the entire medical encounter period when the condition was diagnosed, as well as 2 days before and 10 days after a condition is diagnosed. Clinical subject matter experts defined condition windows as the timeframe around a diagnosis that supportive evidence would likely present itself in the medical record.

**Supporting Evidence**

Within the condition window period, the algorithm may filter cases based on supporting evidence of an AESI. This filter looks for either laboratory test results found in observations, AESI treatment procedures, AESI treatment medications, or procedure or a combination of the 3 supporting evidence with a code that matches a code on to the phenotypes’ concepts code lists. These code lists in the lists aim to include all medical codes that could represent a particular concept, such as administration of epinephrine for an anaphylactic reaction. This AESI supporting evidence filter was applied to all phenotypes except for our febrile seizure AESI phenotype because a review of existing research [32] showed febrile seizure algorithms, in general, had the highest PPV among the selected AESIs. The concepts to build code lists for the supporting evidence were identified using various types of clinical data. Following this, we prioritized improving specificity in the other AESI phenotypes by including filters requiring additional supporting evidence [32-35] and our clinician’s input.

**Vaccine Exposure**

In real-world operation, the algorithm would also include a vaccine exposure and risk window or a period surrounding vaccination in which diagnoses are searched. For the study’s purposes of having sufficient volume and statistical power to estimate operating characteristics of the algorithm, these exposure rules were not included.

**Additional Details**

Ideally, to assess whether these algorithms generalize to other sites, we would have a multisite validation study. Because of the high cost of data agreements, however, we only had data available for a single EHR site. To avoid overfitting and ungeneralizable results, we designed our algorithm development methods to only use our EHR data as a validation set and not use any of it to train, develop, or fine-tune the algorithm. While this does not remove the need for additional external validation, it reduces the likelihood of finding ungeneralizable results. To identify what medical concepts the algorithm should use as evidence, clinicians identified observations, medications, conditions, and procedure concepts from the AESI’s case definition, their relevant clinical experience, or other research from their literature review. A brief description of the AESI and the reference of the case definition used is captured in Table 1, and additional information on the case definition is saved in Table S2 in Multimedia Appendix 1 [21-23,25-27].

An analyst completed a text search for a list of terms for these identified concepts, a list of which is captured in Table S2 in Multimedia Appendix 1, to build the code lists of relevant codes from selected interoperable coding libraries (Textbox 1) [21-23,25-27]. This was accomplished by searching the open-sourced OHDSI Observational Medical Outcomes Partnership concepts table and ATLAS tool (OHDSI community) [36], which is a collection of thousands of interoperable codes and their definitions and descriptions. The table was searched for any definition or description that matched the identified concept for the interoperable code systems that we listed in Textbox 1 and then was reviewed by a clinician for their suitability for the algorithm.

**Textbox 1.** Codes used for each type of clinical data.

### Clinical data and interoperable code lists used

- **Diagnosis**: International Classification of Diseases, Tenth Revision, Clinical Modification, Systematized Nomenclature of Medicine Clinical Terms
- **Medication or immunization**: National Drug Code, RxNorm
- **Procedures**: Current Procedural Terminology, International Classification of Diseases, 10th Revision Procedure Coding System
- **Observations**: Logical Observation Identifiers Names and Codes

The immunization and the diagnosis International Classification of Diseases, 10th Revision, clinical modification (ICD-10) and Systematized Nomenclature of Medicine Clinical Terms code lists have been published on the Value Set Authority Center [37], and the additional observation, medication, and evidence code lists may be added in the future after this study is published.
For a surveillance use case, the algorithms need be run regularly (e.g., daily or weekly) to collect batches of historical cases once all the data are available (as opposed to a real-time implementation to collect cases as they are happening). Because the algorithms were created to prioritize simplicity and interoperability rather than maximize total performance (e.g., metrics beyond PPV such as sensitivity and, negative predictive power, etc), this study aimed for improved performance (measured by PPV) to existing AESI claims-based algorithms. Given our knowledge of how some crucial distinguishing information is part of unstructured clinical notes, which are not considered by the algorithms in this study, we expect further analysis is needed to improve accuracy [38,39]. Natural language processing techniques can improve algorithm performance but greatly increase the deployment complexity across health care organizations. Therefore, no natural language processing techniques were used for any phenotypes designed for this study.

**Study Period**

The study period spanned from January 1, 2018, through May 1, 2022, to ensure that the study’s data sampled patients both before and after the FDA issued the EUA and full licensure for COVID-19 vaccines. We also pulled at least 1 year of historical data for all patients; therefore, our data set includes historical information from January 1, 2017, to January 1, 2018, for all patients with medical encounters in the study period. Patients were included even if there were no clinical events in their historical period.

**Data**

The study population came from a single academic health system in the United States, with EHR medical encounter data from >2.6 million patients and >20.7 million medical encounters for the study period. Table 2 shows the demographic breakdown for age, gender, race, and ethnicity of this population.

The entire EHR population during the study period was eligible to be selected by one of our developed phenotype algorithms. There were no age-related, medical condition–related, or other exclusions on the population for the algorithm to select cases. Clinical data necessary to select and validate cases selected by the algorithm were provided to the study team through a series of EHR data extracts for all patients in the study period. The algorithm required the following clinical data categories:

- demographic
- encounter
- condition
- procedure
- medication
- observation
Table 2. Demographics of academic health system for study population (N=2,666,974).

<table>
<thead>
<tr>
<th>Category and demographic group</th>
<th>Patients, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>96,146 (3.6)</td>
</tr>
<tr>
<td>5-17</td>
<td>224,941 (8.4)</td>
</tr>
<tr>
<td>18-24</td>
<td>224,631 (8.4)</td>
</tr>
<tr>
<td>25-44</td>
<td>840,395 (31.5)</td>
</tr>
<tr>
<td>45-64</td>
<td>689,075 (25.8)</td>
</tr>
<tr>
<td>≥65</td>
<td>591,497 (22.2)</td>
</tr>
<tr>
<td>Missing</td>
<td>289 (0.01)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,167,374 (43.8)</td>
</tr>
<tr>
<td>Female</td>
<td>1,494,096 (56.1)</td>
</tr>
<tr>
<td>Missing</td>
<td>5504 (0.2)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>748,746 (28.1)</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>5834 (0.2)</td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>53,666 (2)</td>
</tr>
<tr>
<td>White</td>
<td>1,030,834 (38.7)</td>
</tr>
<tr>
<td>Other</td>
<td>198,265 (7.4)</td>
</tr>
<tr>
<td>Unknown</td>
<td>629,608 (23.6)</td>
</tr>
<tr>
<td>Declined to answer</td>
<td>21 (0)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>94,207 (3.5)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>1,866,561 (70)</td>
</tr>
<tr>
<td>Unknown</td>
<td>706,206 (26.5)</td>
</tr>
</tbody>
</table>

EHR data extracts were mapped and loaded into an OHDSI Observational Medical Outcomes Partnership database [40]. Medication, observation, and procedure data extracts were requested and loaded into the database only for patients who would not be disqualified by other algorithm criteria. For patients selected to be in the validation sample, these data along with the clinical data for allergies, immunizations, and clinical notes were pulled from the EHR’s FHIR API endpoints, patient by patient, using a custom Python script to loop through the patients in the sample. The data were loaded into a Health Level 7 API (HAPI) FHIR server. We only pulled FHIR data for cases not initially disqualified by the vaccination and diagnosis filters to avoid unnecessary large data transfers and storage. The algorithm flagged potential AESIs that met the specified criteria. Samples of these cases were sent to physicians for validation. **Validation Sample**

Once the algorithm identified cases, a random sample was drawn for each AESI for clinician adjudication. We used stratified sampling to ensure cases during pre– and post–COVID-19 EUA periods were represented (Figure 2). This was due to concerns regarding potential confounding introduced by the COVID-19 vaccines, when attention to possible AESIs or medical charting of AESIs may have shifted. Where possible for each AESI, 100 cases were sampled from the pre–COVID-19 EUA period and 35 from the post–COVID-19 EUA period. If there were <100 or <35 cases during these periods, respectively, the sample would contain all cases the algorithm selected. Febrile seizure was the exception, as we believe the COVID-19 vaccine EUA should not affect the algorithm’s performance because febrile seizure AEs are usually associated with pediatric populations, and the COVID-19 vaccine was not approved for these populations during the study period [27].
Figure 2. Study population CONSORT (Consolidated Standards of Reporting Trials) diagram. EUA: Emergency Use Authorization; GBS: Guillain-Barré syndrome; N/A: not applicable; TTS: thrombosis with thrombocytopenia syndrome.

**Clinical encounters**

*January 1, 2018 to May 1, 2022*

N=20.7 million

- **Cases with condition diagnosis**
  - Myocarditis/Pericarditis (n=1699)
  - Anaphylaxis (n=1332)
  - GBS (n=168)
  - Febrile Seizure (n=626)
  - TTS (n=413)

- **Excluded cases without supporting evidence**
  - Myocarditis/Pericarditis (n=504)
  - Anaphylaxis (n=802)
  - GBS (n=45)
  - Febrile Seizure (n=0)
  - TTS (n=18)

**Cases with supporting evidence**

- Myocarditis/Pericarditis (n=1195)
- Anaphylaxis (n=550)
- GBS (n=123)
- Febrile Seizure (n=626)
- TTS (n=395)

**Pre-COVID-19 vaccine EUA**

*January 1, 2017 to December 11, 2020*

- Myocarditis/Pericarditis (n=770)
- Anaphylaxis (n=366)
- GBS (n=75)
- Febrile Seizure (n=626)
- TTS (n=275)

**Post-COVID-19 vaccine EUA**

*December 11, 2020 to May 1, 2022*

- Myocarditis/Pericarditis (n=425)
- Anaphylaxis (n=184)
- GBS (n=48)
- Febrile Seizure (n=N/A)
- TTS (n=120)

**Pre-EUA cases validated**

- Myocarditis/Pericarditis (n=100)
- Anaphylaxis (n=100)
- GBS (n=75)
- Febrile Seizure (n=100)
- TTS (n=100)

**Post-EUA cases validated**

- Myocarditis/Pericarditis (n=53)
- Anaphylaxis (n=35)
- GBS (n=35)
- Febrile Seizure (n=N/A)
- TTS (n=35)

Case counts sampled in each period were based on the incidence of diagnosis code occurrence within each period, as well as the period covered. In addition, we added negative controls selected randomly from every encounter in the period to establish a baseline comparison for the case validation process. We included negative controls as a quality control step to reduce the chance of quality issues with the data and to review the methods our clinicians were following and not for the purpose of making inferences about the phenotypes’ performance for non-AE cases (eg, through metrics such as sensitivity, negative predictive power, or an overall metric for performance). This study did not focus on the algorithmic identification of undetected AEs or AEs that were not coded properly. The focus of this study was to determine the phenotypes’ PPV. Given the expense of clinicians’ time for validations and the rarity of the AESIs, there would be minimum benefit to this study to have a negative control sample large enough to draw strong inferences. Furthermore, negative case controls would not further validate the utility of the phenotypes as tools for identifying probable AESIs through distributed surveillance. We added 20 negative controls from the pre–COVID-19 EUA period and 7 from the post–COVID-19 EUA period. Physicians were blinded to which cases were controls and which were not.

**Chart Review Process**

The sample of cases used to validate the algorithm was loaded into a chart review tool for clinician review. This allowed the clinicians to sort through the clinical information for a case and record the determination. Each case was assigned to 2 clinicians for review. The clinical validation used a patient’s full clinical history, which included EHR data, including all clinical notes for each case. The full EHR data used for clinician review included data unused by the detection algorithm described in the Computable Phenotype Development section, including different types of data (eg, allergies and clinical notes) and data filtered out (eg, admitting diagnosis and encounters with different care settings).

For each case, the clinician evaluated whether the clinical data evidence met the specified case definition criteria. Relevant patient data for the case window were available and presented to the clinicians in an easy-to-use, browser-based tool with a custom user interface. In the tool, clinicians were able to group
items by type, search across all items and text, and request additional chart data to expand the window and access any available historical patient data, if desired.

All suspected AEs were validated using published case definitions [21-23,25-27] according to the levels of diagnostic certainty: level 1 (definite), level 2 (probable), and level 3 (possible). If a case did not meet one of the levels in the case definition, it was assigned as level 4 (doubtful) or level 5 (ruled out). “Ruled out” is distinct from “doubtful” in that “ruled out” cases have definitive evidence disqualifying them from being a correct diagnosis. If a case was determined to be “definite” or “probable,” it was considered a positive case of the AESI.

In the event of a disagreement between a positive and negative clinical review, a third clinician made a final determination by reviewing the case EHR data. If the clinicians found the structured or unstructured EHR data was insufficient, they marked this in their review by creating a level 3 (possible, insufficient evidence) designation, where an AESI could have occurred, but where there was not enough documentation to fulfill the requirements of the case definition.

Statistical Analysis

PPV of Algorithms

Each algorithm’s PPV was the proportion of positive AEs the algorithm identified that were confirmed by clinical adjudication. PPVs were calculated for each AESI overall, as well as stratified by pre– and post–COVID-19 EUA periods and care setting (inpatient, emergency department, or outpatient). Sensitivity analyses were performed to evaluate the impact of medication use, different case definitions, and levels of evidence. PPVs were calculated in 2 different ways for each AESI algorithm. The first PPV calculated removed all possible cases with insufficient evidence from the denominator (cases labeled “definite” and “probable”/total cases minus any labeled “possible, insufficient evidence” by clinicians). PPV was then calculated with the cases with insufficient evidence added back into the denominator (cases labeled “definite” and “probable”/total cases). Reporting both PPV calculations can help with understanding the performance for different algorithm uses. Algorithm performance should ideally be compared with past literature of detection algorithms for the same AESI.

CI Values

Because PPV is a binominal proportion, we calculated CIs for the PPV using the Agresti-Coull interval [41], which is the recommended method for estimating accurate CIs for binomial proportions such as PPV [42].

Interrater Reliability

Interrater reliability was used to measure the extent to which 2 physicians agreed in their AESI assessment. It was calculated using Cohen κ between the first 2 reviewers. Cohen κ measures the agreement between 2 raters classifying instances into mutually exclusive groups [43].

Stratification Analysis and Sensitivity Analysis

After validation was completed, we conducted a stratification and sensitivity analysis. We selected 2 stratification variables that could reasonably impact the generalizability of the results. First, we stratified the data by pre- and post-EUA date to confirm that the algorithm behavior did not change for AESIs after the COVID-19 vaccine was approved and administered to a large portion of the population. Ideally, the algorithms would perform consistently across these eras, but there are multiple factors that could impact the performance over these time periods. We also stratified the data by the care setting of the AE diagnosis, given that care settings may be associated with varying EHR data elements (eg, emergency departments compared with inpatient settings). Algorithm performance was computed using PPV within each stratum.

We also completed a post hoc sensitivity analysis where we investigated whether the algorithm could be improved, as measured by PPV, through small changes to it or by updating the process for evaluation. These changes were based on insights from clinicians or data analysts reviewing validation results, so results may not generalize to other data sets. However, we did attempt to limit our analysis to decisions that could have been feasibly made without postvalidation insights. The changes to the algorithms were either removing medications, observations, procedures, or diagnosis codes that are not specific enough to the AESI in question or adding logic to further filter out cases by requiring more supporting evidence (Table 3). The stratification or sensitivity analyses are meant as exploratory analyses to prompt additional research, but subgroups often have too small a sample size that have narrow enough CIs for meaningful results.

We also completed a sensitivity analysis on the GBS algorithm to calculate the PPV if we relaxed some of the specific case definition evaluation criteria and if more general evidence was available. We found that the 2 pieces of evidence that the case definition required were often missing in the chart review tool: lack of cerebrospinal fluid (CSF) white blood cell (WBC) count in cases of elevated CSF protein and limited or inconsistent documentation of diminished or absent reflexes. In some of these cases, we saw evidence that a neurologist was consulted and felt there was strong suspicion of GBS despite the missing documentation for these tests. This could be explained by 2 mechanisms.

First, and most likely, this could be due to data loss during the delivery or translation of EHR data to our chart review tool. Because we did not have direct access to the data, our process for obtaining, translating to different common data models or standards, and presenting the data to clinicians using the chart review tool could cause the data for these tests to be incorrectly mapped.
Table 3. Total list of all sensitivity analyses for each adverse event of special interest (AESI).

<table>
<thead>
<tr>
<th>AESI</th>
<th>Data type</th>
<th>Sensitivity analysis</th>
<th>Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocarditis/pericarditis</td>
<td>Medication</td>
<td>Removal of NSAIDs(^a) from our list of qualifying medication supporting evidence</td>
<td>NSAIDs are medications that can be used to treat many different conditions besides myocarditis and pericarditis.</td>
</tr>
<tr>
<td>Myocarditis/pericarditis</td>
<td>Diagnostic code</td>
<td>Stratification by diagnostic code (myocarditis vs pericarditis)</td>
<td>Diagnostic criteria differ for these related conditions and may lead to different performance.</td>
</tr>
<tr>
<td>GBS(^b)</td>
<td>Medication</td>
<td>Removal of gabapentin from our list of qualifying medication supporting evidence</td>
<td>Gabapentin was originally used as supporting evidence of a GBS episode due to its use for nerve pain associated with GBS events [44]. However, it is also used for a variety of other conditions with neuropathic pain and is not specific to GBS.</td>
</tr>
<tr>
<td>GBS</td>
<td>Case definition</td>
<td>Update case definition criteria to allow for a case to be validated as positive if there is a missing documentation for absent or diminished reflexes in the weak limbs, CSF(^c) WBC(^d) count with neurology consult, or clinical note indicating evidence of the test result of GBS more generally</td>
<td>Documentation required for definite or probable GBS as defined by the case definition diagnosis was often missing from our data set due to failure to capture in EHR(^e) or failure to translate to our data set and can be supplemented by an expert’s judgment (eg, a neurologist).</td>
</tr>
<tr>
<td>Febrile seizure</td>
<td>Medication</td>
<td>Addition of medications used to treat fever without suggested evidence, but we believed adding suggested evidence could improve PPV(^f).</td>
<td>The original febrile seizure algorithm did not filter out cases without suggested evidence, but we believed adding suggested evidence could improve PPV.</td>
</tr>
<tr>
<td>Febrile seizure</td>
<td>Observation</td>
<td>Addition of observation of clinician describing the symptoms of seizure activity</td>
<td>The original febrile seizure algorithm did not filter out cases without suggested evidence, but we believed adding suggested evidence could improve PPV.</td>
</tr>
<tr>
<td>TTS(^g)</td>
<td>Diagnostic code</td>
<td>Stratification by most prevalent diagnostic code I81 versus all other codes</td>
<td>Diagnostic criteria differ for these related conditions and may lead to different performance.</td>
</tr>
</tbody>
</table>

\(^{a}\)NSAID: nonsteroidal anti-inflammatory drug.  
\(^{b}\)GBS: Guillain-Barré syndrome.  
\(^{c}\)CSF: cerebrospinal fluid.  
\(^{d}\)WBC: white blood cell.  
\(^{e}\)EHR: electronic health record.  
\(^{f}\)PPV: positive predictive value.  
\(^{g}\)TTS: thrombosis with thrombocytopenia syndrome.

Second, case definition requirements for GBS are extremely strict, and physicians in this study believed that some of these might have represented valid GBS cases while not meeting every requirement. For example, several of the cases with missing CSF WBC count did mention cytoidalbuminologic dissociation (or similar); in the presence of such a clinical statement, we might infer that CSF WBC count was performed and acceptable to meet the case definition criteria despite a missing test result.

Furthermore, in cases where a neurologist felt strongly that GBS was a likely diagnosis, along with other supporting evidence, it may be acceptable to rely on documented progressive and significant muscle weakness, especially with conflicting reflex findings. In these instances, we placed more weight on the clinician review (which may account for any unforeseen difficulties in data processing and the strictness of the case definition), not relying solely on the available (nonmissing) data types of the algorithm for assigning case diagnostic certainty.

Results

Population Sample

Figure 2 illustrates the identification of the study populations and validation sample. From the study population of 20.7 million medical encounters for 2,666,974 patients over the study period, the algorithm selected 1195 (0.04%) cases of myocarditis/pericarditis, 550 (0.02%) of anaphylaxis, 123 (0.005%) of GBS, 626 (0.02%) of febrile seizure, and 395 (0.01%) of TTS. Of these patient cases (n=2,666,974), a stratified, random sample of 135 (0.01%) cases each was selected from myocarditis/pericarditis, 135 (0.01%) from anaphylaxis, and 395 (0.01%) of TTS. Of these patient cases (n=2,666,974), a stratified, random sample of 135 (0.01%) cases each was selected from myocarditis/pericarditis, 135 (0.01%) from anaphylaxis, and 395 (0.01%) from TTS populations. All 75 pre-EUA cases of GBS and a random sample of 35 post-EUA cases were selected to be validated. A random selection of 100 cases from the pre-EUA period were sampled to validate febrile seizure. An additional 27 negative control cases were sampled for each algorithm from the roughly 20.7 million medical encounters not selected by the algorithm in our study period. In total, 20 of these cases were sampled from the period before the COVID-19 vaccine EUA, and the remaining 7 came from the period after the EUA.
Overall PPV and Interrater Reliability Results

Table 4 presents algorithm performance measured by PPV for each of the 5 AESIs using cases that had sufficient evidence and all cases (ie, including cases unable to be confirmed as positive by clinicians due to insufficient evidence). Counts for the number of cases included in each PPV calculation can be found in Table S3 in Multimedia Appendix 1 [21-23,25-27].

Overall PPVs, when removing all cases with insufficient evidence, were highest for anaphylaxis (93.3%, 95% CI 86.4%-97%) and febrile seizure (89%, 95% CI 80%-94.4%), followed by myocarditis/pericarditis (83.5%, 95% CI 74.9%-89.6%) and TTS at unusual sites (70.2%, 95% CI 61.4%-77.6%). The lowest was for GBS (47.2%, 95% CI 35.8%-58.9%). All negative control cases across the 5 phenotypes were correctly classified by the algorithms.

The PPV results from the chart reviews of the validation sample for each AESI are reported for all cases as well as for only cases with sufficient evidence to make a clear by chart reviewers. The frequencies and percentages for insufficient evidence are presented with the stratification results in Table 5. The interrater reliability scores for clinician chart reviews all showed substantial agreement between the clinicians (Table 6). Interrater reliability, measured by Cohen κ, suggests substantial reliability when the value is >0.61, with many similar texts recommending a higher threshold of 0.80 [43].

Table 4. Total validation positive predictive value (PPV) results.

<table>
<thead>
<tr>
<th>AESI and metric</th>
<th>Detected cases, PPV % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Myocarditis/pericarditis</strong></td>
<td></td>
</tr>
<tr>
<td>Cases with sufficient evidence only</td>
<td>83.5 (74.9-89.6)</td>
</tr>
<tr>
<td>All cases</td>
<td>63.7 (55.2-71.4)</td>
</tr>
<tr>
<td><strong>Anaphylaxis</strong></td>
<td></td>
</tr>
<tr>
<td>Cases with sufficient evidence only</td>
<td>93.3 (86.4-97)</td>
</tr>
<tr>
<td>All cases</td>
<td>72.6 (64.4-79.5)</td>
</tr>
<tr>
<td><strong>GBS</strong></td>
<td></td>
</tr>
<tr>
<td>Cases with sufficient evidence only</td>
<td>47.2 (35.8-58.9)</td>
</tr>
<tr>
<td>All cases</td>
<td>30.9 (22.9-40.3)</td>
</tr>
<tr>
<td><strong>TTS</strong></td>
<td></td>
</tr>
<tr>
<td>Cases with sufficient evidence only</td>
<td>70.2 (61.4-77.6)</td>
</tr>
<tr>
<td>All cases</td>
<td>64.4 (55.9-72.1)</td>
</tr>
<tr>
<td><strong>Febrile seizure</strong></td>
<td></td>
</tr>
<tr>
<td>Cases with sufficient evidence only</td>
<td>89 (80-94.4)</td>
</tr>
<tr>
<td>All cases</td>
<td>89 (80-94.4)</td>
</tr>
</tbody>
</table>

aAESI: adverse event of special interest.
bGBS: Guillain-Barré syndrome.
cTTS: thrombosis with thrombocytopenia syndrome.
Table 5. Stratification analysis: validation sample results.

<table>
<thead>
<tr>
<th>AESI^a and metric</th>
<th>Detected cases</th>
<th>Pre-EUA^b period</th>
<th>Post-EUA period</th>
<th>Inpatient</th>
<th>Outpatient</th>
<th>Emergency department</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Myocarditis/pericarditis (n=135)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cases, n</td>
<td>135</td>
<td>100</td>
<td>35</td>
<td>91</td>
<td>26</td>
<td>18</td>
</tr>
<tr>
<td>Total TP^c cases, n (PPV%; 95% CI)</td>
<td>86 (63.7; 55.2-71.4)</td>
<td>68 (68.0; 58.1-76.5)</td>
<td>18 (51; 35-68)</td>
<td>72 (79; 69-86)</td>
<td>10 (38; 21-59)</td>
<td>4 (22; 7-48)</td>
</tr>
<tr>
<td>Total cases with sufficient evidence, n (PPV^d% for TP cases with sufficient evidence; 95% CI)</td>
<td>103 (83.5; 74.9-89.6)</td>
<td>79 (86; 76-92)</td>
<td>24 (75; 53-89)</td>
<td>79 (91; 82-96)</td>
<td>16 (63; 36-84)</td>
<td>8 (50; 15-85)</td>
</tr>
<tr>
<td><strong>Anaphylaxis (n=135)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cases, n</td>
<td>135</td>
<td>100</td>
<td>35</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total TP cases, n (PPV%; 95% CI)</td>
<td>98 (72.6; 64.4-79.5)</td>
<td>70 (70; 60.2-78.3)</td>
<td>28 (80; 63-90.9)</td>
<td>17 (63; 42.9-79.7)</td>
<td>—</td>
<td>81 (75; 65.8-82.4)</td>
</tr>
<tr>
<td>Total cases with sufficient evidence, n (PPV%; 95% CI)</td>
<td>105 (93.3; 86.4-97)</td>
<td>74 (94.6; 86.2-98.4)</td>
<td>31 (90.3; 73.4-98)</td>
<td>19 (89.5; 65.6-99.7)</td>
<td>—</td>
<td>86 (94.2; 86.6-97.9)</td>
</tr>
<tr>
<td><strong>GBS^f (n=110)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cases, n (%)</td>
<td>110</td>
<td>65</td>
<td>45</td>
<td>110</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total TP cases, n (PPV%; 95% CI)</td>
<td>34 (30.9; 22.9-40.3)</td>
<td>24 (40; 25.9-49.5)</td>
<td>20 (44; 30.4-59.4)</td>
<td>34 (30.9; 22.8-40.3)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total cases with sufficient evidence, n (PPV%; 95% CI)</td>
<td>72 (47.2; 35.8-58.9)</td>
<td>52 (46.2; 32.9-60)</td>
<td>20 (50; 28.1-71.9)</td>
<td>72 (47.2; 35.8-58.9)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>TTS^g (n=135)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cases, n</td>
<td>135</td>
<td>100</td>
<td>35</td>
<td>133</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total TP cases, n (PPV%; 95% CI)</td>
<td>87 (64.4; 55.9-72.1)</td>
<td>64 (64; 54-72.9)</td>
<td>23 (66; 48.2-80)</td>
<td>86 (64.7; 56.1-72.4)</td>
<td>1 (100; 0-100)</td>
<td>0 (0; 0-100)</td>
</tr>
<tr>
<td>Total cases with sufficient evidence, n (PPV%; 95% CI)</td>
<td>124 (70.2; 61.4-77.6)</td>
<td>91 (70.3; 60-78.9)</td>
<td>33 (70; 51.6-83.5)</td>
<td>122 (70.5; 61.7-78)</td>
<td>1 (100; 0-100)</td>
<td>1 (100; 0-100)</td>
</tr>
<tr>
<td><strong>Febrile seizure (n=100)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cases, n</td>
<td>100</td>
<td>100</td>
<td>—</td>
<td>1</td>
<td></td>
<td>99</td>
</tr>
<tr>
<td>Total TP cases, n (PPV%; 95% CI)</td>
<td>73 (73; 63.3-80.9)</td>
<td>73; (73; 63.3-80.9)</td>
<td>—</td>
<td>0 (0; 0-100)</td>
<td>—</td>
<td>73 (74; 64.1-81.6)</td>
</tr>
<tr>
<td>Total cases with sufficient evidence, n (PPV%; 95% CI)</td>
<td>83 (88; 78.8-93.6)</td>
<td>83 (88; 78.8-93.6)</td>
<td>—</td>
<td>1 (0; 0-100)</td>
<td>—</td>
<td>82 (89; 80-94.4)</td>
</tr>
</tbody>
</table>

^aAESI: adverse event of special interest.  
bUEA: emergency use authorization.  
cTP: true positive.  
dPPV: positive predictive value.  
eNot applicable.  
fGBS: Guillain-Barré syndrome.  
gTTS: thrombosis with thrombocytopenia syndrome.
Table 6. Interrater reliability.

<table>
<thead>
<tr>
<th>AESI</th>
<th>Total cases validated, n</th>
<th>Interrater reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocarditis/pericarditis</td>
<td>162</td>
<td>0.814</td>
</tr>
<tr>
<td>Anaphylaxis</td>
<td>162</td>
<td>0.770</td>
</tr>
<tr>
<td>GBS(^b)</td>
<td>137</td>
<td>0.832</td>
</tr>
<tr>
<td>TTS(^c) at unusual sites</td>
<td>162</td>
<td>0.851</td>
</tr>
<tr>
<td>Febrile seizure</td>
<td>120</td>
<td>0.965</td>
</tr>
</tbody>
</table>

\(^a\)AESI: adverse event of special interest.
\(^b\)GBS: Guillain-Barré syndrome.
\(^c\)TTS: thrombosis with thrombocytopenia syndrome.

Stratification

To evaluate consistency across pre- and post-EUA periods and care settings, we reported true positive and PPV results for each stratum (Table 5).

None of the algorithms had notable differences between the pre- and post-EUA periods since all 95% CIs had some overlap. However, there were some differences between the PPVs for the 2 periods that could be significant with a larger validation sample. The difference in PPV for myocarditis/pericarditis varied from 68% in the pre-EUA period to 51.4% in the post-EUA period, while anaphylaxis showed the opposite pattern with a 70% PPV in the pre-EUA period that increased to 80% PPV in the post-EUA period.

We also reported stratified results by care setting (Table 5). For myocarditis/pericarditis, the PPV of cases with an inpatient care setting (79.1%, 95% CI 69.4%-86.4%) was notably higher than that from the outpatient (38.5%, 95% CI 21.2%-58.8%) or emergency department (22.2%, 95% CI 6.7%-47.9%) care settings.

Anaphylaxis did not have a large difference across care settings, as the 95% CIs overlapped between the 2 care settings. However, they did show better performance with cases in an emergency department (PPV 75%, 95% CI 65.8%-82.4%) care setting over cases with an inpatient care setting (PPV 63%, 95% CI 42.9%-79.7%). The other AESI algorithms filtered for only 1 care setting or had a vast majority of cases in 1 care setting.

Sensitivity Analysis

Medication and Observation Algorithm Changes

We analyzed whether changes to medication code lists for the myocarditis/pericarditis and GBS algorithms could improve performance. For the myocarditis/pericarditis algorithm, removal of nonsteroidal anti-inflammatory drugs from the medication code lists showed no change in PPV at 83.5% (Table 7), but PPV values were higher for cases selected with the pericarditis instead of myocarditis ICD-10 codes.

For the GBS algorithm, when cases were removed where gabapentin (used for post-GBS pain management) was the only supporting evidence, PPV increased to 38.1% (95% CI 28.2%-49.1%) from 30.9% (95% CI 22.9%-40.3%; Table 8).

Our initial febrile seizure algorithm did not use any supporting evidence to filter out possible false positives since we believed we could get adequate PPV without it.

For our sensitivity analysis, we tested requiring supporting evidence in the condition period, such as the presence of medications for reducing fever such as acetaminophen, observation evidence when the patient’s chief complaint was related to fever or seizure, or the presence of both. When filtered to only cases with either medication or observation evidence, febrile seizure PPV increased significantly to 93.3% (95% CI 84.7%-97.6%) from the original algorithm PPV of 73% (95% CI 63.3%-80.9%), with no overlap in 95% CIs and a \(P\) value of <.001 (Table 9). When the algorithm required both medication and observation evidence, it performed even better (PPV 96.9%, 95% CI 88.5%-99.9%).
Table 7. Sensitivity analysis: myocarditis/pericarditis validation sample results.

<table>
<thead>
<tr>
<th>AESI and sensitivity analysis</th>
<th>Total TP^b cases, n</th>
<th>Selected cases, n (change, n)^c</th>
<th>PPV^d, % (95% CI; change)^d</th>
<th>Selected cases with sufficient evidence, n (change, n)^d</th>
<th>PPV, % (95% CI; change)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Removal of NSAIDs^e</td>
<td>86</td>
<td>135 (0)</td>
<td>63.7 (55.2-71.4; 0)</td>
<td>103 (0)</td>
<td>83.5 (74.9-89.6; 0)</td>
</tr>
<tr>
<td>Pericarditis diagnosis^f</td>
<td>59</td>
<td>82 (–53)</td>
<td>72 (61.1-80.8; +8.3)</td>
<td>67 (–36)</td>
<td>88.1 (77.6-94.3; +4.6)</td>
</tr>
<tr>
<td>Myocarditis diagnosis^f</td>
<td>27</td>
<td>53 (–82)</td>
<td>50.9 (37.4-64.3; –12.8)</td>
<td>36 (–67)</td>
<td>75 (57.9-87.1; –8.5)</td>
</tr>
</tbody>
</table>

^aAESI: adverse event of special interest.  
^bTP: true positive.  
^cPPV: positive predictive value.  
^dValues in parentheses reflect the change due to the modified algorithm features.  
^eNSAID: nonsteroidal anti-inflammatory drug.  
^fAll International Classification of Diseases, Tenth Revision, Clinical Diagnosis codes that the algorithm used were broken into 2 groups: myocarditis (I40.0 infective myocarditis, I40.1 isolated myocarditis, I40.8 other acute myocarditis, I40.9 acute myocarditis, unspecified, and I51.4 Viral myocarditis) and pericarditis (B33.22 viral pericarditis, B33.23 acute rheumatic pericarditis, I30.0 acute nonspecific idiopathic pericarditis, I30.1 infective pericarditis, I30.8 other forms of acute pericarditis, I30.9 acute pericarditis, unspecified, I32 pericarditis in diseases classified elsewhere, and I41 meningococcgal pericarditis).

Table 8. Sensitivity analysis: Guillain-Barré syndrome validation sample results.

<table>
<thead>
<tr>
<th>AESI and sensitivity analysis</th>
<th>Total TP^b cases</th>
<th>Selected cases (change)^c</th>
<th>PPV^d, % (95% CI; change)^d</th>
<th>Selected cases with sufficient evidence (change)^d</th>
<th>PPV, % (95% CI; change)^d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Removal of gabapentin</td>
<td>33</td>
<td>86 (–24)</td>
<td>38.4 (28.6-49.2; 7.5)</td>
<td>53 (–19)</td>
<td>62.3 (48.3-74.5; +15)</td>
</tr>
<tr>
<td>Adjusted case definition</td>
<td>49</td>
<td>110 (0)</td>
<td>44.5 (35.4-54; +13.6)</td>
<td>72 (0)</td>
<td>68.1 (56.3-78; +20.8)</td>
</tr>
<tr>
<td>Adjusted case definition+removal of gabapentin</td>
<td>49</td>
<td>86 (–26)</td>
<td>57.1 (46.2-67.4; +26.2)</td>
<td>68 (–4)</td>
<td>72.1 (60-81.6; +24.8)</td>
</tr>
</tbody>
</table>

^aAESI: adverse event of special interest.  
^bTP: true positive.  
^cPPV: positive predictive value.  
^dValues in parentheses reflect the change due to the modified algorithm features.

Table 9. Sensitivity analysis: febrile seizure.

<table>
<thead>
<tr>
<th>AESI and sensitivity analysis</th>
<th>Total TP^b cases, n</th>
<th>Selected cases, n (change, n)^c</th>
<th>PPV^d, % (95% CI; change)^d</th>
<th>Selected cases with sufficient evidence, n (change, n)^c</th>
<th>PPV^d, % (95% CI; change)^d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases with either medication or observation</td>
<td>70</td>
<td>75 (–25)</td>
<td>93.3 (84.7-97.6; +20.3)</td>
<td>73 (–10)</td>
<td>95.9 (87.9-99.2; +7.9)</td>
</tr>
<tr>
<td>Cases with both medication and observation evidence</td>
<td>63</td>
<td>65 (–35)</td>
<td>96.9 (88.5-99.9; +23.9)</td>
<td>63 (–20)</td>
<td>100 (92.8-100; +12)</td>
</tr>
</tbody>
</table>

^aAESI: adverse event of special interest.  
^bTP: true positive.  
^cValues in parentheses reflect the change due to the modified algorithm features.  
^dPPV: positive predictive value.

Diagnostic Code List Changes

We also analyzed if changing diagnostic codes that were used to identify the AESI might lead to higher performance for the myocarditis/pericarditis and TTS algorithms.

For myocarditis/pericarditis, we found that an algorithm only looking for the myocarditis code (PPV 50.9%, 95% CI 37.4%-64.3%) underperformed an algorithm with just pericarditis codes (PPV 72%, 95% CI 61.1%-80.8%; Table 7). For TTS, we found that the main ICD-10 code I81 for “portal vein thrombosis” (73.5%, 95% CI 64%-81.3%) outperformed all other codes in our code list, including G08 (intracranial and intraspinal phlebitis and thrombophlebitis), I82.0 (Budd-Chiari syndrome), I82.3 (embolism and thrombosis of renal vein), and I82.890 (acute embolism and thrombosis of other specified veins), with a PPV of 36.4% (95% CI 21.3%-54.4%; Table 10).
### Case Definition Validation Criteria

Finally, we analyzed whether a small update to our case definition criteria for the GBS algorithms described in the Stratification Analysis and Sensitivity Analysis section would improve reported performance in Table 7. When we applied both changes, the validation criteria change to the algorithm and removal of gabapentin, as discussed in the Medication and Observation Algorithm Changes section, the algorithm achieved a PPV of 57.1% (95% CI 46.2%-67.4%).

### Discussion

#### Principal Findings

#### Overview

The results of this study show that for 4 out of 5 AEsIs, we can build an interoperable computable phenotype with comparable or improved performance to algorithms in the existing literature. These algorithms are developed using a rules-based approach to facilitate their application and increase the generalizability of performance across EHR databases. For the phenotypes with poorer performance, the issues were often that the case definition required documentation of a test that was lost in our data pipeline, or was not completed, or was not recorded by the treating physician or nurse. While these cases are marked as false positives based on our methodology, they may be true AEs that are lacking the documentation to meet the case definition. Some small updates to the algorithms or the case definition evaluation method could be made to potentially improve the algorithms’ performances, but a more important next step would be to validate our algorithms on other data partners to ensure generalizability of the original algorithms and any updates. Given the need for active AE surveillance, this study is still an important first step toward building an algorithm that can be distributed and implemented on health provider EHR databases and can accurately detect AEs.

The PPV results of the phenotypes, negative control groups, and stratification and sensitivity analysis are discussed in more detail in the following sections. Note our negative control groups and many of the stratification and sensitivity analyses have sample sizes too small to draw strong conclusions as illustrated by the width of the 95% CIs for those results. These were exploratory analyses completed as a supplement to the main findings of the study around the PPV of the algorithms.

#### Myocarditis/Pericarditis

The myocarditis/pericarditis algorithm showed strong PPV performance using cases with sufficient evidence. The literature appears to lack good comparison studies against which to evaluate this algorithm’s performance. A meta-analysis from 2013 reviewed myocarditis/pericarditis algorithm studies and found that none of them evaluated their algorithm by calculating PPV [45].

When myocarditis/pericarditis was segmented via care settings, algorithm performance was highest for inpatient settings, with a PPV of 79.1%. This can be attributed to the availability of supporting clinical data needed for accurate case detection in such settings. Given that inpatient testing is necessary to meet the criteria of the case definition, the algorithm performance matches clinical expectations and adds to its public health importance.

In emergency care settings, myocarditis/pericarditis is often diagnosed for patients with a history of inpatient visits to one or more other health systems. This increases the probability of these patients having additional documentation necessary to meet the case definition. This highlights the role of health information exchanges in supporting public health use cases, improving AE reporting, and enhancing postmarket surveillance.

Myocarditis/pericarditis had a notable difference in PPVs for pre- and post-EUA date. The post-EUA date strata of the sample had a higher percentage of cases coming from the emergency department, which had few cases before EUA. This could be explained by patients being diagnosed during previous inpatient stays in other health systems and a lower threshold to provide a preliminary diagnosis with limited information. This category had a lower PPV on average for myocarditis/pericarditis, likely due to less documentation in an emergency care setting than in an inpatient care setting. This highlights the need for further validation of the algorithm in these settings for an effective public health benefit and to gain confidence that our algorithm is fit for purpose. Because the aim of the algorithms is postvaccination AESI detection in support of public health safety surveillance, any potential degradation in performance in the post-EUA period is a concern. If performance decrease

---

**Table 10. Sensitivity analysis: thrombosis with thrombocytopenia syndrome (TTS).**

<table>
<thead>
<tr>
<th>AESI(^a) and sensitivity analysis</th>
<th>Total TP(^b) cases, n</th>
<th>Selected cases, n (change, n)(^c)</th>
<th>PPV(^d), % (95% CI; change)(^c)</th>
<th>Selected cases with sufficient evidence, n (change, n)(^c)</th>
<th>PPV, % (95% CI; change)(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I81</td>
<td>75</td>
<td>102 (–33)</td>
<td>73.5 (64-81.3; +9.1)</td>
<td>96 (–28)</td>
<td>78.1 (68.6-85.4; +8)</td>
</tr>
<tr>
<td>All other TTS ICD(^e) codes(^f)</td>
<td>12</td>
<td>33 (–102)</td>
<td>36.4 (21.3-54.4; –28)</td>
<td>28 (–96)</td>
<td>42.9 (25.4-62.1; –27.3)</td>
</tr>
</tbody>
</table>

\(^a\)AESI: adverse event of special interest.

\(^b\)TP: true positive.

\(^c\)Values in parentheses reflect the change due to the modified algorithm features.

\(^d\)PPV: positive predictive value.

\(^e\)ICD: International Classification of Diseases.

\(^f\)All other TTS ICD codes include G08, I82.0, I82.3, and I82.890.
in the post-EUA period is driven by postvaccination myocarditis/pericarditis being more likely to have confounding physical findings that could affect how quickly and in which care setting it gets diagnosed, the PPV from this study may not be applicable to a postvaccination version of the phenotype. There is a small overlap in the 2 periods’ PPV 95% CI, and a 2-sample proportion test returns a P value of .08. This suggests that the difference could also be due to statistical noise. However, given the importance of the post-EUA period to the algorithm’s future task and the size of the difference, we suggest validating additional cases in the post-EUA period to confirm whether the algorithm is actually less effective.

Anaphylaxis

In cases with sufficient evidence, our anaphylaxis algorithm performed strongly with a PPV score of 93.3% (95% CI 86.4%-97%). This shows a possible slight improvement over previous anaphylaxis research, although both results were within the 95% CI [33,34]. When stratified by care setting, the algorithm performed better in emergency department care settings. This can be explained due to the anaphylaxis symptoms and treatment being more likely to be well-documented in this setting. Availability of additional evidence increases the PPV of the algorithm. Since anaphylaxis cases related to vaccination are more likely to culminate in visits to the emergency department, the better performance of the algorithm would provide a better public health benefit.

Overall, the performance of the algorithm was moderate compared with that seen in literature. With no obvious avenues for improvement available, no additional sensitivity analyses were applied.

GBS Algorithm

Our initial GBS algorithm showed weak performance for GBS with a PPV of 47.2% (95% CI 35.8%-58.9%). Given existing research on GBS validations, this result is not surprising, since our result is comparable with a study result showing GBS algorithm validation PPV of 29% (95% CI 24%-34%) [35]. We hoped that our algorithm would improve on this study’s results, allowing us to meet the “moderate” performance threshold defined in the Methods section, given that we added additional logic to require suggested evidence and filter out historical diagnoses. However, we believe that the algorithm’s performance could be improved based on the sensitivity analysis results.

An increase in performance was observed when adjusting the case definition interpretation of GBS to allow for more general written clinical notes or neurology consult evidence to replace specific documented test results. The lack of standardization in laboratory results is fraught with challenges such as inconsistent data. The observed improvement in the GBS phenotype highlighted the need for further standardization to have a better impact on public health benefit.

Furthermore, the performance of the GBS algorithm was improved by the exclusion of nonspecific medications such as gabapentin, increasing its public health benefit. Gabapentin is often used to treat generalized neuropathic pain for a variety of conditions other than GBS, including diabetes, and can confound the results.

With both case definition and medication adjustments to the algorithm, the PPV rose to be closer to the moderate performance threshold and an increase over the cited historical study [35]. Because these changes were informed by the cases in the validation study post hoc, they might be overfitted to this validation sample and may not be generalizable. They should be tested in other EHR systems.

The GBS algorithm performed slightly better in the post-EUA period, but the performance of both periods was well within the 95% CI of the other period. The GBS algorithm only applies to the inpatient care setting; therefore, no care setting stratification analysis was performed.

Febrile Seizure

Our febrile seizure algorithm performed strongly, with a PPV score of 89% using cases with sufficient evidence. This performance is in line with existing febrile seizure algorithm validation research [32], where a febrile seizure validation study on the FDA Sentinel database showed a PPV of 70% (95% CI 64%-76%). Our sensitivity analysis suggests that even better performance could possibly be achieved by adding additional filters to select cases with supporting medication and observation evidence, which are well-documented in EHRs. The better performance of the algorithm provides better public health benefits and further supports the use of EHRs in the detection of AEs. For cases that met either or both criteria, the PPV increased. Since these changes to the algorithm happened after the validation was completed, they overstate the general performance increases when applied to a new EHR setting but offer avenues for a future validation study. Future research can test whether stronger performance is possible with these filters and focus on reviewing the algorithm’s application to AEs following pediatric vaccinations.

TTS Algorithm

The TTS algorithm showed moderate performance for PPV at 70.2% which is similar to a separate FDA TTS validation study which estimated the performance at 76.1% (95% CI 67.2-83.2%) [29]. TTS had consistent performance across both pre- and post-EUA periods and did not have enough cases in the outpatient and emergency department care settings for any defensible findings around diagnosis care setting stratification. Our sensitivity analysis revealed that when the AESI was diagnosed with the ICD-10 code I81 (portal vein thrombosis), the algorithm showed a significant increase when compared with the performance of all other ICD codes (73.5%, 95% CI 64%-81.3%, compared with 36.4%, 95% CI 21.3%-54.4%). Although if an increase to specificity is desired at the cost of some sensitivity, the TTS algorithm could be limited to only select the higher performing I81 diagnosis code.

Limitations

There are several limitations to this study. First, it only evaluates general AESIs and not postvaccination AESIs specifically since the algorithms do not require the evidence of vaccine administration as criteria. While this was necessary due to the...
rareness of the postvaccination AESIs in our data, it is possible that
the algorithms perform worse detecting postvaccination
AESIs specifically since they will often present slightly
differently in different populations when occurring after a
vaccine administration. For example, the major presenting
symptoms appeared to resolve faster in cases of myocarditis
after COVID-19 vaccination than in typical viral cases of
myocarditis [9]. To guard against this, we included both pre–
and post–COVID-19 EUA data with the hope that post-EUA
cases would include some postvaccination AESIs. However,
we did not have enough post-EUA cases available to build a
large enough sample size for a comparison with sufficient
statistical power to provide definitive evidence on this topic.
Another limitation in this vein is the general small sample size
for all stratification, sensitivity, and negative control analyses.
We make sure to state that these analyses are exploratory in
nature, and the reader should not form strong conclusions from
them given their small samples size and large CI range. Future
research could address these concerns by identifying a data
source with enough postvaccination AESI cases to complete a
comparably large validation study.

An additional limitation of this study is that it only measures
algorithms’ PPVs instead of investigating other metrics that
could give a better picture of the algorithm’s holistic
performance such as sensitivity and specificity. Specifically,
these other metrics would estimate how many of the total
positive cases are being identified and how well the algorithm
is able to identify cases without the AESIs. However, we believe
that this limitation is necessary for the following reasons: (1)
the main purpose of this study was to assess the PPV of
phenotypes because it answers the most relevant public health
question, if the algorithms will generate a quality detected set
of AE cases for the public health surveillance and (2) a much
higher cost and more extensive data sharing are needed to
properly estimate sensitivity and specificity because of the
required validation sample size necessary for a negative control
group. To calculate PPV, one only needs a sample of the cases
selected by the algorithm. To estimate the sensitivity and
specificity, however, it would be necessary to also validate an
extremely large negative control group sample since the AESI
conditions that the algorithms try to detect are often rare events.
We would expect it to be even more rare for these conditions
of interest for AESIs to happen and not be recorded with types
of structured data elements that are being used in the phenotypes.
In fact, the lack of structured data elements in some negative
control cases led to a clinician asking the research team if
something was wrong because their case had no relevant charted
events to be reviewed. A much larger validation study would
also expose clinicians to a larger set of patient data for cases
that have a low likelihood of having an AE. This approach limits
the interaction with protected health information data until the
algorithms’ PPVs support continued research with broader
samples and methodologies.

Another limitation is that although they were designed to be
simple to deploy, the algorithms are still time-consuming to
apply to different EHR systems. Although a hallmark of this
algorithm is its interoperability, the algorithm logic still must
be applied to the EHR common data model or extracted and
translated into another common data model as was done for this
study. Interoperable codes should be available for all patients,
given the requirement to provide patient data in an interoperable
FHIR standard. However, given the recency of the requirement,
they might not be available in all systems and require some
code translation on the health organization side, especially when
analyzing at the population level. In addition, since the
interoperable codes will only be available through a FHIR API,
this adds another data pull and integration with the EHR system
to obtain these codes for the algorithm.

In the future, the evolving landscape of health IT may facilitate
the public health use cases of detecting and reporting
postvaccination AESIs in a safe and secure manner that protects
patient privacy. This could be achieved by EHRs supporting
secure querying of patient cohorts with probable postvaccination
AESIs using clinical query language [46] or other interoperable
query language. Reducing the burden of automatic detection of
postvaccination AESIs would help public health organizations
improve AE surveillance with minimal additional burden to
health care organizations and providers.

A final limitation of this study is that the algorithms were only
applied to 1 site. Going forward, algorithm performance should
be validated at other sites to ensure their generalizability.
Although the algorithms were generated without prior input
from the data, the study is still limited to 1 health care
organization, and this method could have different operating
characteristics (PPV, sensitivity, etc) at a second location.
Future research can be performed to improve algorithm accuracy
and as stated previously would require additional partner EHR
data systems. To create a better performing algorithm, machine
learning techniques could be used to train the model to identify
specific patterns of data instead of relying on rules-based
methods that incorporate published case definition criteria and
clinical subject matter expert experience. When given enough
data, machine learning approaches generally outperform
rules-based approaches across domains, and some prior research
suggests that this is true in the medical domain as well [47].
However, machine learning methods will not generalize across
EHR systems because the data patterns that machine learning
identifies could be specific to an individual health care
organization. Trying to build a large data set that combines
multisite data is extremely difficult and costly due to concerns
over infrastructure, regulations, privacy, and data
standardization. A method such as federated learning could be
explored to alleviate this problem. Federated learning allows
multiple sites to collaboratively train a global model without
directly sharing data and has been used to train machine learning
algorithms at EHR sites previously [48].

Conclusions
In summary, this study presents strong initial evidence that
creating simple, interoperable, rules-based phenotypes can detect
AESIs on a new data source and that the phenotypes outperform
the PPV outcomes for historical validations studies for these
conditions. The study validates 5 different AESIs to prove that
this approach can work for a broad range of AESIs, while also
highlighting where the approach might be less successful. For
example, the GBS algorithm was built using ICD-10 codes that previous validation studies have demonstrated are not accurate predictors of a GBS case that meets case definition criteria; subsequently, our GBS algorithm performed poorly. The validation study sample sizes for all AESIs allowed for adequate precision to evaluate algorithm PPV against historical studies.

An active surveillance system can enhance vaccine safety and aid in the development and use of safer vaccines and recommendations to minimize the AE risks after vaccination [49]. The algorithms were developed using a method that should be able to be applied to and generalize performance for new EHR databases, but more research is needed to confirm this. If the methodology can be successfully used to detect postvaccination AESI cases across EHR databases, these algorithms could be deployed widely to inform FDA decision-making, promote public safety, and improve public confidence. Going forward, further research and investigation are needed to enhance algorithm performance and integrate the algorithms across health care organizations for active surveillance in the interest of public health.

Acknowledgments
Development of the manuscript benefitted from significant engagement with the Food and Drug Administration (FDA) Center for Biologics Evaluation and Research (CBER) team members and their partners. The authors thank them for their contributions. Additional feedback on the manuscript was provided by IBM Consulting (Stella Muthuri and Brian Goodness), Accenture Consulting (Shayan Hobbi), and Korrin Bishop (writing and editing). This research was funded through the FDA CBER Biologics Effectiveness and Safety Initiative. Several coauthors hold commercial affiliations with Accenture, IBM Consulting, and MedStar Health Research Institute. Accenture (PSH); IBM Consulting (AAH, JP, JB, AS, EM, LDJ, and MD); and MedStar Health Research Institute (AZH and JB) provided support in the form of salaries for authors but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability
The data sets generated during and analyzed during this study are not publicly available, and were only made available to the Food and Drug Administration for the purpose of evaluating algorithms for adverse events of special interest outcomes. For inquiries or questions regarding the data, individual queries should be directed to the corresponding author.

Conflicts of Interest
Authors AAH, JP, JB, AS, EM, LDJ, and MD are or were employed by IBM while participating in the study. PSH is employed by Gevity Consulting, Inc, a part of Accenture. Authors AZH and HJB are employed by MedStar Health Research Institute, and AZH holds an appointment with Georgetown University School of Medicine. These authors have delivered clinical and epidemiology consulting engagement for public and private sector partners. These affiliations did not impact the study design, data collection and analysis, decision to publish, or preparation of the manuscript and do not alter our adherence to JMIR policies on sharing data and materials. The opinions expressed are those of the authors and do not necessarily represent the opinions of their respective organizations. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Multimedia Appendix 1
Search terms and code lists for 5 developed phenotypes and detailed case definitions.

References


44. Standardized data: the OMOP common data model. Observational Health Data Sciences and Informatics. URL: https://www.ohdsi.org/data-standardization/ [accessed 2024-02-22]


Abbreviations

AE: adverse event
AESI: adverse event of special interest
API: application programming interface
BEST: Biologics Effectiveness and Safety Initiative
CBER: Center for Biologics Evaluation and Research
CDC: Centers for Disease Control and Prevention
CSF: cerebrospinal fluid
EHR: electronic health record
EUA: emergency use authorization
FDA: Food and Drug Administration
FHIR: Fast Healthcare Interoperable Resources
GBS: Guillain-Barré syndrome
HAPI: Health Level 7 application programming interface
ICD-10: International Classification of Diseases, 10th Revision, clinical modification
OHDSI: Observational Health Data Sciences and Informatics
PPV: positive predictive value
RWD: real-world data
TTS: thrombosis with thrombocytopenia syndrome
USCDI: United States Core Data for Interoperability
VAERS: Vaccine Adverse Event Reporting System
WBC: white blood cell

Edited by A Mavragani, T Sanchez; submitted 09.06.23; peer-reviewed by B Ru, AS Bhagavathula; comments to author 20.01.24; revised version received 24.02.24; accepted 26.05.24; published 15.07.24.

Please cite as:
Development of Interoperable Computable Phenotype Algorithms for Adverse Events of Special Interest to Be Used for Biologics Safety Surveillance: Validation Study
JMIR Public Health Surveill 2024;10:e49811
URL: https://publichealth.jmir.org/2024/1/e49811
doi:10.2196/49811
PMID:39008361

©Ashley A Holdefer, Jeno Pizarro, Patrick Saunders-Hastings, Jeffrey Beers, Arianna Sang, Aaron Zachary Hettinger, Joseph Blumenthal, Erik Martinez, Lance Daniel Jones, Matthew Deady, Hussein Ezzeldin, Steven A Anderson. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 15.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
The Association Between Personality Traits and Health-Related Quality of Life and the Mediating Role of Smoking: Nationwide Cross-Sectional Study

Jiangyun Chen1,*, PhD; Jiahuan Wan1,*, PhD; Li Gan1, MA; Haomiao Li2, PhD; Yan Zhou3, MA; Siyuan Liu4, PhD; Lan Luo5; Haozheng Zhou4; Xuanhao Yin4; Jinghui Chang1, PhD

Abstract

Background: There are positive and negative correlations in different directions between smoking, personality traits, and health-related quality of life (HRQOL), where smoking may mask the pathway between personality traits and HRQOL. Understanding the masking pathway of smoking between personality traits and HRQOL can elucidate the mechanisms of smoking’s psychosocial effects and provide new ideas for developing tobacco control strategies.

Objective: The purpose of this study was to investigate the correlation between Big Five personality traits and HRQOL and whether smoking mediates the relationship between them.

Methods: This was a cross-sectional study using data from 21,916 respondents from the 2022 Psychology and Behavior Investigation of Chinese Residents survey. Linear regression models were used to analyze the correlations between smoking, Big Five personality traits, and HRQOL while controlling for potential confounders. The mediating role of smoking on the association between Big Five Personality traits and HRQOL was analyzed using the Sobel-Goodman mediation test.

Results: Extraversion (β=0.001; P=0.04), agreeableness (β=0.003; P<0.001), and neuroticism (β=0.003; P<0.001) were positively correlated with HRQOL, whereas openness was negatively correlated with HRQOL (β=−0.001; P=0.003). Smoking was associated with a decrease in HRQOL and mediated the positive effect of HRQOL on extraversion (z=−2.482; P=0.004), agreeableness (z=−2.264; P=0.02), and neuroticism (z=−3.230; P=0.001). Subgroup analyses further showed that smoking mediated the effect of neuroticism on HRQOL in the population with chronic illnesses (z=−2.724; P=0.006), and in the population without chronic illnesses, smoking contributed to the effect of HRQOL on extraversion (z=−2.299; P=0.02), agreeableness (z=−2.382; P=0.02), and neuroticism (z=−2.213; P=0.03).

Conclusions: This study provided evidence that there is a correlation between personality traits and HRQOL. It also found that smoking plays a role in mediating the connection between personality traits and HRQOL. The development of future tobacco control strategies should consider the unique traits of each individual’s personality, highlighting the significance of extraversion, agreeableness, and neuroticism.

(KEYWORDS: Big Five personality; HRQOL; smoking; mediation; tobacco control; China; neuroticism; extraversion; agreeableness; health-related quality of life)

Introduction

A widely recognized theoretical framework in the field of personality psychology is the Big Five personality model, which classifies personality traits into 5 dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism [1]. The stability and validity of the Big Five personality theory have been consistently confirmed by previous research, which has conducted long-term follow-up studies, cross-cultural studies, and comparisons across various age groups. Studies have revealed that the Big Five personality traits are significantly correlated with diverse behaviors and adaptations [2], along
with their interplay with genetic and environmental factors [3,4]. Various characteristics of individuals have been discovered to have a strong association with academic success [5], professional pathways [6], interpersonal connections [7], physical well-being, and psychological well-being [8,9]. The extensive applications of the Big Five personality theory span across multiple domains, encompassing talent management [10], mental health research [11], style of behavior [11,12], clinical practice [13], personal growth, and education [14,15]. By evaluating the Big Five personality traits, scientists have gained insights into individuals’ personality characteristics, offering direction and support. Prior research has additionally indicated that although a person’s character attributes remain relatively consistent, the manifestation of these attributes is not completely rigid. Moreover, an individual’s encounters, surroundings, and cultural upbringing can also influence their character traits and their overall quality of life in terms of health (health-related quality of life [HRQOL]) [16,17].

The Big Five personality theory is extensively used in the domains of health behavior and psychology to examine how personality traits influence behavioral traits and health outcomes in populations [18,19]. Research has confirmed a robust correlation between aspects of the Big Five personality traits and HRQOL. People who have a strong inclination toward openness are generally more welcoming toward novel encounters, have a wider array of hobbies, and exhibit a sense of curiosity. These traits are linked to increased levels of contentment and overall quality of life [20]. Individuals who possess a strong sense of duty and a proactive attitude toward accomplishing tasks tend to have high conscientiousness. This trait is linked to improved mental well-being, reduced negative emotions, and enhanced HRQOL [21]. Individuals who have a high level of extraversion generally have enhanced social abilities, display optimistic emotional expression, and are more inclined to form and sustain positive interpersonal connections. As a result, they experience increased life satisfaction and happiness [22]. People who possess a strongly agreeable nature tend to pay closer attention to the happiness and welfare of those around them. They demonstrate kindness, empathy, and assistance in social engagements, which are linked to improved HRQOL, increased social support, and reduced feelings of isolation [23]. Individuals exhibiting elevated levels of neuroticism are prone to experience anxiety, tension, and negative emotions, which have been linked to diminished levels of life contentment and overall welfare [24]. However, some studies have proposed the concept of “healthy neuroticism,” which refers to individuals with neurotic traits that do not lead to physical and mental health problems [25,26]. The healthy neuroticism theory suggests that neurotic individuals may pursue perfection, possess higher alertness and introspection, and have higher demands on themselves, which may positively influence health behaviors and thus positively impact HRQOL through high alertness to unhealthy behaviors [27].

Personality is also associated with smoking. For example, people with a high level of extraversion are usually more prone to seeking excitement and engaging in daring pursuits, which increases their susceptibility to the temptation of smoking [28]. On the other hand, people with a strong sense of conscientiousness are inclined to be more accountable and have a greater tendency to follow healthy habits, such as refraining from smoking [29]. People with a high level of agreeableness typically exhibit a compliant and cooperative demeanor. As such, those with higher agreeableness exhibit more positive social and prosocial behaviors [30], but they are also more susceptible to being socially influenced to smoke. People with a high level of openness tend to be more inclined to engage in novel activities and are more susceptible to experimenting with smoking [31]. In contrast, people with a high level of neuroticism are prone to feeling anxious and stressed and experiencing negative emotions, which increases their likelihood of initiating smoking and makes it harder for them to quit [32].

Older adults and individuals with underlying diseases are particularly affected by smoking, as it leads to a decline in HRQOL due to health hazards, mental strain, financial strain, and limitations on social activities [33,34]. Multiple research studies have firmly established a robust correlation between tobacco use and the emergence of diverse ailments, such as increasing the burden of cerebral cancer and mortality [35], underscoring its substantial capacity to jeopardize human well-being. According to the World Health Organization Report on the Global Tobacco Epidemic 2021, tobacco remains a significant contributor to untimely fatalities on a global scale [36]. China, being the biggest manufacturer and user of tobacco on a global scale, bears the sole responsibility for around 1 million tobacco-related deaths [37]. The use of tobacco is a significant contributing factor linked to the greatest load of long-term illness, and a decrease in tobacco consumption can result in decreased occurrences of heart disease, stroke, and additional chronic ailments.

Past studies have confirmed that characteristics of an individual’s personality affect both their HRQOL and smoking habits, where smoking is found to have an adverse effect on HRQOL. Nevertheless, the correlation among these 3 variables remains incompletely comprehended, particularly concerning individuals with long-term illnesses. Examining the mechanisms of interaction between Big Five personality traits, smoking, and HRQOL and conducting tobacco control efforts at the level of individual personality traits can provide new perspectives for improving HRQOL. Based on these reasons, this study formulated the following hypotheses:

- **Hypothesis 1**: The Big Five personality traits influence HRQOL.
- **Hypothesis 2**: Smoking is associated with Big Five personality traits and HRQOL, where smoking plays a mediating role between them.

**Methods**

**Participants**

The information used in this investigation was acquired from the 2022 Psychology and Behavior Investigation of Chinese Residents survey. From June to August 2022, a comprehensive survey was carried out in 148 cities; 202 districts and counties; 390 townships, towns, and streets; and 780 communities and villages spanning 23 provinces, 5 autonomous regions, and 4
municipalities under the central government. To ensure the overall representativeness of the study population, the survey used a multistage sampling technique, incorporating stratified sampling at various levels including provincial; city; district and county; township, town, and street; and community and village. Quota sampling was used at the community and village level as well as at the individual level, using quotas that were determined based on sex and age attributes from the data of the Seventh National Population Census. In every city, there was recruitment of at least 1 enumerator or survey team, where each enumerator had the duty of gathering 30 - 90 questionnaires and each survey team had the duty of gathering 100 - 200 questionnaires. The questionnaires were distributed through the web-based Questionnaire Star platform, and if face-to-face surveys were possible in the area, the investigator filled out the questionnaires on site on a one-to-one basis. In the event that face-to-face surveys were impractical due to the constraints of the new coronavirus outbreak, electronic surveys were individually provided to the participants. Participants provided their responses by clicking on the questionnaire link, and they were required to give their informed consent. The study included individuals who were at least 12 years of age, held citizenship in the People’s Republic of China, and were permanent residents of China with an annual out-of-home time of no more than 1 month. Participants who did not fulfill the criteria for this research were disqualified. A total of 23,414 questionnaires were collected for the study, ensuring high quality and national representativeness of the data. After eliminating duplicates and excluding missing data and logically inconsistent outlier data, 21,916 respondents were finally included in this study, with a valid response rate of 93.6% (21,916/23,414). The survey protocol has been published [38].

Ethical Considerations
This study complied with the ethical review rules of the Health Culture Research Center of Shaanxi (JKWH-2022-02). Informed consent was obtained for the investigation. Respondents completed an anonymous, web-based survey in approximately 30 minutes.

Variables

Dependent Variable
The health status of the population was assessed using HRQOL in this study. The measurement of HRQOL was conducted using the EQ-5D-5L traditional scale, which has been proven to be better than its previous version, the EQ-5D-3L, in terms of practicality, upper limit impact, distinguishing ability, and agreement with other measures [39,40]. The EQ-5D-5L scale consists of 5 aspects: mobility, self-care, daily activity performance, pain or discomfort, and anxiety or depression, as specified in Multimedia Appendix 1. There are 5 levels for each dimension, ranging from 1 (no problems) to 5 (extreme problems). The levels of these questions can describe 243 combinations ranging from 11,111 (perfectly good) to 55,555 (perfectly poor). These health states are assigned an index value, known as the health state index (HIS), which reflects the weighting of society’s preference for the health state. The HIS score varied from below 0 (where 0 represents the health state value of death; negative values indicate a health state worse than death) to 1 (representing perfect health), with higher scores indicating better health utility [40]. Based on their health preferences, the HIS value estimates for the Chinese population vary between −0.391 and 1, representing the worst and best outcomes. In this study, the final HIS was obtained according to the utility value conversion formula \((X – \min)/(\max – \min)\) of \([0,1]\) [41].

Independent Variable
The assessment of personality involved the use of the Big Five Inventory (BFI)–10, a condensed variant of the comprehensive BFI-44. In this study, the reliability and validity of the BFI-10 were assessed to confirm its suitability for situations where there are time constraints or it is not feasible to conduct a personality assessment (eg, telephone surveys, etc). The scale consists of 10 items that assess 5 personality dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness. A 5-point Likert scale is used to score each item, with higher scores indicating a stronger trait. Extraversion is assessed by questions 1 and 6, agreeableness is assessed by questions 2 and 7, conscientiousness is assessed by questions 3 and 8, neuroticism is assessed by questions 4 and 9, and openness is assessed by questions 5 and 10. It should be mentioned that questions 1, 3, 5, 7, and 9 are scored in the opposite direction [42].

Mediation Variables
Smoking behavior was assessed by asking respondents about their current smoking habits. Specifically, they were asked, “Do you currently smoke?” Responses were categorized into 5 levels: 0=“No”; 1=“Yes, regular cigarettes”; 2=“Yes, e-cigarettes”; 3=“Yes, both”; and 4=“Ever (quit).” For this study, smoking was divided into 2 categories depending on whether participants were presently smoking or not: 0=never smoked or used to smoke but have stopped, and 1=currently smoking.

Covariates
The variables examined in this research consist of the socioeconomic background of the participants (including sex, age, group, area of residence, ethnicity, political status, religion, household income, educational level, occupation, and social status); family attributes (marital status and family type); lifestyle elements (smoking habits and alcohol intake); and mental health status related to perception of stress, perception of social support, self-confidence, and health knowledge. Detailed information on the definitions and categorization of these variables can be found in Multimedia Appendix 2. The choice of covariates was determined by their correlation with the independent variables, as well as their impact on the association between the independent variables and the dependent variable. Age group and sex were adjusted as fixed covariates. If the other covariates changed the dependent variable by more than 10% with the independent variable or were significantly associated with the dependent variable, they were included as potential confounding factors in the final model. The Empower software (X&Y Solutions) was used to test the selected covariates, which were chosen based on established associations or plausible biological relationships. These covariates include...
ethnicity, political status, religion, area of residence (a special Chinese identifier that impacts various aspects of life in China, such as purchasing a house or a car, public health insurance reimbursement rate, and welfare benefits), household income, education level, occupation, social status, marital status, family type, alcohol consumption, stress perception ability, social support appreciation ability, self-efficacy, and health literacy. Multimedia Appendices 3 and 4 contain detailed findings.

Statistical Analysis

The basic study population description included the presentation of chronic diseases characteristics (yes, no, and total) as mean (SD) for continuous variables and as frequency (%) for categorical variables. To examine variations in the attributes of chronic diseases, a 2-tailed Student t test was used for continuous factors, whereas the χ² test was used for categorical factors. The correlation between smoking, Big Five personality traits, and HRQOL was measured using linear regression models. This was done before and after adjusting for covariates, and the findings were presented as β coefficients along with 95% CIs. The Sobel-Goodman mediation test was used to examine the impact of smoking on Big Five personality traits and HRQOL while taking into account all covariates [43]. Statistical significance was determined using 2-sided P values, with α<.05 as the threshold. The analysis of data was conducted using Stata (version 17; StataCorp).

Results

General Characteristics

The sample analyzed in this study consisted of 21,916 cases. There was an equal distribution of sexes, with 10,958 (50%) participants identifying as male and 10,958 (50%) identifying as female. A total of 71.4% (n=15,647) of the participants fell within the age group of 18 - 59 years, and 56.75% (n=12,437) of them were married. The vast majority of respondents were of Han nationality (n=19,970, 91.12%), had no religion (n=21,058, 96.09%), and had “the masses” as their political status (n=13,912, 63.48%). Over half (n=11,811, 53.89%) of the participants lived in urban regions, with a greater proportion belonging to the high-income bracket (n=8032, 36.65%). The majority (n=15,214, 69.42%) of the respondents reported never drinking alcohol, and the largest proportion (n=9773, 44.59%) had tertiary education. In all, 34.68% (n=7601) were employed, with over half (n=11,574, 52.81%) of the family type being a core family. The prevalence of smoking was 14.87% (n=3258). The average social status of the respondents was close to the upper-middle class (mean 4.35, SD 1.30; out of a total score of 6). Among individuals with chronic illnesses, there was a notable decline in HRQOL (mean 0.92, SD 0.10), which was significantly lower than that of the overall sample (mean 0.96, SD 0.10; P<.001). Individuals with chronic illnesses exhibited a diminished level of extraversion (mean 6.14, SD 1.61; P<.001) compared to that of the overall sample (mean 6.23, SD 1.62). The average rating for agreeableness was 7.00 (SD 1.48). The average score for conscientiousness was 6.76 (SD 1.65), and individuals with chronic diseases exhibited a higher level of conscientiousness (mean 6.98, SD 1.65; P<.001). The average score for neuroticism was 6.27 (SD 1.56). Individuals with chronic illnesses exhibited a diminished level of openness (mean 6.20, SD 1.61), which was significantly lower than that of the overall sample (mean 6.46, SD 1.55; P<.001). The average rating for perceived stress capacity was 6.55 (SD 2.54), indicating an increase among individuals with chronic illnesses (mean 6.63, SD 2.57; P=.02). Moreover, the population with chronic illnesses experienced a decrease in their corresponding competencies, specifically in comprehending social support (mean 15.03, SD 3.78), self-efficacy (mean 7.79, SD 2.42), and health literacy (mean 27.55, SD 5.30), suggesting a decline in these abilities (P<.001). Table 1 contains comprehensive details.
Table. Characteristics of respondents.a.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Overall (N=21,916)</th>
<th>With chronic disease (n=5460)</th>
<th>Without chronic disease (n=21,796)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>12 - 17</td>
<td>2072 (9.45)</td>
<td>136 (2.49)</td>
<td>1936 (11.76)</td>
<td></td>
</tr>
<tr>
<td>18 - 59</td>
<td>15,647 (71.4)</td>
<td>2943 (53.9)</td>
<td>12,704 (77.2)</td>
<td></td>
</tr>
<tr>
<td>≥60</td>
<td>4197 (19.15)</td>
<td>2381 (43.61)</td>
<td>1816 (11.04)</td>
<td></td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>10,958 (50)</td>
<td>2854 (52.27)</td>
<td>8104 (49.25)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10,958 (50)</td>
<td>2606 (47.73)</td>
<td>8352 (50.75)</td>
<td></td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never married</td>
<td>8497 (38.77)</td>
<td>851 (15.59)</td>
<td>7646 (46.46)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>12,437 (56.75)</td>
<td>4059 (74.34)</td>
<td>8378 (50.91)</td>
<td></td>
</tr>
<tr>
<td>Divorce</td>
<td>406 (1.85)</td>
<td>160 (2.93)</td>
<td>246 (1.49)</td>
<td></td>
</tr>
<tr>
<td>Widowed</td>
<td>576 (2.63)</td>
<td>390 (7.14)</td>
<td>186 (1.13)</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.86</td>
</tr>
<tr>
<td>Han nationality</td>
<td>19,970 (91.12)</td>
<td>4972 (91.06)</td>
<td>14,998 (91.14)</td>
<td></td>
</tr>
<tr>
<td>Ethnic minority</td>
<td>1946 (8.88)</td>
<td>488 (8.94)</td>
<td>1458 (8.86)</td>
<td></td>
</tr>
<tr>
<td><strong>Religion, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>None</td>
<td>21,058 (96.09)</td>
<td>5091 (93.24)</td>
<td>15,967 (97.03)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>858 (3.91)</td>
<td>369 (6.76)</td>
<td>489 (2.97)</td>
<td></td>
</tr>
<tr>
<td><strong>Political status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Party member or probationary party member</td>
<td>3179 (14.51)</td>
<td>1059 (19.4)</td>
<td>2120 (12.88)</td>
<td></td>
</tr>
<tr>
<td>Member of the Communist Youth League</td>
<td>4671 (21.31)</td>
<td>457 (8.37)</td>
<td>4214 (25.61)</td>
<td></td>
</tr>
<tr>
<td>Other parties</td>
<td>154 (0.7)</td>
<td>59 (1.08)</td>
<td>95 (0.58)</td>
<td></td>
</tr>
<tr>
<td>The masses</td>
<td>13,912 (63.48)</td>
<td>3885 (71.15)</td>
<td>10,027 (60.93)</td>
<td></td>
</tr>
<tr>
<td><strong>Area of residence, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.007</td>
</tr>
<tr>
<td>Urban</td>
<td>11,811 (53.89)</td>
<td>2856 (52.31)</td>
<td>8955 (54.42)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>10,105 (46.11)</td>
<td>2604 (47.69)</td>
<td>7501 (45.58)</td>
<td></td>
</tr>
<tr>
<td><strong>Family income, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Low</td>
<td>7229 (32.99)</td>
<td>2050 (37.55)</td>
<td>5179 (31.47)</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>6655 (30.37)</td>
<td>1628 (29.82)</td>
<td>5027 (30.55)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>8032 (36.65)</td>
<td>1782 (32.64)</td>
<td>6250 (37.98)</td>
<td></td>
</tr>
<tr>
<td><strong>Alcohol intake, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>15,214 (69.42)</td>
<td>3277 (60.02)</td>
<td>11,937 (72.54)</td>
<td></td>
</tr>
<tr>
<td>All the time</td>
<td>3266 (14.9)</td>
<td>1019 (18.66)</td>
<td>2247 (13.65)</td>
<td></td>
</tr>
<tr>
<td>Used to drink, but does not drink now</td>
<td>2148 (9.80)</td>
<td>947 (17.34)</td>
<td>1201 (7.3)</td>
<td></td>
</tr>
<tr>
<td>Did not drink in the past, but drinks now</td>
<td>1288 (5.88)</td>
<td>217 (3.97)</td>
<td>1071 (6.51)</td>
<td></td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Primary school and below</td>
<td>3412 (15.57)</td>
<td>1528 (27.99)</td>
<td>1884 (11.45)</td>
<td></td>
</tr>
</tbody>
</table>
Correlation Analysis

Before accounting for covariates, the linear regression model revealed a detrimental association between smoking and HRQOL ($\beta=-.028; P<.001$). Additionally, HRQOL exhibited significant correlations with extraversion ($P<.001$), agreeableness ($P<.001$), conscientiousness ($P<.001$), neuroticism ($P<.001$), and openness ($P=.001$). After making adjustments for various factors such as sex, age range, ethnicity, political status, religion, area of residence, household income, place of residence, education level, occupation, social status, marital status, family structure, alcohol consumption, ability to perceive stress, ability to appreciate social support, self-confidence, and health knowledge, the findings indicated that smoking still had a negative association with HRQOL ($\beta=-.016; P<.001$). Additionally, extraversion ($\beta=.04$), agreeableness ($\beta=.003; P<.001$), neuroticism ($\beta=.003; P<.001$), and openness ($\beta=.001; P=.003$) were all significantly linked to HRQOL. However, in the population with chronic illnesses, the findings of the model, after accounting for covariates, indicated that only tobacco use ($\beta=-.021; P<.001$), agreeableness ($\beta=.003; P=.005$), and neuroticism ($\beta=.005; P<.001$) exhibited a correlation with HRQOL. Table 2 provides the comprehensive details.

Before accounting for covariates, the initial linear regression analysis revealed that smoking had a negative correlation with...
agreeableness ($\beta = -0.107; P < 0.001$), neuroticism exhibited a positive correlation with smoking ($\beta = 0.185; P < 0.001$), and openness displayed a negative correlation with smoking ($\beta = -0.247; P < 0.001$). The findings after accounting for covariates indicated that extraversion ($\beta = 0.077; P = 0.02$), agreeableness ($\beta = 0.059; P = 0.04$), and neuroticism ($\beta = 0.089; P = 0.004$) exhibited a positive association with smoking. After accounting for covariates, the model demonstrated a noteworthy impact of neuroticism ($\beta = 0.155; P = 0.004$) within the group of individuals with chronic illnesses. Table 3 provides a comprehensive overview of the detailed information.

Table 3. Linear regression analysis for the associations of health-related quality of life (HRQOL) with smoking and Big Five personality traits.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Overall (N=21,916)</th>
<th>With chronic disease (n=5460)</th>
<th>Without chronic disease (n=21,796)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjust-</td>
<td>Adjust-</td>
<td>P value</td>
</tr>
<tr>
<td></td>
<td>ed, $\beta^a$ (95% CI)</td>
<td>ed, $\beta$ (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.028 (-0.032 to -0.024)</td>
<td>-0.016 (-0.020 to -0.012)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Extraversion</td>
<td>0.004 (0.003 to 0.005)</td>
<td>0.001 (0.000 to 0.002)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.008 (0.007 to 0.009)</td>
<td>0.003 (0.002 to 0.003)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.006 (0.005 to 0.006)</td>
<td>0.001 (0.000 to 0.001)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>0.008 (0.007 to 0.009)</td>
<td>0.003 (0.003 to 0.004)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Openness</td>
<td>0.001 (0.001 to 0.002)</td>
<td>-0.001 (-0.002 to 0.000)</td>
<td>.001</td>
</tr>
</tbody>
</table>

$^a$ $\beta$: beta coefficient.
$^b$ Adjusting for sex, age group, marital status, ethnicity, religion, political status, chronic disease, area of residence, family income, alcohol intake, education level, occupation, social status, family type, health literacy, self-efficacy, perceived stress, and perceived social support.
$^c$ Adjusting for sex, age group, marital status, ethnicity, religion, political status, area of residence, family income, alcohol intake, education level, occupation, social status, family type, health literacy, self-efficacy, perceived stress, and perceived social support.
there was a positive correlation between agreeableness and extraversion (β=.005; P=.02), whereas smoking showed a negative correlation with HRQOL (β=−.013; P<.001). Additionally, agreeableness was positively correlated with HRQOL (β=.003; P<.001). Smoking mediated −1.5% of the effect HRQOL had on agreeableness (z=−2.64; P=.002). Neuroticism exhibited a positive correlation with smoking (β=.006; P<.001), whereas smoking showed a negative correlation with HRQOL (β=−.013; P<.001). Additionally, neuroticism displayed a positive correlation with HRQOL (β=.003; P<.001). Smoking mediated −2.3% of the effect HRQOL had on neuroticism (z=−3.230; P=.001). Figure 1 displays the ultimate mediation model.
The mediating effect of smoking on Big Five personality traits and health-related quality of life (HRQOL), as explored by the Sobel-Goodman mediation test.

<table>
<thead>
<tr>
<th>Extraversion&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Agreeableness&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Conscientiousness&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Neuroticism&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Openness&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>P value</td>
<td>Value</td>
<td>P value</td>
<td>Value</td>
</tr>
<tr>
<td>Big Five personality trait → smoking, β</td>
<td>0.005</td>
<td>.002</td>
<td>0.004</td>
<td>.02</td>
</tr>
<tr>
<td>Smoking → HRQOL, β</td>
<td>–0.012</td>
<td>&lt;.001</td>
<td>–0.013</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Indirect effect, β</td>
<td>–0.000</td>
<td>.004</td>
<td>–0.000</td>
<td>.002</td>
</tr>
<tr>
<td>Direct effect, β</td>
<td>0.000</td>
<td>.24</td>
<td>0.003</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Total effect, β</td>
<td>0.000</td>
<td>.30</td>
<td>0.003</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Proportion of the total effect that is mediated</td>
<td>–0.133</td>
<td>__&lt;sup&gt;b&lt;/sup&gt;</td>
<td>–0.015</td>
<td>—</td>
</tr>
<tr>
<td>Sobel-Goodman mediation test</td>
<td>–2.842</td>
<td>.004</td>
<td>–2.264</td>
<td>.02</td>
</tr>
</tbody>
</table>

<sup>a</sup> Adjusting for sex, age group, marital status, ethnicity, religion, political status, chronic disease, area of residence, family income, alcohol intake, education level, occupation, social status, family type, health literacy, self-efficacy, perceived stress, and perceived social support.

<sup>b</sup> Not applicable.
Figure 1. Mediating model of smoking in the association between Big Five personality traits and health-related quality of life (HRQOL).

Subgroup Analysis

In Table 5, the subgroup analysis revealed that smoking acted as a mediator between Big Five personality traits and HRQOL in individuals with chronic diseases. In particular, among individuals with chronic diseases, there was a positive correlation between neuroticism and smoking ($\beta=.012; P=.001$), whereas smoking showed a negative correlation with HRQOL ($\beta=-.020; P<.001$). Additionally, neuroticism exhibited a positive correlation with HRQOL ($\beta=.005; P<.001$). Smoking mediated $-5.1\%$ of the effect HRQOL had on neuroticism ($z=-2.724, P=.006$). Smoking mediated $-13.6\%$ of the effect HRQOL had on extraversion ($z=-2.299; P=.02$), $-1.7\%$ of the effect HRQOL had on agreeableness ($z=-2.382; P=.02$), and $-1.5\%$ of the effect HRQOL had on neuroticism ($z=-2.213; P=.03$) among the population without chronic illnesses. The extent to which smoking behavior mediated the relationship between agreeableness and HRQOL was slightly greater than that of neuroticism, whereas smoking behavior was the least relevant on the relationship between extraversion and HRQOL. Figure 2 displays the ultimate mediation model for the subgroup of the population with chronic diseases.
Table. Subgroup analysis of mediation models for Big Five personality traits associated with health-related quality of life (HRQOL), as mediated by smoking.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Extraversion&lt;sup&gt;a&lt;/sup&gt; Value</th>
<th>Extraversion&lt;sup&gt;a&lt;/sup&gt; P value</th>
<th>Agreeableness&lt;sup&gt;a&lt;/sup&gt; Value</th>
<th>Agreeableness&lt;sup&gt;a&lt;/sup&gt; P value</th>
<th>Conscientiousness&lt;sup&gt;a&lt;/sup&gt; Value</th>
<th>Conscientiousness&lt;sup&gt;a&lt;/sup&gt; P value</th>
<th>Neuroticism&lt;sup&gt;a&lt;/sup&gt; Value</th>
<th>Neuroticism&lt;sup&gt;a&lt;/sup&gt; P value</th>
<th>Openness&lt;sup&gt;a&lt;/sup&gt; Value</th>
<th>Openness&lt;sup&gt;a&lt;/sup&gt; P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>With chronic disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Five personality</td>
<td>0.007</td>
<td>.03</td>
<td>0.001</td>
<td>.69</td>
<td>-0.003</td>
<td>.44</td>
<td>0.012</td>
<td>.001</td>
<td>0.005</td>
<td>.16</td>
</tr>
<tr>
<td>τ→smoking, β</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking→HRQOL, β</td>
<td>-0.019</td>
<td>&lt;.001</td>
<td>-0.019</td>
<td>&lt;.001</td>
<td>-0.019</td>
<td>&lt;.001</td>
<td>-0.020</td>
<td>&lt;.001</td>
<td>-0.019</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Indirect effect, β</td>
<td>-0.000</td>
<td>.51</td>
<td>-0.000</td>
<td>.69</td>
<td>0.000</td>
<td>.45</td>
<td>-0.000</td>
<td>.06</td>
<td>-0.000</td>
<td>.18</td>
</tr>
<tr>
<td>Direct effect, β</td>
<td>0.001</td>
<td>.26</td>
<td>0.004</td>
<td>&lt;.001</td>
<td>0.003</td>
<td>.01</td>
<td>0.005</td>
<td>&lt;.001</td>
<td>-0.001</td>
<td>.35</td>
</tr>
<tr>
<td>Total effect, β</td>
<td>0.001</td>
<td>.32</td>
<td>0.004</td>
<td>&lt;.001</td>
<td>0.003</td>
<td>.01</td>
<td>0.005</td>
<td>&lt;.001</td>
<td>-0.001</td>
<td>.31</td>
</tr>
<tr>
<td>Proportion of the total effect that is mediated</td>
<td>-0.133</td>
<td>____&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.006</td>
<td>—</td>
<td>0.018</td>
<td>—</td>
<td>-0.051</td>
<td>—</td>
<td>0.083</td>
<td>—</td>
</tr>
<tr>
<td>Sobel-Goodman test</td>
<td>-1.948</td>
<td>.051</td>
<td>-0.397</td>
<td>.69</td>
<td>0.760</td>
<td>.45</td>
<td>-2.724</td>
<td>.006</td>
<td>-1.328</td>
<td>.18</td>
</tr>
<tr>
<td>Without chronic disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big Five personality</td>
<td>0.004</td>
<td>.007</td>
<td>0.005</td>
<td>.005</td>
<td>0.003</td>
<td>.04</td>
<td>0.004</td>
<td>.01</td>
<td>-0.002</td>
<td>.25</td>
</tr>
<tr>
<td>τ→smoking, β</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking→HRQOL, β</td>
<td>-0.009</td>
<td>&lt;.001</td>
<td>-0.009</td>
<td>&lt;.001</td>
<td>-0.009</td>
<td>&lt;.001</td>
<td>-0.009</td>
<td>&lt;.001</td>
<td>-0.009</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Indirect effect, β</td>
<td>-0.000</td>
<td>.02</td>
<td>-0.000</td>
<td>.02</td>
<td>-0.000</td>
<td>.07</td>
<td>-0.000</td>
<td>.03</td>
<td>0.000</td>
<td>.26</td>
</tr>
<tr>
<td>Direct effect, β</td>
<td>0.000</td>
<td>.44</td>
<td>0.003</td>
<td>&lt;.001</td>
<td>0.001</td>
<td>.02</td>
<td>0.003</td>
<td>&lt;.001</td>
<td>-0.002</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Total effect, β</td>
<td>0.000</td>
<td>.50</td>
<td>0.003</td>
<td>&lt;.001</td>
<td>0.001</td>
<td>.03</td>
<td>0.003</td>
<td>&lt;.001</td>
<td>-0.002</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Proportion of the total effect that is mediated</td>
<td>-0.136</td>
<td>—</td>
<td>-0.017</td>
<td>—</td>
<td>-0.023</td>
<td>—</td>
<td>-0.015</td>
<td>—</td>
<td>-0.008</td>
<td>—</td>
</tr>
<tr>
<td>Sobel-Goodman test</td>
<td>-2.299</td>
<td>.02</td>
<td>-2.382</td>
<td>.02</td>
<td>-1.841</td>
<td>.07</td>
<td>-2.213</td>
<td>.03</td>
<td>1.121</td>
<td>.26</td>
</tr>
</tbody>
</table>

<sup>a</sup> Adjusting for sex, age group, marital status, ethnicity, religion, political status, registered permanent residence, family income, alcohol intake, education level, work status, social status, household type, health literacy, self-efficacy, perceived stress, and perceived social support.

<sup>b</sup> Not applicable.
**Discussion**

**Principal Findings**

Prior research has established robust connections between tobacco use and HRQOL [44], along with the correlations between personality traits and both smoking and HRQOL [45,46]. Nevertheless, there has been no research conducted so far that has investigated the correlation among these 3 variables. The impact of smoking on the relationship between the Big Five personality traits and HRQOL was investigated in our research. We used data from a nationwide cross-sectional survey conducted in China, which encompassed 32 provinces, including autonomous regions and centrally governed municipalities.

Through the process of mediation decomposition, we were able to provide additional evidence that smoking mediates the relationship between Big Five personality traits and HRQOL. The results of our study indicate that there is not only a direct correlation between Big Five personality traits and HRQOL but also an indirect correlation through smoking. Subgroup analyses for people with chronic conditions were also carried out.

The study findings indicated that extraversion, agreeableness, and neuroticism have a positive correlation with HRQOL, whereas openness has a negative correlation. Previous research has demonstrated that the Big Five personality traits impact well-being, life satisfaction, and self-identity, and thus HRQOL, through different personality traits [46]. People who possess...
elevated levels of extraversion and agreeableness have a greater tendency to form enduring and beneficial social connections, which can help reduce psychological strain and feelings of isolation [22, 23]. Moreover, they possess a sunnier perspective on existence and exhibit enhanced resilience in the face of life’s obstacles and stresses [47]. Surprisingly, our research discovered a positive correlation between high neuroticism and HRQOL, which contradicts previous study results. There is a widespread belief that elevated neuroticism is linked to feelings of anxiety, depression, and various other negative emotions [48], ultimately resulting in a diminished quality of life [24, 49]. However, according to the theory of healthy neuroticism, people with healthy neurotic traits aim for flawlessness and exhibit a heightened state of vigilance and self-reflection toward detrimental actions. As a result, their chances of encountering physical and psychological health issues are reduced [25, 26]. Conversely, individuals who possess elevated levels of openness are not constrained by conventional notions and routines, and they might experience remorse for impulsive or erroneous behaviors [50]. Excessive attention to and maintenance of interpersonal relationships can also diminish life satisfaction and happiness [51].

Furthermore, our study provided evidence for the favorable correlations among extraversion, agreeableness, and neuroticism personality traits with smoking. Personality traits that promote social interactions, such as being friendly and reliable, are linked to high levels of extraversion and agreeableness [52]. As smoking is often perceived as a social behavior in some settings, these individuals may be more susceptible to social influence to start or quit smoking [53, 54]. Conversely, people with elevated levels of neuroticism have a greater susceptibility to anxiety and stress [55]. Since nicotine can provide temporary relief from these emotions [56], individuals with high neuroticism may be inclined to use smoking as a coping mechanism, increasing their likelihood of initiating smoking or becoming regular smokers.

The act of smoking has been associated with a decrease in different aspects of HRQOL and the emergence of multiple long-term illnesses [57]. The results of our study indicated a consistent negative correlation between smoking and HRQOL, which was observed in both the entire population and the subgroup with chronic diseases, aligning with previous research. Furthermore, our research suggests that smoking mediates the relationship between extraversion, agreeableness, and neuroticism personality traits and HRQOL. High levels of extraversion and agreeableness are positively correlated with HRQOL, but they are also strongly associated with smoking, as these personality types are more likely to smoke due to a need for social interaction [58]. Smoking has a more significant correlation with HRQOL compared to the favorable correlations of high extraversion and agreeableness, thus smoking’s mediation weakens this positive effect. Furthermore, our research discovered that smoking can mediate the positive impact of highly healthy neuroticism on HRQOL. Prior research demonstrated that high neuroticism can be positively associated with HRQOL by promoting “healthy neuroticism” or introspection; it is also strongly associated with smoking, as people with high neuroticism are more likely to become dependent on tobacco for anxiety relief and experience symptoms of tobacco dependence [32]. The negative effect of smoking on HRQOL is greater than that of high neuroticism. Due to the mediation of smoking, the positive effect of high neuroticism on HRQOL is also weakened.

The findings from the subgroup analysis additionally indicate that smoking plays a mediating role in connecting neuroticism and HRQOL in the population with chronic diseases. However, the mediating effect in the population with chronic diseases remains similar to that of the overall population. The different mediating effects of personality traits in the populations with and without chronic diseases may be due to several factors. For example, the act of smoking is a major contributor to long-term health conditions, and people who have chronic illnesses may experience mental health issues such as neuroticism and anxiety due to their ailment. Furthermore, persistent illnesses frequently necessitate extended periods of therapy and medication, potentially leading to the formation of a neurotic character trait [59]. The findings of this research suggest that although the influence of mediation was minimal for certain traits, the Sobel-Goodman test produced noteworthy outcomes, indicating the existence of mediated routes. Considering the limited impact magnitudes, it is conceivable that alternative mechanisms are at play.

To summarize, our study supports the notion that smoking acts as a mediator in the relationship between the Big Five personality traits and HRQOL. Therefore, using a single tobacco control plan for the entire community may not be the best course of action; instead, tailored smoking cessation tactics based on various personality qualities can be taken into consideration. For individuals with high levels of extraversion and agreeableness, interventions such as smoking cessation environments; legislation; and support from partners, friends, or support groups can greatly increase the chances of successfully quitting smoking [60]. Providing emotional support, actively listening to their emotions and uncertainties, and assisting them in discovering suitable emotion management methods such as deep breathing and relaxation exercises to manage mood fluctuations during the process of quitting smoking could potentially yield greater advantages for individuals exhibiting elevated levels of agreeableness [61]. For people with high neuroticism, it is more important to promote self-reflection among people with high neuroticism and shape healthy neuroticism by sharing the dangers of smoking and health knowledge; provide anxiety management techniques such as deep breathing, meditation, or relaxation training to help them cope with anxiety and stress during the process of quitting smoking; and emphasize internal factors such as self-efficacy in interventions [62, 63]. It is noteworthy that the correlation coefficients between personality traits and HRQOL in this study were small, and that HRQOL may provide a critical research perspective not from a clinical but from a psychosocial point of view, as HRQOL covers a wealth of information and personality traits are a potential factor influencing HRQOL. We initially explored the pathway through which Big Five personality traits influence individual HRQOL, and this pathway does exist. In addition, the mechanisms by which Big Five personality traits acts on HRQOL may be
complex, and some mediating effects may be overshadowed by direct effects.

**Limitation**

Although this study revealed a mediating role of smoking in the relationship between Big Five personality traits and HRQOL, it is important to acknowledge the existence of certain constraints that need to be considered. First, this study has the inherent limitations of cross-sectional studies in inferring causality. Because a cross-sectional study is conducted at a specific point in time, it can only reveal correlations between variables and cannot directly determine causality. Thus, although our cross-sectional study found associations between Big Five personality traits, smoking, and HRQOL, these results were not sufficient to suggest a causal relationship between them. Future research can explore the potential reciprocal association between the Big Five personality traits and HRQOL using longitudinal and prospective studies, thereby further validating and explaining our findings. Second, the correlation coefficients and mediating effects of our study were not very large, and further exploration needs to be made in the future as to exactly how Big Five personality traits affect HRQOL and how smoking mediates the relationship between Big Five personality traits and HRQOL. Finally, since all variables were reported by the participants themselves, there is a possibility of recall and cognitive biases being present, which could impact the precision of factors associated with health and personality. Furthermore, the formation of an individual’s character requires a significant amount of time, and as one matures, their character tends to become more steadfast and influenced by their surroundings. As a result, personality scores may have some bias in their immediate outcomes.

**Conclusion**

This study demonstrated that smoking mediates the relationship between extraversion, agreeableness, and neuroticism personality traits and HRQOL. Additionally, smoking can mediate the effect neuroticism have on HRQOL in a population with chronic illnesses. In the future, when creating tobacco control strategies, it is important to consider the impact of personality, as suggested by these findings. We hope that our study will contribute to increasing the global smoking cessation rate and reducing the incidence of chronic diseases caused by smoking. This could assist in the advancement of campaigns promoting smoke-free initiatives and aid in the creation of a healthier and smoke-free atmosphere.

**Acknowledgments**

The National Natural Science Foundation of China (grant 72204107), the Basic and Applied Basic Research Foundation of Guangdong (grant 2021A1515110743), the Science and Technology Program of Guangzhou (grants 2024A04J02668 and 2023A04J2267), the Medical Scientific Research Foundation of Guangdong (grants A2023125 and B2023441), the Student Innovation and Entrepreneurship Training Program of Guangdong Province (grant S202312121138), the Key Laboratory of Philosophy and Social Sciences of Guangdong Higher Education Institutions for Health Polices Research and Evaluation (grant 2015WSY0010), and the Public Health Service System Construction Research Foundation of Guangzhou (2021 - 2023) provided support for this project. The study design, data collection, data analysis, data interpretation, and report writing were not influenced by the funders. The authors in charge were granted complete access to all of the study’s data and held the ultimate responsibility for the determination to submit the findings for publication.

**Authors’ Contributions**

J Chen, YW, and J Chang played a role in formulating and designing the study. Data acquisition involved HL, SL, HZ, and XY. Data analysis was completed by JW, YZ, and LL. The initial version of the manuscript was written by J Chen, LG, and JW. J Chang and YW assisted in overseeing the analysis of data and the revision of the manuscript. All authors participated in the revision of the paper and endorsed the submitted final version.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**
Health-related quality of life (HRQOL) questions and scores.
[DOCX File, 14 KB - publichealth_v10i1e51416_app1.docx ]

**Multimedia Appendix 2**
Variable description.
[DOCX File, 17 KB - publichealth_v10i1e51416_app2.docx ]

**Multimedia Appendix 3**
The covariate selection process (step 1): analyzing the relationship between each covariate and health-related quality of life (HRQOL).
[DOCX File, 18 KB - publichealth_v10i1e51416_app3.docx ]
References


29. Chen et al. JMIR Public Health Surveill 2024 | vol. 10 | e51416 | p.664


Abbreviations

BFI: Big Five Inventory
HIS: health state index
HRQOL: health-related quality of life
information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: Physical exercise is one of the main nonpharmacological treatments for most pathologies. In addition, physical exercise is beneficial in the prevention of various diseases. The impact of physical exercise has been widely studied; however, existing meta-analyses have included diverse and heterogeneous samples. Therefore, to our knowledge, this is the first meta-analysis to evaluate the impact of different physical exercise modalities on telomere length in healthy populations.

Objective: In this review, we aimed to determine the effect of physical exercise on telomere length in a healthy population through a meta-analysis of randomized controlled trials.

Methods: A systematic review with meta-analysis and meta-regression of the published literature on the impact of physical exercise on telomere length in a healthy population was performed. PubMed, Cochrane Library, SCOPUS, Web of Science, and Embase databases were searched for eligible studies. Methodological quality was evaluated using the Risk Of Bias In Nonrandomized Studies of Interventions and the risk-of-bias tool for randomized trials. Finally, the certainty of our findings (closeness of the estimated effect to the true effect) was evaluated using Grading of Recommendations, Assessment, Development, and Evaluations (GRADE).

Results: We included 9 trials that met the inclusion criteria with fair methodological quality. Random-effects model analysis was used to quantify the difference in telomere length between the exercise and sham groups. Meta-analysis showed that exercise did not significantly increase telomere length compared with the control intervention (mean difference=0.0058, 95% CI −0.05 to 0.06; P=.83). Subgroup analysis suggested that high-intensity interventional exercise significantly increased telomere length compared with the control intervention in healthy individuals (mean difference=0.15, 95% CI 0.03-0.26; P=.01). Furthermore, 56% of the studies had a high risk of bias. Certainty was graded from low to very low for most of the outcomes.
Conclusions: The findings of this systematic review and meta-analysis suggest that high-intensity interval training seems to have a positive effect on telomere length compared with other types of exercise such as resistance training or aerobic exercise in a healthy population.

Trial Registration: PROSPERO CRD42022364518; http://tinyurl.com/4fwb85ff

KEYWORDS
meta-analysis; aging; exercise; older; telomere length

Introduction

Background

Telomeres are nucleoprotein structures located at the ends of eukaryotic chromosomes and are of critical importance both in the maintenance of genomic stability and in the processes of tumor suppression and aging [1]. In most eukaryotic cells, telomeres consist of tandem repeats of a guanine-rich sequence (TTAAGGG) that develop at the end of chromosomes in the 5’ to 3’ direction, with a complementary cytidine-rich chain [2].

Telomeric sequences may vary between species; however, every organism possesses the same repetitive sequence for all telomeres. At birth, the telomeres of human somatic cells contain approximately 15 kilobases. In the absence of telomerase, an average of 25 to 200 bases are lost from the telomeric ends at each cell division [3]; when the length of the telomere reaches below a critical limit, cell division ceases, and the cell ages and dies [4]. The main function of telomeres is to protect the ends of chromosomes and prevent their degradation and fusion while maintaining genomic stability [5,6].

Several studies have suggested that short telomere length is associated with progressive acceleration of aging, including an increase in age-related diseases such as osteoporosis, cancer, and dementia [7-9]. Therefore, it seems evident that controlling telomere length could be a key factor in the aging process and health care.

Regular physical exercise is one of the main nonpharmacological strategies used to prevent the onset of age-related diseases. Physical exercise is defined as planned, structured, repetitive, and purposeful physical activity, that is, for the improvement or maintenance of one or more components of physical fitness [10]. Werner et al [11] observed that endurance athletes have a larger telomere size than inactive controls. Moreover, physically active middle-aged twins have longer telomeres than inactive siblings [12]. Therefore, it seems clear that regular physical exercise is essential for healthy aging and supporting positive mental health. It can help delay, prevent, or manage many costly and difficult chronic diseases faced by older adults [13]. It can also reduce the risk of premature death and moderate or severe functional limitations in older adults [14].

However, few prospective studies have evaluated the effect of physical exercise on telomere length. In addition, these studies had large methodological differences, such as heterogeneous samples, different physical exercise modalities, and varied time and duration of the interventions. To date, only one meta-analysis [15] has studied the relationship between different physical exercise modalities and telomere length; however, the populations analyzed in the qualitative and quantitative analyses were heterogeneous. Therefore, the results should be cautiously interpreted.

Objective

This systematic review and meta-analysis aimed to study the impact of different physical exercise modalities on telomere length in prospective studies (clinical trials and randomized controlled trials [RCTs]) in which the study sample comprised a healthy population without any type of pathology.

Methods

This systematic review and meta-analysis was conducted in accordance with the guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Multimedia Appendix 1) [16].

Literature Search

To identify relevant studies on the impact of physical exercise on telomere length, we conducted a systematic literature search using the following English-language databases (until September 2022): PubMed, Web of Science, SCOPUS, Embase, and Cochrane Library. The search was performed independently by 2 researchers (JLSG and VNL). The search strategy used to identify all potential studies using the following terms is detailed in Multimedia Appendix 2 [17-25].

We also manually searched the references cited in selected articles or reviews to identify relevant studies.

Study Selection

We used the Population, Intervention, Comparison, Outcomes, Time, and Study design as a framework to formulate eligibility criteria (Textboxes 1 and 2) [26].
**Textbox 1. Population, Intervention, Comparison, Outcomes, Time, and Study design framework.**

- **Population**: healthy adults with no neurological disease
- **Intervention**: interventions with exercise as the main focus were selected
- **Compare**: control group not performing physical exercise
- **Outcomes**: telomere length was assessed using both peripheral blood and saliva samples
- **Time**: no temporal restrictions were applied to the duration of the intervention or outcome measures
- **Studies**: only randomized controlled trials (RCTs) and controlled trials were included

**Textbox 2. Inclusion and exclusion criteria.**

**Inclusion criteria**
- **Article type**: Randomized controlled trials or controlled trials
- **Language**: English
- **Population**: Healthy population
- **Type of intervention**: Aerobic exercise, resistance training, or high-intensity interval training
- **Outcome**: Measurement of telomere size by peripheral blood or saliva collection

**Exclusion criteria**
- **Article type**: Case studies, systematic reviews, and meta-analyses
- **Language**: Spanish, Chinese, French
- **Population**: Population with neoplastic processes, neurodegenerative diseases, and cognitive alterations
- **Type of intervention**: Any other type of nonexercise intervention
- **Outcome**: Any other type of measure that purports to measure aging but is not telomeric length

**Data Extraction**

Two investigators (JLSG and VNL) independently extracted data. A standardized methodology was used to obtain data from studies that met the inclusion criteria. Data were obtained for the first author, year of publication, design, patient characteristics, intervention protocol and timing, study outcomes (telomere size before and after intervention), and the telomere size calculation technique. In addition, the means and SDs of the study results were obtained. The authors of the included studies were contacted via email to access potentially unclear data. If no responses were received, the data were excluded from the analysis.

**Interrater Reliability**

Interrater reliability for screening, risk of bias assessment, and quality of the evidence rating were assessed using percentage agreement and Cohen κ coefficient [27,28]. There was strong agreement between the reviewers for the screening records and full texts (98.51% agreement rate and κ=0.91) and the risk of bias assessment (92.86% agreement rate and κ=0.83).

**Risk of Bias and Assessment Methodological Quality of the Studies**

Two reviewers independently assessed the risk of bias of the included studies (SVR and VNL).

The risk of bias in nonrandomized studies of interventions (NRSIs) was assessed using the Risk Of Bias In Nonrandomized
Studies of Interventions (ROBINS-I) [29]. This tool assesses the risk of bias in NRSI results. The types of NRSIs that can be assessed with this tool are quantitative studies that estimate the efficacy (harm or benefit) of an intervention and did not use randomization to assign units (individuals or groups of individuals) to comparison groups. ROBINS-I considers 6 domains: randomization process (D1), bias arising from period and carryover effects (D5), deviations from the intended interventions (D3), missing outcome data (D4), and selection of the reported result (D5).

In contrast, a revised tool was used to assess the risk of bias in randomized clinical trials (risk-of-bias tool for randomized trials; RoB2) [30]. The tool was structured into 5 domains through which bias could be introduced into the outcome. These were identified on the basis of empirical evidence and theoretical considerations. Because the domains cover all types of bias that may affect the results of randomized trials, each domain is mandatory; and no additional domains should be added. The 5 domains for individual randomized trials (including crossover trials) were bias arising from the randomization process (D1), bias due to deviations from intended interventions (D2), bias due to missing outcome data (D3), bias in measurement of the outcome (D4), and bias in selection of the reported result (D5) [31,32].

Overall Quality of Evidence
The overall quality of evidence was based on classifying the results into levels of evidence according to the Grading of Recommendations Assessment, Development, and Evaluation (GRADE), which is based on five domains: (1) study design, (2) imprecision, (3) indirect, (4) inconsistency, and (5) publication bias.

Evidence was categorized into the following four levels accordingly: (1) high quality: further research is very unlikely to change our confidence in the estimate of effect, and all 5 domains are also met; (2) moderate quality: further research is likely to have an important impact on our confidence and might change the estimate of effect, and 1 of the 5 domains is not met; (3) low quality: further research is very likely to have an important impact on our confidence and is likely to change the estimate of effect, and 2 of the 5 domains are not met; and (4) very low quality: any estimate of effect is very uncertain, and 3 of the 5 domains are not met [31,32].

Statistical Analysis
Mean differences (MDs) after the intervention were used to compare values between the exercise and control groups, with a 95% CI. To obtain the effect size, the MD between the groups was converted to the standardized MD with a 95% CI. Statistical significance was set at $P < .05$. The individual effect size of each study and calculation of the overall effect size are presented as forest plots.

The restricted maximum likelihood method estimated the variance of between-study heterogeneity; the presence of between-study heterogeneity was assessed with the Cochran Q statistic test (with $P < .05$ considered significant) and the degree of heterogeneity with the inconsistency index ($I^2$) [33]. An $I^2$ value between 0% and 25% was considered to represent small heterogeneity, between 25% and 75% medium heterogeneity, and >75% large heterogeneity [34]. $I^2$ complements the $Q$ test, although it has the same power problems when the number of studies is small [34]. When the $Q$ test was significant ($P < .10$) and the $I^2$ result was >25%, indicating heterogeneity between studies, the random-effects model was applied in the meta-analysis. When heterogeneity was >25% according to the $I^2$ statistic, outliers (studies whose 95% CI cutoff was lower and greater than the pooled 95% CI upper and lower cutoff) and influential case analysis were performed using the analysis according to the graph of Baujat et al [35] (graph showing the contribution of each study to the overall heterogeneity compared with its contribution to the overall pooled result). The identified studies were flagged as outliers or influential cases and were removed. A subgroup analysis was performed according to the type of exercise used (resistance training, aerobic exercise, or high-intensity interval training [HIIT]). An a priori meta-regression analysis was performed on the variables of exercise intensity and duration, as well as the year of publication and methodological quality, to evaluate whether these variables influenced the overall effect size.

Skewness was assessed using a contour-enhanced funnel plot in analyses consisting of at least 5 studies, indicating the possible publication bias of small studies small studies with negative results. In the absence of publication bias, the plot resembled a symmetrical funnel-shape.

Studies were analyzed with R software (R Foundation for Statistical Computing) [36], using the Metafor package [37] as detailed by Harrer et al [38], and with the computer software Review Manager (version 4.1; The Cochrane Collaboration).

Results
Study Selection and Characteristics
The search found 3102 records, of which 1612 were duplicates and 1490 were screened by title and abstract. We found 30 studies that were potentially relevant and excluded 21 studies after screening their full reports. Finally, 9 studies met the eligibility criteria and were included in the qualitative analysis, and 7 studies, which included 1320 participants, were included in the quantitative analysis. The entire screening process is shown in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram in Figure 1.
Qualitative Summary of the Included Studies

All 9 studies were intervention studies (RCTs or controlled trials) and were of fair to good methodological quality according to the ROBINS-I or RoB2 scale (Figure 2 [17-20,22-25]). These studies were conducted in Germany [17,18], Canada [19], the United States [20], the United Kingdom [21], Iran [22], Brazil [23], Spain [24] and South Korea [25]. A total of 1320 participants were included, including both men and women, with the latter being represented to a greater extent (98%).

Regarding the type of exercise, 4 studies performed resistance training [17,18,21,25], 4 used aerobic exercises [17,19,20,23], 3 used HIIT [17,22,23], and 1 used combined training (aerobic plus resistance training) [24]. Regarding protocol duration, 1 study conducted a 2-week intervention [22], 1 study conducted a 4-week intervention [23], 4 studies conducted a 6-week intervention [17,18,24,25], and 3 studies conducted a 12-week intervention [19-21]. The intensity information for each protocol is presented in Table 1.
Figure 2. Risk of bias.
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Group; sample size; age (y), mean (SD)</th>
<th>Sample size, n</th>
<th>Protocol intervention</th>
<th>Duration (wk)</th>
<th>Laboratory techniques and procedures</th>
<th>Quality score (PE-Dro)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noori-mofrad and Ebrahim [22], 2018</td>
<td>RCT²</td>
<td>G1: 15; 19.7 (1.1), G2: 15; 20.13 (0.64)</td>
<td>G1: 0, G2: 0</td>
<td>G1: HIIT*, 6-10 laps of HIIT 30 s at an intensity of 150%-175% of Pmax. G2: control group</td>
<td>30</td>
<td>PCR. The DNA from these cells was extracted using standard salting out-protease K method. The concentration and quality of the extracted DNA were examined using NanoDrop at wavelengths of 260 and 280 nm, and the ratio of the 2 wavelengths was used. Two PCR reactions were performed for each sample, the first reaction for telomeric DNA fragment and the second for its control gene, acid ribosomal phosphoprotein. DNA telomeric length was calculated based on the ratio of telomere to control gene.</td>
<td>4/11</td>
</tr>
<tr>
<td>Ribeiro et al [23], 2021</td>
<td>RCT</td>
<td>G1: 0, G2: 0, G3: 0</td>
<td>G1: 30; 28.50 (5.76), G2: 28; 29.14 (5.26), G3: 29; 28.97 (4.32)</td>
<td>G1: control group, G2: continuous aerobic physical training. Treadmill training, 3 times a week for 16 weeks, from 30 min in the first week to 50 min in the last week. G3: intermittent aerobic physical training. Treadmill training, 3 times a week for 16 weeks, from 30 min in the first week to 50 min in the last week</td>
<td>87</td>
<td>PCR. DNA integrity was accessed by agarose gel stained with GelRead, and the concentration was determined using the NanoDrop 2000c spectrophotometer. Telomere length was determined by calculating the telomere to single-copy gene ratio using DCT²/Ctelomere/Csingle gene). The telomere length was expressed as the relative T/S ratio, normalized to the mean of the T/S ratio of the reference sample</td>
<td>6/11</td>
</tr>
<tr>
<td>Shin et al [25], 2008</td>
<td>RCT</td>
<td>G1: 8; (46.8) (6.4), G2: 8; 46.8 (6.4), G3: 0</td>
<td>G1: 0; 28.50 (5.76), G2: 28; 29.14 (5.26), G3: 29; 28.97 (4.32)</td>
<td>G1: aerobic exercise. 3 d/wk. 10 min warm-up, 45-min treadmill walk/run at 60% VO2max², 5 min cool-down. G2: did not participate in any form of regular exercise. G3: control group</td>
<td>31</td>
<td>PCR. The final telomere oligo-primer concentration were tel1, 270 nM; tel2, 900 nM. The final 36B4 (single copy gene oligo-primer concentrations were 36B4a, 300 nM; 36B4d, 500 nM. Relative T/S values were determined by sample T/S values compared with reference DNA T/S values</td>
<td>5/11</td>
</tr>
<tr>
<td>Werner et al [17], 2019</td>
<td>RCT</td>
<td>G1: 12; 50.2 (7.4), G2: 9; 49.5 (7), G3: 10; 48.4 (6.5), G4: 14; 48.1 (7.5)</td>
<td>G1: 23; 50.2 (7.4), G2: 17; 49.5 (7), G3: 19; 48.4 (6.5), G4: 20; 48.1 (7.5)</td>
<td>G1: control group, G2: aerobic endurance training. 3 d/wk, 45-min session G3: interval training. 3 d/wk, 45-min session G4: resistance training. 3 d/wk, 45-min session</td>
<td>114</td>
<td>PCR. DNA concentrations were quantified photometrically to ensure sufficient quantity and purity. PCR data were exported to Microsoft Excel, formatted, and analyzed with the comparative Ct method (2-ΔΔCt) to calculate T/S ratios and thereby relative differences in the amount of telomere repeat DNA between the individual pre- vs poststudy time points</td>
<td>5/11</td>
</tr>
<tr>
<td>Study</td>
<td>Design</td>
<td>Group: sample size; age (y), mean (SD)</td>
<td>Sample size, n</td>
<td>Protocol intervention</td>
<td>Duration (wk)</td>
<td>Laboratory techniques and procedures</td>
<td>Quality score (PESDro°)</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------</td>
<td>----------------------------------------</td>
<td>----------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>--------------------------------------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Sánchez-González et al [24], 2021</td>
<td>RCT</td>
<td>Male: G1: 0; G2: 0; G3: 49.5 (7); G4: 48.4 (6.5); G5: 48.1 (7.5)</td>
<td>74</td>
<td>G1: aerobic-resistance training. 3 d/wk.</td>
<td>24</td>
<td>PCR. DNA was determined by measuring the absorbance at 260 nm using a NanoDropTM 2000/2001 spectrophotometer. The purity of the DNA was analyzed based on the A260/280 absorbancy ratio, where an optimal purity ratio ranged between 1.8 and 2.0. The Ct comparative method was used to calculate the relative expression levels of each amplicon</td>
<td>4/11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female: G1: 41; 72.0 (4.13); G2: 33; 71.21 (4.32); G3: 17; 49.3 (7); G4: 19; 48.4 (6.5); G5: 20; 48.1 (7.5)</td>
<td></td>
<td>G2: control group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G3: aerobic endurance training. 3 d/wk, 45-min session</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G4: interval training. 3 d/wk, 45-min session</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G5: resistance training. 3 d/wk, 45-min session</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigendorf et al [18], 2019</td>
<td>RCT</td>
<td>Male: G1: 0; G2: 0; G3: 146; 53.0 (4.9); G4: 145; 52.8 (4.7)</td>
<td>291</td>
<td>G1: resistance training. 210 min of resistance training per week for 6 months</td>
<td>24</td>
<td>PCR. For assessment of telomere length, genomic DNA was extracted from peripheral blood mononuclear cells using QIAamp DNA Mini kit (Qiagen, Hilden, Germany). Telomere length was calculated as abundance of telomeric template vs a single copy gene (36B4)</td>
<td>7/11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female: G1: 146; 53.0; G2: 145; 52.8</td>
<td></td>
<td>G2: control group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nickels et al [21], 2022</td>
<td>CTh</td>
<td>Male: G1: 0; G2: 0; G3: 11; 50.8 (7.5); G4: 11; 49.3 (6.1)</td>
<td>22</td>
<td>G1: pilates training. Minimum of 2 sessions of 1 h for 12 months.</td>
<td>52</td>
<td>PCR. Whole blood was utilized as the starting material for DNA extraction and the concentration and purity were evaluated by spectrophotometry. Intra-assay coefficient of variation for calculated T/S ratio was 4.6%. Interassay coefficient of variation for calculated T/S ratio was 2.8%</td>
<td>4/11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female: G1: 11; 50.8; G2: 11; 49.3</td>
<td></td>
<td>G2: control group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mason et al [20], 2013</td>
<td>RCT</td>
<td>Male: G1: 0; G2: 0; G3: 117; 58; G4: 117; 58</td>
<td>439</td>
<td>G1: did not receive intervention</td>
<td>52</td>
<td>PCR. The DNA from these cells was extracted using standard salting out-proteinase K method. The concentration and quality of the extracted DNA were examined using Nano drop at wavelengths of 260 and 280 nm, and the ratio of the 2 wavelengths were used. Two PCR reactions were performed for each sample, the first reaction for telomeric DNA fragment and the second for its control gene, acid ribosomal phosphoprotein. DNA telomere length was calculated based on the ratio of telomere to control gene</td>
<td>4/11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female: G1: 87; 58; G2: 118; 58; G3: 117; 58; G4: 117; 58</td>
<td></td>
<td>G2: calorie-reduced diet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G3: aerobic exercise. 45-min moderate to vigorous [≥ 4 METs] exercise at heart rate of 70%-85%; 5 d/wk.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G4: aerobic exercise. 45-min moderate to vigorous [≥ 4 METs] exercise at heart rate of 70%-85%; 5 d/wk; +calorie-reduced diet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedenreich et al [19], 2019</td>
<td>RCT</td>
<td>Male: G1: 0; G2: 0; G3: 99; 60.4; G4: 113; 60</td>
<td>212</td>
<td>G1: did not receive intervention</td>
<td>52</td>
<td></td>
<td>8/11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female: G1: 99; 60.4; G2: 113; 60</td>
<td></td>
<td>G2: control group</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

https://publichealth.jmir.org/2024/1/e46019
Quality score (PEDro)

| Study Design | Group: sample size; age (y), mean (SD) | Sample size, n | Protocol intervention | Duration (wk) | Laboratory techniques and procedures | Quality score (PEDro)

- G1: aerobic exercise ≥45 min for 5 d/wk (supervised 3 d/wk by ALPHA) Trial exercise trainers+unsupervised 2 d/wk
- Sedentary individuals (exercise <90 min/wk or 90–120 min/wk maximum)

PCR Sample reactions were set up in triplicate using the EpMotion 5075 (Eppendorf, United States), containing 20 ng of template DNA, Power SYBR Green PCR

aPEDro: Physiotherapy Evidence Database.
bRCT: randomized controlled trial.
cHIIT: high-intensity interval training.
dPCR: polymerase chain reaction.
eDCt: delta cycle threshold.
fT/S: telomere/single gene.
gVO2max: Volume of Oxygen Maximum.
hCT: controlled trial.
iMET: metabolic equivalents.

Risk of Bias

Owing to the design of the included studies, 8 were analyzed using the RoB2, and 1 study was analyzed using the ROBINS-I. As assessed by the RoB2 and ROBINS-I, 56% (5/8) of the studies showed a high risk of bias, 33% (2/8) showed some concerns, and 11% (1/8) showed a low risk of bias. The item with the highest risk of bias was “deviations from the intended interventions” in which 45% (3/8) of the studies showed a high risk of bias, and the item “missing data” had 33% (2/8) of the studies that showed a high risk of bias in therapist blinding (Figure 2).

Effects of Exercise on Telomere Length

The meta-analysis showed that overall, exercise did not produce a significant increase in telomere length compared with that of the control groups, which did not exercise (MD 0.02, 95% CI −0.10 to 0.13; P=.77; N=1058; Figure 3 [17-25]). The restricted maximum likelihood method estimated a between-study heterogeneity variance of $\tau^2=0.0034$ and an $\hat{I}^2$ value of 70%, indicating significant heterogeneity among the studies included in the analysis ($P<.01$).

When performing an analysis of influential cases in the heterogeneity and outlier studies (random-effects model), we detected 2 influential cases (Figures 4 and 5) [17-25]: the study by Sánchez-González et al [24] (which was also considered an outlier) and the study by Friedenreich et al [19]. Excluding the influential cases from the meta-analysis resulted in reduced heterogeneity between studies (23%) and did not affect the results of the meta-analysis (Table 2).

The subgroup analysis according to the type of exercise showed significant differences between the groups ($P=.05$). Resistance training (MD −0.02, 95% CI −0.01 to 0.05; $P=.54; \hat{I}^2=16\%$) and aerobic exercise (MD −0.01, 95% CI −0.0 to 0.06; $P=.64; \hat{I}^2=0\%$) groups showed no significant differences compared with the control group, but the HIIT group showed significant differences compared with the control group, with a greater telomere length observed in the HIIT group (MD 0.15, 95% CI 0.03-0.26; $P=.01; \hat{I}^2=0\%$; Figure 6 [17,18,20-23,25]).

Meta-regression analysis showed no relationship between exercise intensity and duration, year of publication, and methodological quality of the included studies ($P<.05$).
Figure 3. Forest plot of the effect of exercise on telomere length. Forest plot of the results of a random-effects meta-analysis is shown as mean differences with 95% CI for the comparison of mean telomere length in the exercise and control groups. The shaded square represents the point estimate for the individual study and the study weight in the high-intensity group. Diamond represents the overall mean difference between studies.

Figure 4. Funnel plot.
Figure 5. Influence of pooled result.

Table 2. Meta-analysis without influential and outlier cases.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>MD (95% CI)</th>
<th>P value</th>
<th>Heterogeneity, $I^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main analysis</td>
<td>0.02 (−0.10 to 0.13)</td>
<td>.77</td>
<td>70</td>
</tr>
<tr>
<td>Outliers removed$^b$</td>
<td>0.0058 (−0.05 to 0.06)</td>
<td>.83</td>
<td>30</td>
</tr>
<tr>
<td>Influential cases removed$^c$</td>
<td>0.02 (−0.03 to 0.07)</td>
<td>.50</td>
<td>23</td>
</tr>
</tbody>
</table>

$^a$MD: mean difference.

$^b$Removed as outliers: Sánchez-González et al [24].

$^c$Removed as influential studies: Sánchez-González et al [24] and Friedenreich et al [19].
Figure 6. Subgroup analysis of the effect of exercise on telomere length. Forest plot of the results of a random-effects meta-analysis shown as mean differences with 95% CI for the comparison of mean telomere length in the exercise group and the control group, performing subgroup analyses for each type of exercise included (resistance training, aerobic exercise, and high-intensity interval training). The shaded square represents the point estimate for each study and the weight of the study in the meta-analysis. Diamond represents the overall mean difference between studies.

### Analysis of Publication Bias

The contour-enhanced funnel plot showed asymmetry, which indicated heterogeneity among the included studies. Most of the studies included in this analysis were not significant; therefore, publication bias was ruled out (Figure 7).

Figure 7. Contour-enhanced funnel plot of the included studies. Dispersion of effect size. x-axis: observed effect sizes. y-axis: inverse SE (higher values on the y-axis represent lower SEs). Slight asymmetry, indicating possible publication bias. Inside to outside (0-2). White region $P>.05$; dark gray region $P<.10$; intermediate gray region $P=.05$; outer gray region $P=.001$. 
Quality of Evidence

Table 3 provides the details of the GRADE assessment. Three levels of evidence were downgraded due to the serious risk of bias and high heterogeneity (inconsistency) of the results, which suggests a very small level of evidence regarding the effects of overall physical exercise modalities on telomere length. In the subgroup analysis, inconsistency was rated as not serious for the 3 exercise modalities, and the level of evidence depended on the risk of bias, being moderate for resistance training and small for aerobic exercise and HIIT.

Table 3. Grading of Recommendations Assessment, Development, and Evaluation assessment.

<table>
<thead>
<tr>
<th>Exercise modality, studies, and sample size</th>
<th>Risk of biasa</th>
<th>Inconsistencyb</th>
<th>Indirectnessc</th>
<th>Imprecisiond</th>
<th>Publication biasc</th>
<th>MDf (95% CI)</th>
<th>Quality of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 trials (n=1058)</td>
<td>Very serious</td>
<td>Serious (I^2=70%)</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Not serious</td>
<td>0.02 (−0.10 to 0.13)</td>
<td>Very small</td>
</tr>
<tr>
<td>Resistance training</td>
<td></td>
<td>No serious (I^2=16%)</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Not serious</td>
<td>−0.01 (−0.07 to 0.04)</td>
<td>Moderate</td>
</tr>
<tr>
<td>4 trials (n=376)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aerobic exercise</td>
<td></td>
<td>Not serious (I^2=0%)</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Not serious</td>
<td>0.01 (−0.04 to 0.06)</td>
<td>Small</td>
</tr>
<tr>
<td>3 trials (n=281)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-intensity interval training</td>
<td></td>
<td>Not serious (I^2=0%)</td>
<td>Not serious</td>
<td>Not serious</td>
<td>Not serious</td>
<td>0.15 (0.03 to 0.26)</td>
<td>Small</td>
</tr>
<tr>
<td>3 trials (n=115)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aNo: most information is from results at low risk of bias; serious: crucial limitation for one criterion or some limitations for multiple criteria sufficient to lower confidence in the estimate of effect; very serious: crucial limitation for one or more criteria sufficient to substantially lower confidence in the estimate of effect.
bSerious: I^2>40%; very serious: I^2>80%.
cNo indirectness of evidence was found in any study.
dOn the basis of sample size. “Serious,” n<250 participants; “very serious,” n<250 and the estimated effect is little or absent.
eOn the basis of funnel plots. No publication bias was found. Funnel plots are not shown because the number of trials was less than 10.
fMD: mean difference.

Discussion

Principal Findings

This meta-analysis aimed to examine the effects of different types of exercise on telomere length in healthy individuals. To date, this is the only study to investigate the impact of different physical exercise modalities in a healthy population. Overall, 9 studies with 1320 participants were eligible; of them, 1299 (98%) were female participants and 91 (2%) were male participants. A total of 199 participants performed resistance exercises, 270 performed aerobic exercises, 73 performed HIIT, and 41 performed mixed exercises.

The pooled effect sizes across all telomere length outcomes showed that exercise did not significantly increase telomere length compared with the control conditions (Figure 3). This finding was robust, with little statistical heterogeneity between studies (I^2=23%). Subgroup analysis suggested that HIIT was the only type of exercise that significantly increased telomere length in exercisers compared with the nonintervention group (MD 0.15, 95% CI 0.03-0.26; P=.01; I^2=0%), with a medium effect size (standardized MD 0.41, 95% CI 0.02-0.8; P=.04; I^2=0%). Meta-regression analyses showed that exercise intensity and duration, year of publication, and methodological quality did not influence the observed effect size. Furthermore, when we compared exercise prescription, considering intensity and duration, with LTL gain, no relationship could be established between LT and exercise intensity and duration. The 2 studies that showed the greatest improvement in LT [17,24] were not those in which the exercise prescription was more intense and...
longer in duration. Similarly, the methodological quality of the studies was not related to the LT gain. The study with the highest methodological quality [19] does not show a significant correlation with LT gain. Therefore, we could not establish a causal relationship between exercise prescription, methodological quality, and LT gain.

To our knowledge, this is the first study to conduct a systematic review and meta-analysis of RCTs to investigate the effects of different physical exercise modalities on telomere length in a healthy population. A recent review by Song et al [15] concluded that aerobic exercise for ≥6 months had a significant effect on the rate of telomere length shortening. However, that review included studies with heterogeneous study populations (breast cancer, polycystic ovarian syndrome, or healthy individuals). This difference in results with our study is probably because in our review, we have only included healthy populations to try to better clarify the possible impact of different physical exercises on telomere length. Telomeric shortening can be accelerated by factors that induce oxidative stress and inflammation [39]; neoplastic processes [40]; psychological disorders [41]; and chronic diseases, such as diabetes or cardiovascular disease [7,42]. Therefore, it seems clear that it is necessary to study the impact of physical exercise in specific populations because of the large number of factors that can influence telomere length.

As previously discussed, our results indicate that HIIT is the type of exercise that appears to have the most beneficial effect on LT. HIIT is characterized by short intermittent bursts of vigorous exercise interspersed with periods of low-intensity recovery [43]. This type of training has sufficient evidence to show that it is a good option for improving cardiovascular health in both healthy individuals and individuals with cardiometabolic diseases [44,45]. However, according to our results, this type of physical exercise significantly increases the length of telomeres, as intense exercise causes an increase in the total oxidative state and external production of free radicals that can lead to oxidative stress [46]. Some studies have suggested that the effects of physical exercise on LT may be represented by an inverted U-shaped dose-response [47,48]. High- or low-intensity levels (too much or too little) may have deleterious effects on the immune system and produce free radicals, thereby accelerating the aging process [49].

The different methodologies used (type of exercise and intensity), time of intervention, lack of homogeneity in the populations studied, and large number of variables that can influence LT could be the cause of the differences in the results of the different studies. Therefore, it is necessary to continue investigating the role of different modalities of physical exercise on LT in different populations by having as much control as possible over the variables that can influence telomere size in RCTs.

Limitations and Recommendations for Future Studies

This study has several limitations. The main limitation was the small number of studies with a small sample size that performed a physical exercise intervention to assess telomere length compared to that of a control group. The included studies were heterogeneous in several aspects. The participants who underwent the interventions were healthy individuals of various ages, the vast majority of whom were women, and this might have influenced the results. The intervention protocol was heterogeneous and included exercise protocols of different intensities and application times (times/wk), some of which were incomplete. Heterogeneity was also present in the main outcome measures, showing disparities among the included studies both at baseline and postintervention measurements, although the methods used for assessing telomere length were the same.

Future research is recommended to evaluate the effects of high-intensity exercise interventions in various healthy age groups to evaluate the effect of these interventions in people with different pathologies and to establish the clinical relationship between the increase in telomere length and variables of clinical relevance.

Recommendations for Clinical Practice

The recommendation to incorporate regular exercise, particularly through HIIT, at least 3 times a week for a sustained period, emphasizes the commitment to the preservation of health and prevention of premature aging.

Conclusions

The findings of this systematic review and meta-analysis suggest that HIIT seems to have a positive effect on telomere length compared to other types of exercise, such as resistance training or aerobic exercise, in a healthy population. The results should be interpreted with caution because of the low quality of evidence.

Data Availability

The original contributions presented in the study are included in the article and supplementary material, and further inquiries can be directed to the corresponding author.

Authors’ Contributions

JLSG and JLSR designed the study, participated in the research, drafted the manuscript, and supervised the study. SVR participated in the operation and drafted the manuscript. RJV participated in the study, revised the article, and supervised the manuscript. RGS participated in the operation and revised the manuscript. CITG contributed to study design, participated in the research, and drafted the manuscript. CRP participated in the operation and drafted the manuscript. VNL and JMV participated in the operation, drafted the manuscript, collected data, and performed the analysis.
Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) checklist.

References


Abbreviations

GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HIIT: high-intensity interval training
MD: mean difference
NRSI: nonrandomized studies of interventions
RCT: randomized controlled trial
RoB2: risk-of-bias tool for randomized trials
ROBINS-I: Risk Of Bias In Nonrandomized Studies of Interventions

©Juan Luis Sánchez-González, Juan Luis Sánchez-Rodríguez, Sergio Varela-Rodríguez, Rogelio González-Sarmiento, Cristina Rivera-Picón, Raúl Juárez-Vela, Clara Isabel Tejada-Garrido, Javier Martín-Vallejo, Víctor Navarro-López. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 09.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.

https://publichealth.jmir.org/2024/1/e46019
JMIR Public Health Surveill 2024 | vol. 10 | e46019 | p.683
(page number not for citation purposes)
Global, Regional, and National Burden of Myocarditis in 204 Countries and Territories From 1990 to 2019: Updated Systematic Analysis

Qingyu Kong1, MD; Xue Xu1, MM; Meng Li1, MD; Xiao Meng2, MD; Cuifen Zhao1*, MD; Xiaorong Yang3,4*, PhD

1Department of Pediatrics, Qilu Hospital of Shandong University, Jinan, Shandong, China
2Department of Cardiology, Qilu Hospital of Shandong University, Jinan, Shandong, China
3Clinical Epidemiology Unit, Qilu Hospital of Shandong University, Jinan, China
4Clinical Research Center of Shandong University, Qilu Hospital, Cheeloo College of Medicine, Jinan, China
*these authors contributed equally

Corresponding Author:
Xiaorong Yang, PhD
Clinical Epidemiology Unit
Qilu Hospital of Shandong University
107 Wenhuaixi Road
Jinan, 250012
China
Phone: 86 53182166672
Fax: 86 53186927544
Email: yangxiaorong@sdu.edu.cn

Abstract

Background: Myocarditis is characterized by high disability and mortality, and imposes a severe burden on population health globally. However, the latest global magnitude and secular trend of myocarditis burden have not been reported.

Objective: This study aimed to delineate the epidemiological characteristics of myocarditis burden globally for optimizing targeted prevention and research.

Methods: Based on the Global Burden of Disease Study 2019, the myocarditis burden from 1990 to 2019 was modeled using the Cause of Death Ensemble tool, DisMod-MR, and spatiotemporal Gaussian regression. We depicted the epidemiology and trends of myocarditis by sex, age, year, region, and sociodemographic index (SDI). R program version 4.2.1 (R Project for Statistical Computing) was applied for all statistical analyses, and a 2-sided P-value of <.05 was considered statistically significant.

Results: The number of incident cases (1,268,000) and deaths (32,450) associated with myocarditis in 2019 increased by over 1.6 times compared with the values in 1990 globally. On the other hand, the age-standardized incidence rate (ASIR) and age-standardized mortality rate (ASMR) decreased slightly from 1990 to 2019. The disability-adjusted life years (DALYs) decreased slightly in the past 3 decades, while the age-standardized DALY rate (ADSR) decreased greatly from 18.29 per 100,000 person-years in 1990 to 12.81 per 100,000 person-years in 2019. High SDI regions always showed a more significant ASIR. The ASIR slightly decreased in all SDI regions between 1990 and 2019. Middle SDI regions had the highest ASMR and ADSR in 2019. Low SDI regions had the lowest ASMR and ADSR in 2019. The age-standardized rates (ASRs) of myocarditis were higher among males than among females from 1990 to 2019 globally. All ASRs among both sexes had a downward trend, except for the ASMR among males, which showed a stable trend, and females had a more significant decrease in the ADSR than males. Senior citizens had high incident cases and deaths among both sexes in 2019. The peak numbers of DALYs for both sexes were noted in the under 1 age group in 2019. At the national level, the estimated annual percentage changes in the ASRs had significant negative correlations with the baseline ASRs in 1990.

Conclusions: Globally, the number of incident cases and deaths associated with myocarditis have increased significantly. On the other hand, the ASRs of myocarditis showed decreasing trends from 1990 to 2019. Males consistently showed higher ASRs of myocarditis than females from 1990 to 2019 globally. Senior citizens gradually predominated in terms of myocarditis burden. Policymakers should establish targeted control strategies based on gender, region, age, and SDI; strengthen aging-related health research; and take notice of the changes in the epidemic characteristics of myocarditis.
myocarditis; global burden; temporal trend; systematic analysis; incidence; mortality; disability-adjusted life years

Introduction

Cardiovascular diseases can lead to high disability and mortality, which impose severe burdens on population health globally [1,2]. Among cardiovascular diseases, myocarditis has variable clinical presentations and outcomes, including chest pain with uncomplicated clinical manifestations, new onset or worsening heart failure, chronic dilated cardiomyopathy, and sudden death [3]. Myocarditis caused up to 9% of sudden deaths among cardiovascular events in athletes in the United States [4]. Moreover, 16% to 20% of cases of sudden infant death syndrome were documented to be associated with myocarditis [5-7]. Globally, 46,490 deaths were estimated to be due to myocarditis in 2017 [8]. Myocarditis characterized by high mortality represents an enormous public health burden.

Data for specific clinical settings are available to estimate the burden of myocarditis. Data showed that 3% of cases of chest pain in the emergency department were attributed to acute myocarditis and pericarditis [9]. In a prospective registry of northeastern Italy, autopsy studies showed a 12% incidence of myocarditis among young people who died suddenly [10]. The prevalence of myocarditis was reported to be 1.14% among patients with advanced cancers after therapy with immune checkpoint inhibitors (ICIs) [11]. The prevalence rate of myocarditis in hospitalized patients infected with COVID-19 was reported to be 2.4 per 1000 patients, considering definite or probable cases [12]. Hospital mortality was found to be much higher among patients with myocarditis and COVID-19 than among those with myocarditis but without COVID-19 [13]. The nationwide incidence of myocarditis in Israel due to the BNT162b2 mRNA COVID-19 vaccine was 2.7 per 100,000 persons [14]. In the last 2 decades, the diagnosis of myocarditis has become reasonably accurate among patients, including those from low-risk populations, through the use of noninvasive examinations, including high-sensitive troponin levels and cardiac magnetic resonance imaging (CMRI) [15,16]. In the United States, the incidence of myocarditis showed a gradual increase from 95 (in 2005) to 144 (in 2014) per 1 million persons [17]. It is quite clear that many factors can influence the global burden of myocarditis.

The Global Burden of Disease Study (GBD) 2019 is the most recent data set for evaluating the epidemiological levels and trends of 369 diseases along with 87 risk factors globally, and it has incorporated new data sets and modeling strategies compared with the GBD 2017 [18,19]. For example, the only country-level covariate of myocarditis used for the GBD 2019 was the Healthcare Access and Quality Index on excess mortality, which differs from the GBD 2017. Since the evaluation of the global burden of myocarditis using data from the GBD 2017, no comprehensive statistics on the global epidemiological trend of myocarditis have been calculated [8]. Due to the inclusion of more raw data and the application of more robust statistical methods in the GBD 2019, the results for myocarditis burden in the GBD 2019 differ dramatically from those in the GBD 2017, indicating that new statistics and analyses of the global burden of myocarditis need to be reported timely to guide its prevention and control. This research summarizes the latest epidemiological characteristics of myocarditis, including incidence, death, disability-adjusted life years (DALYS), and changing trends by sex, age, and sociodemographic index (SDI), in 204 countries and territories from 1990 to 2019, using data from the GBD 2019. Our results will help to optimize the niche-targeting prevention and intervention of myocarditis that rely on concrete characteristics across the world.

Methods

Data Source

Previous studies have described the primary data explanations and analytic approaches of the GBD 2019 in detail [18-20]. Researchers can extract the reproducible statistical codes and analysis process online [21]. We briefly introduce the estimation methods specific to myocarditis. The Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) were followed to analyze the GBD database in every step (Report Checklist)[22]. Myocarditis was defined as a clinical diagnosis for the GBD 2019 estimation [2,19]. The International Classification of Diseases version 9 (ICD-9) and version 10 (ICD-10) codes were adopted to identify myocarditis. The determination of myocarditis was based on the disease codes 422-422.9 in ICD-9 and B33.2, I40-I41.9, and I51.4 in ICD-10. To assess the myocarditis burden, more than 250 primary data sources were screened out according to the GBD inclusion criteria. The myocarditis burden from 1990 to 2019 was modeled using the Cause of Death Ensemble tool, DisMod-MR, and spatiotemporal Gaussian regression. Data were obtained from the Institute for Health Metrics and Evaluation to characterize the global burden of myocarditis by sex and 5-year age group across the world in the past 3 decades. According to the SDI (a comprehensive index based on income per person, years of education, and fertility), the world is divided into 5 SDI regions (high, high-middle, middle, low-middle, and low) for assessing the myocarditis burden across different geographies. Furthermore, 204 countries and territories were divided into 21 GBD regions, including Western Europe, Tropical Latin America, and East Asia.

Ethical Considerations

The GBD 2019 is a publicly available database, and all data were anonymous. Our study protocol was approved by the Institutional Review Board of Qilu Hospital of Shandong University with approval number KYLL-202011(KS)-239.

Statistical Analysis

The burden of myocarditis was appraised based on the age-standardized incidence rate (ASIR), age-standardized...
mortality rate (ASMR), and age-standardized DALY rate (ASDR) by calendar year, sex, and region. The estimated annual percentage change (EAPC) was determined based on a regression calculation by fitting the natural logarithm of the age-standardized rate (ASR) with the historical year to characterize the long-term trend in the ASR of myocarditis burden [23]. The calculation formula is as follows:

\[
\ln(\text{ASR}) = \alpha + \beta \times \text{historical year} + \epsilon \tag{1}
\]

The EAPC can be used to depict the changing trends of the ASR in a specific population and a certain time interval. The EAPC and its 95% CI are calculated using the following formula:

\[
100 \times (\exp(\beta) - 1) \tag{2}
\]

An increasing trend was recognized when the EAPC and the minimum of the 95% CI were positive. On the contrary, a decreasing trend was recognized when the EAPC and the maximum of the 95% CI were negative. Otherwise, the trend of the ASR was considered to be stable. The Spearman rank correlation with \(\rho\) coefficient was used to estimate the influence of the baseline ASR in 1990 and the SDI in 2019 all over the world on the EAPC in myocarditis burden. R program version 4.2.1 (R Project for Statistical Computing) was used for all statistical analyses, and a 2-sided \(P\)-value of < .05 was considered statistically significant.

**Results**

**Global Burden and Temporal Trend of Myocarditis**

Across the world, the number of incident cases of myocarditis increased from 780,400 (95% uncertainty interval [UI] 620,600-951,200) in 1990 to 1,265,800 (95% UI 1,021,700-1,531,500) in 2019, with an increase of 62.20%. On the other hand, the ASIR decreased slightly from 16.74 (95% UI 13.46-20.34) to 16.00 (95% UI 13.01-19.28) per 100,000 person-years over the 3 decades, with an EAPC of −0.23 (95% CI −0.26 to −0.21) (Table 1; Figure 1). Simultaneously, the number of worldwide deaths caused by myocarditis increased from 19,620 (95% UI 15,690-26,770) in 1990 to 32,450 (95% UI 23,160-37,090) in 2019, with an increase of 65.39%. On the other hand, the ASMR decreased marginally from 0.46 (95% UI 0.38-0.60) to 0.43 (95% UI 0.31-0.50) per 100,000 person-years over the 3 decades, with an EAPC of −0.09 (95% CI −0.39 to 0.21) (Table 2; Figure 1). Globally, the DALYs decreased slightly. However, the ASDR decreased greatly from 18.29 (95% UI 13.81-27.58) per 100,000 person-years in 1990 to 12.81 (95% UI 10.53-14.72) per 100,000 person-years in 2019, with an EAPC of −1.19 (95% CI −1.33 to −1.04) (Table 3; Figure 1).
Table 1. Incidence and age-standardized incidence rate of myocarditis in 1990 and 2019, and the estimated annual percentage change from 1990 to 2019.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>1990</th>
<th>2019</th>
<th>EAPC in the ASIR from 1990 to 2019, value (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASIR/100,000 (95% UI)</td>
<td>Incident cases×10^4 (95% UI)</td>
<td>ASIR/100,000 (95% UI)</td>
</tr>
<tr>
<td><strong>Global</strong></td>
<td>16.74 (13.46 to 20.34)</td>
<td>78.04 (62.06 to 95.12)</td>
<td>16.00 (13.01 to 19.28)</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>20.06 (16.21 to 24.31)</td>
<td>45.55 (36.39 to 55.72)</td>
<td>19.06 (15.51 to 22.89)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>13.66 (11.04 to 16.67)</td>
<td>32.49 (25.87 to 39.72)</td>
<td>13.10 (10.62 to 15.84)</td>
</tr>
<tr>
<td><strong>SDI region</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High</strong></td>
<td>18.72 (15.07 to 22.80)</td>
<td>17.26 (13.85 to 21.06)</td>
<td>17.75 (14.71 to 21.22)</td>
</tr>
<tr>
<td><strong>High-middle</strong></td>
<td>16.58 (13.32 to 20.18)</td>
<td>18.29 (14.53 to 22.28)</td>
<td>16.01 (12.94 to 19.34)</td>
</tr>
<tr>
<td><strong>Middle</strong></td>
<td>16.39 (13.25 to 19.96)</td>
<td>22.98 (18.20 to 28.36)</td>
<td>15.79 (12.79 to 19.10)</td>
</tr>
<tr>
<td><strong>Low-middle</strong></td>
<td>15.93 (12.90 to 19.29)</td>
<td>13.71 (10.95 to 16.81)</td>
<td>15.68 (12.73 to 18.97)</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>15.35 (12.43 to 18.68)</td>
<td>5.77 (4.60 to 7.12)</td>
<td>15.29 (12.38 to 18.59)</td>
</tr>
<tr>
<td><strong>GBD region</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High-income Asia Pacific</strong></td>
<td>20.66 (16.78 to 24.99)</td>
<td>3.76 (3.02 to 4.58)</td>
<td>20.07 (16.38 to 24.09)</td>
</tr>
<tr>
<td><strong>High-income North America</strong></td>
<td>19.91 (16.01 to 24.29)</td>
<td>6.36 (5.08 to 7.78)</td>
<td>18.21 (15.37 to 21.45)</td>
</tr>
<tr>
<td><strong>Western Europe</strong></td>
<td>17.56 (14.19 to 21.33)</td>
<td>8.14 (6.55 to 9.93)</td>
<td>17.47 (14.21 to 21.14)</td>
</tr>
<tr>
<td><strong>Australasia</strong></td>
<td>16.51 (13.25 to 19.98)</td>
<td>0.36 (0.29 to 0.44)</td>
<td>16.45 (13.31 to 19.83)</td>
</tr>
<tr>
<td><strong>Southern Latin America</strong></td>
<td>15.34 (12.43 to 18.64)</td>
<td>0.72 (0.58 to 0.88)</td>
<td>15.36 (12.45 to 18.66)</td>
</tr>
<tr>
<td><strong>Andean Latin America</strong></td>
<td>14.05 (11.38 to 17.11)</td>
<td>0.40 (0.32 to 0.49)</td>
<td>14.07 (11.41 to 17.10)</td>
</tr>
<tr>
<td><strong>Tropical Latin America</strong></td>
<td>15.76 (12.72 to 19.19)</td>
<td>1.90 (1.52 to 2.33)</td>
<td>15.72 (12.68 to 19.15)</td>
</tr>
<tr>
<td><strong>Central Latin America</strong></td>
<td>15.06 (12.18 to 18.32)</td>
<td>1.84 (1.48 to 2.26)</td>
<td>14.97 (12.11 to 18.20)</td>
</tr>
<tr>
<td><strong>Caribbean</strong></td>
<td>14.21 (11.50 to 17.19)</td>
<td>0.43 (0.35 to 0.52)</td>
<td>14.19 (11.48 to 17.16)</td>
</tr>
<tr>
<td><strong>Eastern Europe</strong></td>
<td>17.64 (14.19 to 21.42)</td>
<td>4.27 (3.40 to 5.19)</td>
<td>17.74 (14.28 to 21.53)</td>
</tr>
<tr>
<td><strong>Central Europe</strong></td>
<td>16.30 (13.13 to 19.83)</td>
<td>2.11 (1.68 to 2.58)</td>
<td>16.48 (13.27 to 20.05)</td>
</tr>
<tr>
<td><strong>Central Asia</strong></td>
<td>14.84 (11.86 to 17.95)</td>
<td>0.88 (0.70 to 1.07)</td>
<td>14.87 (11.88 to 17.98)</td>
</tr>
<tr>
<td><strong>North Africa and Middle East</strong></td>
<td>12.01 (9.68 to 14.61)</td>
<td>3.14 (2.46 to 3.91)</td>
<td>12.05 (9.72 to 14.66)</td>
</tr>
<tr>
<td><strong>South Asia</strong></td>
<td>16.28 (13.20 to 19.69)</td>
<td>13.40 (10.6 to 16.52)</td>
<td>16.18 (13.12 to 19.58)</td>
</tr>
<tr>
<td><strong>Southeast Asia</strong></td>
<td>15.14 (12.27 to 18.41)</td>
<td>5.71 (4.54 to 7.03)</td>
<td>15.14 (12.26 to 18.38)</td>
</tr>
<tr>
<td><strong>East Asia</strong></td>
<td>17.92 (14.47 to 21.97)</td>
<td>19.2 (15.15 to 23.81)</td>
<td>16.86 (13.73 to 20.36)</td>
</tr>
<tr>
<td>Characteristic</td>
<td>1990</td>
<td></td>
<td>2019</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>ASIR/100,000 (95% UI)</td>
<td>Incident cases×10&lt;sup&gt;d&lt;/sup&gt; (95% UI)</td>
<td>ASIR/100,000 (95% UI)</td>
</tr>
<tr>
<td>Oceania</td>
<td>15.30 (12.32 to 18.49)</td>
<td>0.08 (0.06 to 0.09)</td>
<td>15.30 (12.31 to 18.50)</td>
</tr>
<tr>
<td>Western Sub-Saharan Africa</td>
<td>15.48 (12.54 to 18.87)</td>
<td>2.14 (1.71 to 2.63)</td>
<td>15.37 (12.45 to 18.71)</td>
</tr>
<tr>
<td>Eastern Sub-Saharan Africa</td>
<td>15.26 (12.33 to 18.61)</td>
<td>2.00 (1.58 to 2.47)</td>
<td>15.25 (12.32 to 18.59)</td>
</tr>
<tr>
<td>Central Sub-Saharan Africa</td>
<td>14.59 (11.72 to 17.82)</td>
<td>0.56 (0.44 to 0.69)</td>
<td>14.52 (11.67 to 17.70)</td>
</tr>
<tr>
<td>Southern Sub-Saharan Africa</td>
<td>15.72 (12.72 to 19.06)</td>
<td>0.63 (0.51 to 0.78)</td>
<td>15.73 (12.72 to 19.08)</td>
</tr>
</tbody>
</table>

<sup>a</sup>EAPC: estimated annual percentage change.
<sup>b</sup>ASIR: age-standardized incidence rate.
<sup>c</sup>UI: uncertainty interval.
<sup>d</sup>SDI: sociodemographic index.
<sup>e</sup>GBD: Global Burden of Disease Study.
Figure 1. Absolute counts and age-standardized rates of the global myocarditis burden by sex/SDI region from 1990 to 2019. (A) Incidence; (B) Mortality; (C) DALYs. ASDR: age-standardized DALY rate; ASIR: age-standardized incidence rate; ASMR: age-standardized mortality rate; DALYs: disability-adjusted life years; SDI: sociodemographic index.
Table 2. Deaths and age-standardized mortality rate of myocarditis in 1990 and 2019, and the estimated annual percentage change from 1990 to 2019.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>1990</th>
<th>2019</th>
<th>EAPC(^a) in the ASMR(^b) from 1990 to 2019, value (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASMR/100,000 (95% UI)</td>
<td>Deaths x 10(^3) (95% UI)</td>
<td>ASMR/100,000 (95% UI)</td>
</tr>
<tr>
<td>Global</td>
<td>0.46 (0.38 to 0.60)</td>
<td>19.62 (15.69 to 26.77)</td>
<td>0.43 (0.31 to 0.50)</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>1990</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.50 (0.41 to 0.67)</td>
<td>10.18 (8.02 to 14.38)</td>
<td>0.48 (0.32 to 0.57)</td>
</tr>
<tr>
<td>Female</td>
<td>0.42 (0.30 to 0.59)</td>
<td>9.44 (6.40 to 14.25)</td>
<td>0.38 (0.25 to 0.45)</td>
</tr>
<tr>
<td>SDI(^d) region</td>
<td>2019</td>
<td>1990</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>0.26 (0.21 to 0.33)</td>
<td>2.32 (1.87 to 2.93)</td>
<td>0.27 (0.21 to 0.31)</td>
</tr>
<tr>
<td>High-middle</td>
<td>0.59 (0.45 to 0.74)</td>
<td>5.50 (4.36 to 7.15)</td>
<td>0.56 (0.40 to 0.65)</td>
</tr>
<tr>
<td>Middle</td>
<td>0.67 (0.54 to 0.94)</td>
<td>7.80 (6.16 to 11.9)</td>
<td>0.59 (0.35 to 0.72)</td>
</tr>
<tr>
<td>Low-middle</td>
<td>0.39 (0.28 to 0.47)</td>
<td>2.85 (1.69 to 3.93)</td>
<td>0.41 (0.24 to 0.50)</td>
</tr>
<tr>
<td>Low</td>
<td>0.31 (0.21 to 0.44)</td>
<td>1.14 (0.57 to 1.91)</td>
<td>0.24 (0.15 to 0.37)</td>
</tr>
<tr>
<td>GBD(^e) region</td>
<td>2019</td>
<td>1990</td>
<td></td>
</tr>
<tr>
<td>High-income Asia Pacific</td>
<td>0.33 (0.22 to 0.38)</td>
<td>0.53 (0.38 to 0.60)</td>
<td>0.21 (0.16 to 0.25)</td>
</tr>
<tr>
<td>High-income North America</td>
<td>0.17 (0.14 to 0.25)</td>
<td>0.48 (0.40 to 0.73)</td>
<td>0.26 (0.18 to 0.30)</td>
</tr>
<tr>
<td>Western Europe</td>
<td>0.25 (0.18 to 0.32)</td>
<td>1.20 (0.85 to 1.54)</td>
<td>0.34 (0.19 to 0.43)</td>
</tr>
<tr>
<td>Australasia</td>
<td>0.36 (0.3 to 0.48)</td>
<td>0.07 (0.06 to 0.10)</td>
<td>0.30 (0.25 to 0.42)</td>
</tr>
<tr>
<td>Southern Latin America</td>
<td>0.28 (0.19 to 0.35)</td>
<td>0.12 (0.09 to 0.15)</td>
<td>0.24 (0.19 to 0.32)</td>
</tr>
<tr>
<td>Andean Latin America</td>
<td>0.27 (0.15 to 0.37)</td>
<td>0.06 (0.03 to 0.09)</td>
<td>0.12 (0.09 to 0.16)</td>
</tr>
<tr>
<td>Tropical Latin America</td>
<td>0.28 (0.23 to 0.41)</td>
<td>0.31 (0.24 to 0.44)</td>
<td>0.30 (0.25 to 0.46)</td>
</tr>
<tr>
<td>Central Latin America</td>
<td>0.12 (0.09 to 0.13)</td>
<td>0.14 (0.11 to 0.17)</td>
<td>0.14 (0.09 to 0.17)</td>
</tr>
<tr>
<td>Caribbean</td>
<td>0.38 (0.24 to 0.66)</td>
<td>0.12 (0.07 to 0.24)</td>
<td>0.36 (0.24 to 0.56)</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>0.41 (0.24 to 0.53)</td>
<td>0.85 (0.55 to 1.06)</td>
<td>0.53 (0.33 to 0.67)</td>
</tr>
<tr>
<td>Central Europe</td>
<td>1.05 (0.69 to 1.34)</td>
<td>1.24 (0.79 to 1.58)</td>
<td>1.12 (0.80 to 1.36)</td>
</tr>
<tr>
<td>Central Asia</td>
<td>0.47 (0.38 to 0.64)</td>
<td>0.25 (0.21 to 0.35)</td>
<td>0.78 (0.61 to 1.11)</td>
</tr>
<tr>
<td>North Africa and Middle East</td>
<td>0.52 (0.34 to 0.85)</td>
<td>1.60 (0.94 to 3.31)</td>
<td>0.36 (0.26 to 0.65)</td>
</tr>
<tr>
<td>South Asia</td>
<td>0.24 (0.16 to 0.33)</td>
<td>1.38 (0.97 to 1.82)</td>
<td>0.21 (0.15 to 0.29)</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>0.42 (0.28 to 0.53)</td>
<td>1.22 (0.84 to 1.88)</td>
<td>0.36 (0.23 to 0.44)</td>
</tr>
<tr>
<td>East Asia</td>
<td>1.07 (0.81 to 1.68)</td>
<td>8.90 (7.10 to 13.44)</td>
<td>0.90 (0.53 to 1.11)</td>
</tr>
<tr>
<td>Characteristic</td>
<td>1990 ASMR/100,000 (95% UI)</td>
<td>Deaths×10^3 (95% UI)</td>
<td>2019 ASMR/100,000 (95% UI)</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>---------------------------</td>
<td>----------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Oceania</td>
<td>0.30 (0.17 to 0.44)</td>
<td>0.02 (0.01 to 0.03)</td>
<td>0.33 (0.18 to 0.50)</td>
</tr>
<tr>
<td>Western Sub-Saharan Africa</td>
<td>0.41 (0.17 to 0.70)</td>
<td>0.34 (0.15 to 0.55)</td>
<td>0.24 (0.17 to 0.31)</td>
</tr>
<tr>
<td>Eastern Sub-Saharan Africa</td>
<td>0.24 (0.11 to 0.37)</td>
<td>0.51 (0.20 to 0.96)</td>
<td>0.14 (0.05 to 0.33)</td>
</tr>
<tr>
<td>Central Sub-Saharan Africa</td>
<td>0.35 (0.21 to 0.53)</td>
<td>0.15 (0.06 to 0.36)</td>
<td>0.24 (0.08 to 0.50)</td>
</tr>
<tr>
<td>Southern Sub-Saharan Africa</td>
<td>0.30 (0.20 to 0.38)</td>
<td>0.12 (0.07 to 0.17)</td>
<td>0.22 (0.16 to 0.33)</td>
</tr>
</tbody>
</table>

<sup>a</sup>EAPC: estimated annual percentage change.

<sup>b</sup>ASMR: age-standardized mortality rate.

<sup>c</sup>UI: uncertainty interval.

<sup>d</sup>SDI: sociodemographic index.

<sup>e</sup>GBD: Global Burden of Disease Study.
Table 3. Disability-adjusted life years (DALYs) and age-standardized DALY rate of myocarditis in 1990 and 2019, and the estimated annual percentage change from 1990 to 2019.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>1990 ASDR/100,000 (95% CI)</th>
<th>2019 ASDR/100,000 (95% CI)</th>
<th>EAPC(^a) in the AS-DR(^b) from 1990 to 2019, value (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td></td>
<td></td>
<td>−1.19 (−1.33 to −1.04)</td>
</tr>
<tr>
<td>Male</td>
<td>18.29 (13.81 to 27.58)</td>
<td>12.81 (10.53 to 14.72)</td>
<td>97.72 (80.38 to 112.68)</td>
</tr>
<tr>
<td>Female</td>
<td>20.14 (15.47 to 29.74)</td>
<td>15.21 (11.44 to 18.35)</td>
<td>57.03 (43.10 to 69.38)</td>
</tr>
<tr>
<td>SDI(^e) region</td>
<td></td>
<td></td>
<td>−1.54 (−1.71 to −1.36)</td>
</tr>
<tr>
<td>High</td>
<td>11.97 (10.52 to 15.53)</td>
<td>11.63 (9.24 to 12.97)</td>
<td>12.47 (9.76 to 14.00)</td>
</tr>
<tr>
<td>High-middle</td>
<td>21.17 (17.74 to 29.91)</td>
<td>15.39 (12.24 to 18.54)</td>
<td>23.62 (18.89 to 28.37)</td>
</tr>
<tr>
<td>Middle</td>
<td>25.58 (19.32 to 41.28)</td>
<td>15.81 (12.12 to 19.05)</td>
<td>34.62 (26.93 to 42.53)</td>
</tr>
<tr>
<td>Low-middle</td>
<td>14.05 (8.24 to 19.74)</td>
<td>11.51 (8.27 to 13.44)</td>
<td>17.90 (13.17 to 21.06)</td>
</tr>
<tr>
<td>Low</td>
<td>12.08 (6.69 to 18.73)</td>
<td>8.76 (6.37 to 12.81)</td>
<td>9.06 (6.48 to 12.81)</td>
</tr>
<tr>
<td>GBD(^f) region</td>
<td></td>
<td></td>
<td>−1.10 (−1.19 to −1.02)</td>
</tr>
<tr>
<td>High-income Asia Pacific</td>
<td>14.81 (12.35 to 18.98)</td>
<td>9.54 (8.26 to 12.05)</td>
<td>1.94 (1.53 to 2.27)</td>
</tr>
<tr>
<td>High-income North America</td>
<td>11.02 (9.18 to 15.42)</td>
<td>14.60 (10.72 to 16.65)</td>
<td>4.93 (3.57 to 5.67)</td>
</tr>
<tr>
<td>Western Europe</td>
<td>8.40 (6.42 to 11.05)</td>
<td>8.74 (5.71 to 10.29)</td>
<td>5.47 (3.40 to 6.65)</td>
</tr>
<tr>
<td>Australasia</td>
<td>18.11 (15.31 to 21.14)</td>
<td>13.34 (11.1 to 18.06)</td>
<td>0.42 (0.35 to 0.58)</td>
</tr>
<tr>
<td>Southern Latin America</td>
<td>11.36 (8.69 to 14.82)</td>
<td>7.91 (6.47 to 10.93)</td>
<td>0.55 (0.45 to 0.74)</td>
</tr>
<tr>
<td>Andean Latin America</td>
<td>7.98 (4.06 to 11.17)</td>
<td>3.49 (2.62 to 4.65)</td>
<td>0.21 (0.16 to 0.28)</td>
</tr>
<tr>
<td>Tropical Latin America</td>
<td>12.35 (9.06 to 17.42)</td>
<td>11.02 (8.83 to 16.26)</td>
<td>2.26 (1.84 to 3.33)</td>
</tr>
<tr>
<td>Central Latin America</td>
<td>4.89 (3.79 to 5.92)</td>
<td>5.25 (3.92 to 6.77)</td>
<td>1.25 (0.95 to 1.61)</td>
</tr>
<tr>
<td>Caribbean</td>
<td>18.76 (9.05 to 45.08)</td>
<td>17.09 (9.31 to 33.57)</td>
<td>0.76 (0.44 to 1.43)</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>11.3 (9.31 to 16.17)</td>
<td>16.52 (12.98 to 22.05)</td>
<td>4.05 (3.00 to 5.09)</td>
</tr>
<tr>
<td>Central Europe</td>
<td>27.56 (18.86 to 33.05)</td>
<td>25.68 (17.61 to 31.76)</td>
<td>4.28 (2.86 to 5.38)</td>
</tr>
<tr>
<td>Central Asia</td>
<td>17.71 (14.63 to 24.7)</td>
<td>26.7 (20.54 to 41.35)</td>
<td>2.43 (1.83 to 3.87)</td>
</tr>
<tr>
<td>North Africa and Middle East</td>
<td>25.72 (15.08 to 54.1)</td>
<td>15.28 (10.67 to 27.02)</td>
<td>8.70 (5.99 to 15.50)</td>
</tr>
<tr>
<td>South Asia</td>
<td>7.16 (5.05 to 9.45)</td>
<td>6.64 (4.10 to 9.52)</td>
<td>10.84 (8.53 to 14.01)</td>
</tr>
</tbody>
</table>

\(^a\) EAPC = Estimated Annual Percentage Change
\(^b\) AS-DR = Age-standardized Disability-adjusted Life Years
\(^c\) UI = Uncertainty Interval
\(^d\) DALYs = Disability-adjusted Life Years
\(^e\) SDI = Socioeconomic Development Index
\(^f\) GBD = Global Burden of Disease
regions from 1990 to 2019 (Table 1). The largest 3 increases in myocarditis changed slightly or remained stable across all GBD regions. The ASIR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019: 18.72 and 16.58 per 100,000 person-years in 1990, and 17.75 and 16.01 per 100,000 person-years in 2019, respectively (Table 1). The leading ASMR in 2019 was recorded in the United States, with an EAPC of 0.13. On the contrary, the fastest decline in the ASIR was noted in Qatar, with an EAPC of −1.52 (Table 3).

The ASMR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019: 18.72 and 16.58 per 100,000 person-years in 1990, and 17.75 and 16.01 per 100,000 person-years in 2019. The ASIR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019. The ASIR slightly decreased in all SDI regions between 1990 and 2019 (Table 1). The ASIR was noted in the United States, with an EAPC of 0.13. On the contrary, the fastest decline in the ASIR was noted in Qatar, with an EAPC of −1.52 (Table 3).

The ASMR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019: 18.72 and 16.58 per 100,000 person-years in 1990, and 17.75 and 16.01 per 100,000 person-years in 2019. The ASIR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019. The ASIR slightly decreased in all SDI regions between 1990 and 2019 (Table 1). The ASIR was noted in the United States, with an EAPC of 0.13. On the contrary, the fastest decline in the ASIR was noted in Qatar, with an EAPC of −1.52 (Table 3).

The ASMR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019: 18.72 and 16.58 per 100,000 person-years in 1990, and 17.75 and 16.01 per 100,000 person-years in 2019. The ASIR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019. The ASIR slightly decreased in all SDI regions between 1990 and 2019 (Table 1). The ASIR was noted in the United States, with an EAPC of 0.13. On the contrary, the fastest decline in the ASIR was noted in Qatar, with an EAPC of −1.52 (Table 3).

The ASMR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019: 18.72 and 16.58 per 100,000 person-years in 1990, and 17.75 and 16.01 per 100,000 person-years in 2019. The ASIR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019. The ASIR slightly decreased in all SDI regions between 1990 and 2019 (Table 1). The ASIR was noted in the United States, with an EAPC of 0.13. On the contrary, the fastest decline in the ASIR was noted in Qatar, with an EAPC of −1.52 (Table 3).

The ASMR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019: 18.72 and 16.58 per 100,000 person-years in 1990, and 17.75 and 16.01 per 100,000 person-years in 2019. The ASIR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019. The ASIR slightly decreased in all SDI regions between 1990 and 2019 (Table 1). The ASIR was noted in the United States, with an EAPC of 0.13. On the contrary, the fastest decline in the ASIR was noted in Qatar, with an EAPC of −1.52 (Table 3).

The ASMR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019: 18.72 and 16.58 per 100,000 person-years in 1990, and 17.75 and 16.01 per 100,000 person-years in 2019. The ASIR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019. The ASIR slightly decreased in all SDI regions between 1990 and 2019 (Table 1). The ASIR was noted in the United States, with an EAPC of 0.13. On the contrary, the fastest decline in the ASIR was noted in Qatar, with an EAPC of −1.52 (Table 3).

Variation in the Myocarditis Burden at Regional and National Levels

The ASIR of myocarditis was always the highest in high and high-middle SDI regions from 1990 to 2019: 18.72 and 16.58 per 100,000 person-years in 1990, and 17.75 and 16.01 per 100,000 person-years in 2019, respectively (Table 1). Low SDI regions had the lowest ASIR of 15.29 per 100,000 person-years in 2019. The ASIR slightly decreased in all SDI regions between 1990 and 2019 (Table 1). The leading ASMR in 2019 was observed in middle SDI regions at 0.59/100,000, and high-middle SDI regions ranked second at 0.56/100,000. Low SDI regions had the lowest ASMR of 0.24/100,000 in 2019. The largest increase in the ASMR was observed in low-middle SDI regions, followed by high SDI regions. On the other hand, the other 3 regions presented downward trends in the ASMR (lowest EAPC of −0.97 in low SDI regions) (Table 2). The highest ASDR in 2019 was observed in middle SDI regions, and the lowest was observed in low SDI regions. The ASDRs in 5 SDI regions all dropped from 1990 to 2019, with the highest decrease in middle SDI regions (EAPC=−1.52) (Table 3).

With respect to 21 GBD regions, high-income Asia Pacific, high-income North America, and Eastern Europe were the top 3 regions with the highest ASIRs in 2019 (range from 17.74/100,000 to 20.07/100,000). On the contrary, the North Africa and Middle East, Andean Latin America, and Caribbean GBD regions had the lowest ASIRs in 2019 (range from 12.05/100,000 to 14.19/100,000) (Table 1). The ASIR of myocarditis changed slightly or remained stable across all GBD regions from 1990 to 2019 (Table 1). The largest 3 increases in the ASMR were observed in Central Asia, high-income North America, and Western Europe. On the contrary, the largest 3 decreases in the ASMR were observed in Andean Latin America, Western Sub-Saharan Africa, and Eastern Sub-Saharan Africa (Table 2). The changing trends in the ASMR were different across all GBD regions (Table 2). The highest 3 ASDRs in 2019 appeared in Central Asia, Central Europe, and East Asia (range from 25.17 to 26.70 per 100,000 person-years). On the contrary, the lowest 3 ASDRs in 2019 appeared in Andean Latin America, Central Latin America, and South Asia (range from 3.49 to 6.64 per 100,000 person-years) (Table 3). There were significant regional differences in the trends of ASDRs across all the GBD regions, with the most obvious increase in Central Asia (EAPC=2.05) and the largest decrease in Andean Latin America (EAPC=−3.16) (Table 3).

In 2019, there were 3 countries with an ASIR exceeding 20/100,000, including Austria, Japan, and Sweden, while Lebanon had the lowest ASIR (Multimedia Appendix 1; Figure 2). The difference in the ASMR of myocarditis was nearly 60 times across the world in 2019, with Romania showing the highest value (3.32/100,000) and Tajikistan showing the lowest value (0.05/100,000) (Multimedia Appendix 1; Figure 3). Similarly, the difference in the ASDR of myocarditis was nearly 30 times across the world in 2019, with Romania showing the highest value and Tajikistan showing the lowest value (Multimedia Appendix 1; Figure 3). From 1990 to 2019, 100 out of 204 countries and territories showed a rising ASIR. The largest annual increment in the ASIR was noted in Qatar, with an EAPC of 0.13. On the contrary, the fastest decline in the ASIR was noted in the United States, with an EAPC of −0.93.
(Multimedia Appendix 2; Figure 2). The EAPC in the ASMR was the highest in Kazakhstan (EAPC=9.93) and the lowest in Ghana (EAPC=−4.94) from 1990 to 2019 (Multimedia Appendix 2; Figure 3). The EAPC in the ASDR was the highest in Kazakhstan (EAPC=8.70) and the lowest in Serbia (EAPC=−4.70) from 1990 to 2019 (Multimedia Appendix 2; Figure 3).

**Figure 2.** Maps of the global incidence and temporal trends of myocarditis in 204 countries and territories. (A) ASIR of myocarditis around the world in 2019; (B) EAPC in the ASIR from 1990 to 2019. ASIR: age-standardized incidence rate; EAPC: estimated annual percentage change.
Variation in the Myocarditis Burden in Both Genders and 5-Year Age Groups

From 1990 to 2019, the ASIR of myocarditis among men was higher than that among women (19.06/100,000 vs 13.01/100,000 in 2019) (Table 1). Similar to the finding of the ASIR, the ASMR and ASDR of myocarditis were higher among men than among women. Over the past 3 decades, all ASRs among both sexes had downward trends, except for the ASMR among males, which showed a stable trend (EAPC=0.07) (Table 2). Markedly, the decrease in the ASDR from 1990 to 2019 was greater among females (EAPC=−1.54) than among males (EAPC=−0.91) at the global level (Table 3).
In 2019, the incident cases displayed a bimodal distribution with age and peaked the highest in the 65-69 age group. A relatively lower peak was shown in the 30-34 age group with incident cases (Figure 4). The number of deaths peaked in the 80-84 age group among males and in the 85-89 age group among females. The under 1 age group among both sexes had the largest number of deaths younger than 20 years (Figure 4). The number of DALYs in 2019 peaked in the under 1 age group among both sexes (Figure 4).

Figure 4. Age-standardized counts and rates of the myocarditis burden by sex/SDI region in 2019. (A) Incidence; (B) Mortality; (C) DALYs. DALYs: disability-adjusted life years; SDI: sociodemographic index.

The incidence and mortality rates per 100,000 person-years all showed approximately increasing trends with age among both sexes in 2019 (Figure 4). The DALY rate per 100,000 person-years showed a decreasing trend before the 10-14 age group and an increasing trend after the 10-14 age group with age among both sexes (Figure 4). Over the past 3 decades, the ASIRs among males and females were decreasing over all ages (EAPC<0) (Multimedia Appendix 3). From 1990 to 2019, the population older than 69 years among both sexes showed increasing trends in the ASMR and ASDR with age (EAPC>0) (Multimedia Appendix 3).

From 1990 to 2019, all SDI regions showed an increasing tendency in the absolute incident cases among the population over 30 years old (Multimedia Appendix 4). On the other hand, the ASIR showed a decreasing trend in high SDI regions among the population over 50 years old. Over the past 30 years, the ASIR in the other 4 SDI regions changed slightly across all age groups (Multimedia Appendix 4). Deaths were obviously rising in 5 SDI regions among the residents over 40 years old, especially in high-middle and middle SDI regions. Conversely, the number of deaths was decreasing in all SDI regions among the residents younger than 4 years, especially in high-middle, middle, and low-middle SDI regions (Multimedia Appendix 5). From 1990 to 2019, the ASMR in all SDI regions was steadily rising across all age groups over 85 years, except in low SDI regions where there was a decrease (Multimedia Appendix 5).
The number of DALYs was decreasing in all SDI regions among the population younger than 14 years, especially in high-middle, middle, and low-middle SDI regions (Multimedia Appendix 6). The ASDR showed a similar trend as the ASMR in 5 SDI regions across all age groups over 85 years. The ASDR was obviously decreasing in the under 1 age group in all SDI regions over the past 30 years (Multimedia Appendix 6).

The EAPCs in myocarditis ASIRs varied largely among age groups in the 5 SDI regions during the past 30 years. The highest EAPC in the ASIR in high SDI regions was noted between the ages of 14 and 19 years. In the other 4 SDI regions, the highest EAPC in the ASIR was noted in the age group of older than 95 years (Multimedia Appendix 4). Globally, the EAPCs in the ASMR and ASDR showed rising trends with age, and similar trends could be found in low-middle, middle, and high-middle SDI regions. The highest EAPCs in the ASMR and ASDR in high SDI regions were noted between the ages of 44 and 54 years. The EAPCs in the ASMR and ASDR in low SDI regions were always negative at all ages (Multimedia Appendices 5 and 6).

Potential Factors of Changing Trends

We found negative correlations between the EAPC and the ASIR ($\rho = -0.15; P = 0.03$), ASMR ($\rho = -0.23; P < 0.001$), and ASDR ($\rho = -0.28; P < 0.001$) in 1990 (Multimedia Appendices 7-9). On the other hand, there were no correlations between the EAPCs in the ASIR, ASMR, and ASDR and the SDI (Multimedia Appendices 7-9). The analysis of the annual ASIR of myocarditis from 1990 to 2019 and the SDI across all 21 GBD regions indicated that the ASIR in most GBD regions increased or decreased slightly, except for high-income North America that showed an obviously increasing trend first, a substantial decrease next, and an increase again (Multimedia Appendix 7). The ASMR and ASDR in Central Asia, Central Europe, and Southeast Asia changed significantly, while that in the other GBD regions changed slightly (Multimedia Appendices 8 and 9).

Discussion

Principal Findings

The study estimated the global burden of myocarditis from 1990 to 2019 systematically and comprehensively, which indicated the progressive and prominent influence on public health due to myocarditis. Compared with 1990, the numbers of incident cases and deaths of myocarditis both increased by above 1.6 times across the world, while the global DALYs decreased. All ASRs showed declining trends with negative EAPCs over the past 30 years. The myocarditis burden will certainly increase because of the increasing number of cases and deaths. Myocarditis is a kind of myocardial inflammatory disease diagnosed by established criteria [15,24,25]. The clinical manifestations of severe myocarditis are often fulminant, with sudden onset, extensive myocardial damage, and accompanied by fatal arrhythmia, contributing to a high risk of sudden death [26-30]. The burden of myocarditis is a public health problem of widespread concern, particularly under the background of SARS-CoV-2 infection and mRNA vaccines [31-33]. Since the pandemic, many epidemiological studies have concluded that SARS-CoV-2 increased the incidence of myocarditis. Studies also showed that the use of novel mRNA platforms led to a higher number of reported cases than with previous platforms. Hence, fundamental data on the burden of myocarditis is helpful for investigating the burden induced by SARS-CoV-2 [34,35].

Analyses based on the differences in sex, age, year, SDI quintiles, GBDR regions, and nations depicted the heterogeneities of myocarditis burden. Health policymakers should take notice of the significant implications on public health related to myocarditis globally.

This research found that higher SDI regions always showed a more significant ASIR from 1990 to 2019, which was not analyzed in a previous study of the GBD 2017 [8]. The high incidence of myocarditis in high and high-middle SDI regions may be related to better health care infrastructure and medical resources, population aging, the introduction of highly sensitive troponin and CMRI examinations, and extensive use of ICIs and vaccines [36]. Middle SDI regions accounted for the largest ASMR and ASDR in 2019, and low-middle SDI regions had the largest increase in the ASMR. The main reasons for high mortality and disability in middle SDI regions and the fastest increase in low-middle SDI regions may be delayed diagnosis, imperfect health care systems, and limited medical resources. Ventricular assist devices or extracorporeal membrane oxygenation (ECMO) need to be urgently applied in patients with cardiogenic shock or severe ventricular dysfunction due to myocarditis [37,38]. However, mechanical support devices have not been widely applied in middle and low-middle SDI regions. We also found that all ASRs in low SDI regions were the lowest, which may be due to insufficient health care infrastructure and resources. Definitely, the medical health care system and quality of data sources need to be improved in lower SDI regions for reducing the burden of myocarditis.

At the GBD level, high-income Asia Pacific, high-income North America, and East Asia accounted for the top 3 highest ASIRs in 2019. At the national level in 2019, the ASIR in Austria, Japan, and Sweden exceeded 20/100,000. The EAPCs in the ASMR and ASDR were the highest in Kazakhstan in the past 3 decades. Furthermore, there were significant negative correlations between the EAPCs in ASRs and the baseline ASRs in our results, which demonstrated that countries and territories with higher ASIRs, ASMRs, and ASDRs in 1990 went through more rapid decreases or slow increases in the ASRs of myocarditis from 1990 to 2019. Policymakers of countries and territories with higher ASRs were more likely to attach importance to the prevention of myocarditis and facilitate more reasonable policy formulation and resource allocation. In 2019, there were significant regional differences in the trends of ASRs, which provided accumulating evidence for the epidemiological transition of myocarditis across all GBD regions, nations, and territories. Policymakers should facilitate country-specific health research to allocate limited medical resources more reasonably based on the national information on the myocarditis disease burden.

The ASRs of myocarditis were higher among males than among females from 1990 to 2019 globally. Over the past 3 decades, all ASRs among both sexes showed downward trends, except for the ASMR among males, which showed a stable trend, and
females had a more significant decrease in the ASDR than males at the global level. Previous studies have already displayed a slightly higher incidence rate of myocarditis in men [8,29]. More recently, mRNA-based COVID-19 vaccines were reported to be responsible for an increased risk of myocarditis on the basis of passive surveillance reporting in the United States, especially among adolescent boys and young men after the second vaccine dose [39]. Men were more likely to experience a severe myocarditis developmental trajectory. Differences in innate immunity between men and women may contribute to a different prognosis. Testosterone and estrogen play important roles in the immune response. Testosterone can increase the affinity of the virus to myocytes and trigger a T helper type 1 immune response [40], inhibit the quantity of anti-inflammatory cells, and upregulate cardiac fibrotic remodeling genes [41]. Men with myocarditis have been found to have higher levels of heart failure biomarkers, creatine kinase, myoglobin, and T helper 17–associated cytokines [42]. On the other hand, women can develop a better regulatory immune response. Estrogen can promote the differentiation of T cells to T helper 2 cells and reduce the T helper type 1 immune response [43].

For estimating the age characteristics of the burden of myocarditis more accurately, we adopted a more refined age grouping scheme than that in the previous study of the GBD 2017 [8]. We found that the highest peak of deaths occurred in the under 1 age group among those aged younger than 20 years, and the number of DALYs in 2019 peaked in the under 1 age group among both sexes. There were 2 peaks in the incidence of myocarditis in childhood, and a high incidence was found in infancy and adolescence, with infants having poor prognoses [44]. Infantile myocarditis was more prone to causing severe ventricular dysfunction, which led to a higher probability of requiring mechanical circulatory support compared with that among patients of other ages. Myocarditis was proved to have caused 9% of cases of sudden infant death syndrome on autopsy [6]. It is noteworthy that infants are more susceptible to enteroviruses characterized by direct myocardial injury. The virus can trigger a cascade of hyperinflammatory responses and induce more robust immune responses among adolescents and young adults, especially among males. Previous research indicated that children and young adults had a higher incidence of myocarditis and mortality attributed to myocarditis compared with middle-aged and older adults [8,45]. However, our data showed that the incidence and mortality rates had increasing trends with age, and the DALY rate had a decreasing trend before the 10-14 age group and an increasing trend after the 10-14 age group with age in 2019. Differences in etiology, pathogenesis, and diagnosis strategies may be important reasons for the differences in the burden of myocarditis among children, adults, and elderly people. Senior citizens had high ASRs and numbers of cases in 2019, which demonstrated that elderly people gradually predominated in terms of the burden of myocarditis. The burden of myocarditis shifted to elderly people mainly due to population aging, ICI therapy, hypoinnununity, and more concomitant underlying diseases. Policymakers should be aware of the high DALYs in infants and the shifting of the burden to elderly people.

Limitations
This study has several limitations. First, the study may be prone to sampling bias. The GBD model of myocarditis was reconstructed based on a large number of varying quality, sparse, or limited scope data sources. This can affect the accuracy and applicability of the findings across all regions and populations. The possibility of missed cases of myocarditis, which are either undiagnosed or unreported, could also lead to an underestimation of the actual disease burden. Therefore, the estimation of the burden of myocarditis may deviate from the actuality to some extent, especially in some underdeveloped regions with scarce prior information. Second, longitudinal data were lacking in this study. The absence of data extending beyond the timeframe limits the understanding of recent trends and the ability to evaluate the effect of newer developments in the diagnosis, treatment, and prevention of myocarditis, especially after SARS-CoV-2, with some studies reporting its relation to an increased risk of myocarditis [34]. Third, the classification of myocarditis was not available in this study. There are many causes of myocarditis, among which viruses are the most common. Multiple types of viruses, including enteroviruses, adenoviruses, erythroparvoviruses, viruses from the Herpesviridae family, and SARS-CoV-2, can lead to myocarditis. Other etiologies include autoimmunity, vaccines, ICIs, and exposure to toxins and drugs. The clinical course of myocarditis caused by different factors may be different, and further investigations are necessary. Finally, the Dallas criteria in 1987 for myocarditis considered endomyocardial biopsy to be the standard diagnosis [46,47]. However, the diagnostic strategy has changed with the introduction of highly sensitive troponin and CMRI examinations in the past 20 years. In routine clinical practice, it is often sufficient to establish the diagnosis of myocarditis with the combination of symptoms and signs, laboratory examinations, and imaging studies [3]. It is unclear if the diagnostic criteria of myocarditis were kept the same among different areas. Further investigations are needed to evaluate the impact of changes and regional differences in the diagnostic workup for the assessment of myocarditis burden.

Conclusions
In general, myocarditis remains an important cause of early death and chronic disability, and it negatively impacted the global disease burden from 1990 to 2019. The numbers of incident cases and deaths associated with myocarditis have increased significantly. On the other hand, the ASRs of myocarditis showed decreasing trends from 1990 to 2019. High SDI regions showed more significant ASIRs, while middle SDI regions showed the highest ASMRs and ASDRs in 2019. Males consistently showed higher ASRs of myocarditis than females from 1990 to 2019 globally. Senior citizens had high incident cases and deaths among both sexes in 2019. Peak numbers of DALYs for both sexes were noted in the under 1 age group in 2019. At the national level, the EAPCs in ASRs had significant negative correlations with the baseline ASRs in 1990. Policymakers should develop targeted control strategies based on gender, region, age, and SDI; strengthen aging-related health research; and take notice of the changes in the epidemic characteristics of myocarditis. Targeted control strategies should be developed to reduce the high DALYs in infants and the
increasing burden of myocarditis in the elderly population based on the diversity of myocarditis burden explicated in our study.

Acknowledgments

We would like to thank the institute for the health metrics and evaluation staff, and its Global Burden of Disease Study collaborators who prepared the publicly available data. This work was supported by the National Natural Science Foundation of China (81970319); Taishan Scholars Program of Shandong Province (tsqn202103170); and Key Research and Development Program of Shandong Province (2019GSF108186). The funders were not involved in the collection, analysis, or interpretation of data, or the writing or submission of this report.

Data Availability

The Global Burden of Disease Study 2019 is a publicly available database [48], and we fully complied with data usage requirements.

Authors' Contributions

QYK, CFZ, and XRY contributed to the conception and design of the work. QYK, XX, ML, XM, CFZ, and XRY contributed to the acquisition and analysis of data for the work. QYK, CFZ, and XRY contributed to the visualization of the results. All authors contributed to the interpretation of the results. QYK and XRY drafted the manuscript. All authors critically revised the manuscript. All authors gave final approval and agreed to be accountable for all aspects of the work ensuring integrity and accuracy. CFZ and XRY are co-corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The incident cases, deaths, disability-adjusted life years, and corresponding burden rate of myocarditis in 204 countries and territories in 2019.

[DOCX File, 52 KB - publichealth_v10i1e46635_app1.docx]

Multimedia Appendix 2
Estimated annual percentage change in the burden rate of myocarditis in 204 countries and territories from 1990 to 2019.

[DOCX File, 38 KB - publichealth_v10i1e46635_app2.docx]

Multimedia Appendix 3
EAPC in the age-standardized rates of the myocarditis burden across all age groups by sex in 2019. (A) Incidence; (B) Mortality; (C) DALYs. DALYs: disability-adjusted life years; EAPC: estimated annual percentage change.

[PNG File, 565 KB - publichealth_v10i1e46635_app3.png]

Multimedia Appendix 4
Changes in the incidence of myocarditis across all age groups in global and SDI regions from 1990 to 2019. (A) Number of incident cases; (B) Age-standardized incidence rate; (C) EAPC in the age-standardized incidence rate. EAPC: estimated annual percentage change; SDI: sociodemographic index.

[PNG File, 642 KB - publichealth_v10i1e46635_app4.png]

Multimedia Appendix 5
Changes in the mortality of myocarditis across all age groups in global and SDI regions from 1990 to 2019. (A) Number of deaths; (B) Age-standardized mortality rate; (C) EAPC in the age-standardized mortality rate. EAPC: estimated annual percentage change; SDI: sociodemographic index.

[PNG File, 615 KB - publichealth_v10i1e46635_app5.png]

Multimedia Appendix 6
Changes in the DALYs of myocarditis across all age groups in global and SDI regions from 1990 to 2019. (A) Number of DALYs; (B) Age-standardized DALY rate; (C) EAPC in the age-standardized DALY rate. DALYs: disability-adjusted life years; EAPC: estimated annual percentage change; SDI: sociodemographic index.

[PNG File, 607 KB - publichealth_v10i1e46635_app6.png]
Multimedia Appendix 7
Factors affecting the EAPC in the ASIR of myocarditis and the changing trends of the ASIR across all GBD regions from 1990 to 2019. The size of the circle indicates the number of deaths associated with myocarditis. (A) Correlation between the EAPC in the ASIR and the corresponding ASIR in 1990; (B) Correlation between the EAPC in the ASIR and the SDI in 2019; (C) Changing trends of the ASIR across all GBD regions from 1990 to 2019. ASIR: age-standardized incidence rate; EAPC: estimated annual percentage change; GBD: Global Burden of Disease Study; SDI: sociodemographic index.

Multimedia Appendix 8
Factors affecting the EAPC in the ASMR of myocarditis and the changing trends of the ASMR across all GBD regions from 1990 to 2019. The size of the circle indicates the number of deaths associated with myocarditis. (A) Correlation between the EAPC in the ASMR and the corresponding ASMR in 1990; (B) Correlation between the EAPC in the ASMR and the SDI in 2019; (C) Changing trends of the ASMR across all GBD regions from 1990 to 2019. ASMR: age-standardized mortality rate; EAPC: estimated annual percentage change; GBD: Global Burden of Disease Study; SDI: sociodemographic index.

Multimedia Appendix 9
Factors affecting the EAPC in the ASDR of myocarditis and the changing trends of the ASDR across all GBD regions from 1990 to 2019. The size of the circle indicates the number of DALYs of myocarditis. (A) Correlation between the EAPC in the ASDR and the corresponding ASDR in 1990; (B) Correlation between the EAPC in the ASDR and the SDI in 2019; (C) Changing trends of the ASDR across all GBD regions from 1990 to 2019. ASDR: age-standardized DALY rate; DALYs: disability-adjusted life years; EAPC: estimated annual percentage change; GBD: Global Burden of Disease Study; SDI: sociodemographic index.

References


Abbreviations

ASDR: age-standardized disability-adjusted life year rate
ASIR: age-standardized incidence rate
ASMR: age-standardized mortality rate
ASR: age-standardized rate
CMRI: cardiac magnetic resonance imaging
DALYs: disability-adjusted life years
EAPC: estimated annual percentage change
GBD: Global Burden of Disease Study
ICD: International Classification of Diseases
ICI: immune checkpoint inhibitor
SDI: sociodemographic index
UI: uncertainty interval
Challenges and Implications for Menopausal Health and Help-Seeking Behaviors in Midlife Women From the United States and China in Light of the COVID-19 Pandemic: Web-Based Panel Surveys

Bobo Hi Po Lau¹, PhD; Catherine So Kum Tang¹, PhD; Eleanor Holroyd², PhD; William Chi Wai Wong³, MD

¹Department of Counselling and Psychology, Mrs Dorothy Koo and Dr Ti Hua Koo Centre for Interdisciplinary Evidence-Based Practice and Research, Hong Kong Shue Yan University, Hong Kong, China (Hong Kong)
²Faculty of Health and Environmental Sciences, Auckland University of Technology, New Zealand, New Zealand
³Department of Family Medicine and Primary Care, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
William Chi Wai Wong, MD
Department of Family Medicine and Primary Care, School of Clinical Medicine
Li Ka Shing Faculty of Medicine
The University of Hong Kong
3/F, Ap Lei Chau Clinic
161 Main Street, Ap Lei Chau
Hong Kong
China (Hong Kong)
Phone: 852 25185650
Email: wongwcw@hku.hk

Abstract

Background: The global population of women of menopausal age is quickly increasing. The COVID-19 pandemic has led to an accelerated increase in the use of telehealth services, especially technological solutions targeting women’s health. Understanding the factors behind midlife women’s help-seeking behaviors amidst the pandemic will assist in the development of person-centered holistic telehealth solutions targeting menopausal and postreproductive health.

Objective: This study aimed to compare the factors underlying help-seeking for menopausal distress among midlife women in the United States and China.

Methods: We conducted 2 web-based panel surveys in the United States using Amazon Mechanical Turk and in China using Credamo between July and October 2022. A total of 1002 American and 860 Chinese women aged between 40 and 65 years took part in the survey. The survey was designed based on the Health Belief Model with questions related to their menopausal knowledge, perceived severity of menopausal symptoms, perceived susceptibility to menopausal distress, perceived benefits of help-seeking, perceived COVID-19– and non–COVID-19–related barriers against help-seeking, self-efficacy, and motivation to seek help. Structural equations models were fitted for the data using full information maximum likelihood to manage missing data.

Results: Knowledge was not directly related to help-seeking motivation in both samples. Among the Chinese sample, knowledge was negatively related to perceived severity but positively related to COVID-19–related barriers; in turn, higher perceived severity, benefits, COVID-19–related barriers, and self-efficacy and lower non–COVID-19–related barriers were related to more motivation to seek help. In the US sample, knowledge was negatively related to perceived severity, susceptibility, benefits, barriers (COVID-19– and non–COVID-19–related), and self-efficacy; in turn, higher self-efficacy, COVID-19–related barriers, and benefits were associated with more help-seeking motivation. The factors explained 53% and 45.3% of the variance of help-seeking motivation among the American and Chinese participants, respectively.

Conclusions: This study revealed disparate pathways between knowledge, health beliefs, and the motivation for help-seeking among American and Chinese midlife women with respect to menopausal distress. Our findings show that knowledge may not directly influence help-seeking motivation. Instead, perceived benefits and self-efficacy consistently predicted help-seeking motivation. Interestingly, concern over COVID-19 infection was related to higher help-seeking motivation in both samples.
Hence, our findings recommend the further development of telehealth services to (1) develop content beyond health education and symptom management that serves to enhance the perceived benefits of addressing women’s multidimensional menopausal health needs, (2) facilitate patient–care provider communication with a focus on self-efficacy and a propensity to engage in help-seeking behaviors, and (3) target women who have greater midlife health concerns in the postpandemic era.

(JMIR Public Health Surveill 2024;10:e46538) doi: 10.2196/46538

KEYWORDS

menopause; help-seeking; health beliefs; telehealth; COVID-19; women; menopausal health; women’s health; online survey; health beliefs; awareness; digital health; symptom management; health education

Introduction

The population of women of menopausal age is increasing. According to the United Nations, women aged 50 years or older constitutes 26% of the global female population, compared to 21% 10 years ago [1]. Globally, it is predicted that the population of midlife women will reach 1.65 billion by 2050, from about 980 million in 2020, suggesting a burgeoning global demand for menopausal health care. The European Menopause and Andropause Society therefore proposed a Healthy Menopause [2] framework that links menopausal health to the World Health Organization’s healthy aging framework [3] to empower women to make salutary choices for their postreproductive health through personalized interdisciplinary care. Although menopause is often perceived as having ambiguous temporal boundaries and requiring broad population-based interventions, studies across cultures have revealed discrepancies in the understanding of the menopause experience between women in premenopause and postmenopause, pointing to insufficient knowledge and understanding together with inaccurate expectations [4-6].

Midlife women often manage their symptoms through self-initiated lifestyle modifications, such as relaxation, healthy diet, regular exercises, and supplements, or simply tolerating them; yet, they tended to be conservative with seeking medical help and associated services [7-11]. A study reviewing medical records in the United States found that while most menopausal women eventually saw doctors for their worrisome symptoms [12], 40% and 13% of them had neither medication nor any therapeutic interventions documented, respectively, and about 50% of women delayed seeking medical help for more than 6 months after symptom onset. In a recent study in Shanghai, China, only 36.8% of women who had menopausal symptoms went to see a doctor, and most commonly went to internal medicine, rather than gynecology or a specialized menopausal service [13]. Coupling with the patients’ lack of knowledge about menopause and associated care options, a substantial gap in professional education may lead health care professionals to either feeling uncomfortable or less-than-competent in discussing menopause, lacking a holistic perspective about the multidimensional distress of their patients [14].

The COVID-19 pandemic accelerated the rapid rise and accompanying advances in telehealth services [15], which was documented to have increased 38-fold in the total claims volume in the United States in late 2020 when compared to early 2020 [16], with psychiatry and endocrinology services ranked numbers 1 and 3, respectively. Similarly, in China alone, there has been a 900% increase in users and 800% increase in telehealth visits by new users from December 2019 to January 2020, shortly after the emergence of the COVID-19 pandemic [17]. Reflecting on the first 2 years of the COVID-19 pandemic, a global survey with 293 professionals in different fields revealed increased confidence in telehealth [18]. With respect to improving perimenopausal and postmenopausal women’s well-being, Menotech, as a branch of Femtech that encompasses primarily smartphone apps and telemedicine platforms [19-21] aimed at managing symptoms, hormonal replacement therapy, and triaging low-risk women for regular checkups [22], has flourished in recent years in North America, Europe, and Australia. The increased usage of telehealth presents both opportunities and barriers for midlife women seeking health education and care, underscoring the need to examine the related influencing factors from a cross-cultural perspective.

Both China and the United States witnessed a clear spike in telehealth use during COVID-19 outbreaks [23,24]. Both countries also saw marked increases in their population of women entering menopause in recent years. In China, the total number of women in menopause has already exceeded 200 million and is expected to reach 280 million by 2030 [25]. In the United States, about 1.3 million women are entering menopause annually, with over 51 million women going through or having gone through the transition at present [26,27].

Menopause is a universal phenomenon that ensues similar long-term health consequences across populations, yet with geographical differences in the nuanced experiences of its symptoms [13,28]. Palacios et al [29] found that Asian women reported a lower age of onset (42.1-49.5 years) than their North American counterparts (50.5-51.4 years), but a similar prevalence of overall symptoms (36%-50% in North America and 22%-63% in Asia). Chinese women also tend to experience more irritability and emotional distress instead [30]. In general, Asian women are likely to endorse more nonmedicalized views toward menopause, seeing it as a natural progression of life, yet perceive greater difficulties in disclosing and discussing their symptoms with their families, peers, and health professionals. Compared to their Western counterparts, they have been shown to enlist a repertoire of both Western and traditional Asian coping methods to manage with their physical and emotional symptoms [31]. The use of Chinese medicine to manage menopausal symptoms has been such an established practice among Chinese
women that evidence-based Chinese medicine clinical practice guidelines are already available [32]. The purpose of menopausal health care is 2-fold—relieving the multidimensional distress associated with menopause as well as nurturing holistic health for the postreproductive lifespan [33]. With the life expectancy of women in industrialized countries reaching over 80 years and large populations of women facing the onset of menopause at about 50 years of age, understanding the factors behind menopausal help-seeking is important for fostering midlife women’s health and well-being.

In this light, this study used web-based survey panels to explore the differences in help-seeking motivation and the associated factors among midlife women in the United States and China during the pandemic.

Methods

Design and Sampling

This study involved 2 web-based surveys with samples drawn from the United States and China. The US data were collected from Amazon Mechanical Turk (MTurk), while those from China were collected from Credamo Inc. The 2 platforms were chosen considering they are popular and well-established in conducting survey research in their respective countries at the time of data collection, and that they support a similar range of question types, participant screening procedures, and arrangements for distributing incentives. For MTurk, Robinson et al [34] reported that the panel included about 226,500 individuals based in the United States, while Credamo Inc declared over 3 million individuals on their panel based in China as of January 2023 [35]. The advantages of using web-based panels include efficient screening for participants with the desired demographic characteristics out of a vast pool, as well as ensuring that participants have experience filling in web-based surveys to safeguard data quality. In this study, which was conducted against the backdrop of the increase in telehealth use during the COVID-19 pandemic, we were particularly interested in the help-seeking motivation of midlife women who would be potential patrons of telehealth. For both samples, the inclusion criteria were being female and aged between 40 and 65 years, which was the normative window for menopausal transition ±5 years [36,37]. To safeguard data quality, we requested participants to have a HIT approval rate of ≥98% on Amazon MTurk or a historical approval rate and credibility score of ≥80% on Credamo Inc. These benchmarks were recommended by the respective platforms. On the 2 panels, the surveys were advertised and made available to eligible participants only. Informed consent was collected via the internet prior to the surveys. Participants were reimbursed for their participation at nominal prices according to the recommendations of the panels and handled by the panels.

Ethical Considerations

Ethical approval was obtained from the Human Research Ethics Committee of Hong Kong Shue Yan University (HREC 22-04(1)).

Instruments

Participants’ knowledge about menopause was evaluated on a 23-item scale. The first 22 items were adopted from Noroozi et al [38] and Gebretatros et al [39]. These items covered aspects including age and factors affecting the onset and symptom severity, health impacts, and symptoms. An item on hormonal replacement therapy was added as it frequently appears on surveys about menopausal attitudes and knowledge [5,40,41]. Participants scored 1 for a correct answer and the total score ranged from 0 to 23. Perceived severity of menopausal distress was measured by the Greene Climacteric Scale [42], which has 21 items covering physical, psychological, vasomotor, and sexual symptoms. Participants were instructed to answer on a scale from 0 (not at all) to 7 (extremely) on how bothered they were by these symptoms. The scale score ranged from 0 to 63 with good reliability (US: 0.93; China: 0.95). Items of perceived susceptibility, perceived benefits, perceived barriers, self-efficacy, and motivation to seek suitable help for menopausal needs were constructed with reference to the Champion’s Health Belief Model Scale [43] (Table 1 and Textbox 1). The items were face-validated by an expert group comprising 2 general practitioners, a gerontologist, experts in women’s health and mental health, and a medical social worker. Participants were asked to indicate their agreement on a Likert scale from 1 (strongly disagree) to 7 (strongly agree). We examined the perceived COVID-19–related barriers (ie, concerns about contracting COVID-19 from the clinic and spreading it to family members) and perceived non–COVID-19–related barriers (ie, financial, time, and informational barriers and embarrassment from help-seeking) separately. Understanding midlife and older women’s concerns about contracting COVID-19 from medical clinics, separating these 2 categories of barriers was intended to support finding new ways of providing menopausal services, such as telehealth adaptations during a pandemic like the COVID-19 pandemic.

### Table 1. Internal consistency of items assessing perceived susceptibility, perceived benefits, and perceived barriers for menopausal distress.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Items, n</th>
<th>Cronbach α</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>United States</td>
</tr>
<tr>
<td>Perceived susceptibility</td>
<td>2</td>
<td>.84</td>
</tr>
<tr>
<td>Perceived benefits</td>
<td>2</td>
<td>.81</td>
</tr>
<tr>
<td>Perceived COVID-19–related barriers</td>
<td>2</td>
<td>.90</td>
</tr>
<tr>
<td>Perceived non–COVID-19–related barriers</td>
<td>7</td>
<td>.94</td>
</tr>
</tbody>
</table>
**Textbox 1.** Items for perceived susceptibility, perceived benefits, perceived barriers, self-efficacy, and motivation to seek suitable help for menopausal distress.

<table>
<thead>
<tr>
<th>Perceived susceptibility</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• I am likely to suffer from the negative physical impacts of menopause.</td>
<td></td>
</tr>
<tr>
<td>• I am likely to suffer from the negative psychosocial impacts of menopause.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Perceived benefits</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Seeing a doctor/therapist will help me cope more effectively with my menopause.</td>
<td></td>
</tr>
<tr>
<td>• Seeing a doctor/therapist will reduce the negative impacts of menopause on me.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Perceived COVID-19–related barriers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• I have been afraid of catching COVID from attending medical care/a clinic.</td>
<td></td>
</tr>
<tr>
<td>• I have been worried about passing COVID to my loved ones from attending medical care/a clinic.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Perceived non–COVID-19–related barriers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• It has been hard for me to pay for the care for my menopause.</td>
<td></td>
</tr>
<tr>
<td>• I don’t know where to get help for my menopausal symptoms.</td>
<td></td>
</tr>
<tr>
<td>• Talking about menopause with a care professional is too embarrassing.</td>
<td></td>
</tr>
<tr>
<td>• I don’t have time for attending medical care or therapies for my menopausal symptoms.</td>
<td></td>
</tr>
<tr>
<td>• I have more important problems to attend to other than my menopause.</td>
<td></td>
</tr>
<tr>
<td>• I don’t have the encouragement I need from my loved ones to attend to my menopausal distress.</td>
<td></td>
</tr>
<tr>
<td>• I did not get care for my menopausal distress as I don't know what the doctor/therapist will do to help.</td>
<td></td>
</tr>
</tbody>
</table>

Self-efficacy

|  |
|----------------|---|
| • I am confident that I will get the help I need for my menopausal health needs. |  |

Motivation

|  |
|----------------|---|
| • I am motivated to get suitable help for my menopausal health needs. |  |

**Statistical Analysis**

Independent *t* tests were used to compare the factors affecting the motivation to seek help across the 2 samples. A structural equations model (SEM) was used to explore the associations among variables according to the Health Belief Model [44]. A model was fitted for each sample, since there was no prior study suggesting either identical or distinctive patterns of associations among the variables. The model has been fitted with the built-in SEM function of JASP (ver 0.16.4.0, The JASP Team) based on the *lavaan* R package, and the missing data were handled by full information maximum likelihood [45]. Instead of imputing or replacing the missing data, all the data in the sample were used to obtain the population parameters that maximize the likelihood functions. A satisfactory model fit was indicated by a nonsignificant result on the chi-square test, comparative fit index (CFI), and Tucker-Lewis Index (TLI) of over 0.95, as well as a root-mean-square error of approximation (RMSEA) and standardized root-mean-square residuals (SRMR) of 0.08 or less [45,46]. Since a large sample size tends to yield a significant chi-square test result regardless of the model fit, we still accepted the model if it fulfilled the last 4 criteria but with a significant chi-square test result. All tests used *P*<.05 to indicate statistical significance.

**Results**

The sample characteristics are presented on Table 2. The US sample had 1002 women (mean age 48.0, SD 6.3 years) collected in 6 days in late September 2022, while the Chinese sample had 860 women (mean age 48.6, SD 5.5 years) collected on 3 separate dates in July and October 2022. Most participants had tertiary education (US: n=909, 90.7%; China: n=567, 65.9%), were married (US: n=852, 85.4%; China: n=835, 97%), and had full-time occupations (US: n=805, 80.8%; China: n=662, 77%). The median income of the US participants was between US $50,000 and US $75,000 per annum (n=382, 38.1%) and the participants were predominantly White (n=835, 83.3%). The median income of the Chinese sample was between ¥15,000 and ¥20,000 (US $2104.82) and ¥20,000 (US $2806.24) per month (US $23,078-$30,769 per annum; n=217, 25.2%). As for their menopausal statuses, 53.2% (n=533), 35.8% (n=359), and 11% (n=110) of US participants were premenopausal, postmenopausal, and perimenopausal, respectively, while the respective ratios were 37.3% (n=321), 31.1% (n=267), and 31.6% (n=272) for Chinese participants. While the number of self-reported chronic illnesses were comparable across the 2 samples, US participants were more likely to use all categories of health care services except for Chinese medicine and
complementary and alternative therapies, for which Chinese participants reported greater use.

A measurement model was fitted for the 4 subscales with self-created items (perceived susceptibility, perceived benefits, perceived COVID-19–related barriers, and perceived non–COVID-19–related barriers) using confirmatory factor analysis. Satisfactory fit for 4 correlated latent factors was reported for both samples (US: $\chi^2_{59}=128.8$, $P<.001$; CFI=0.99; TLI=0.99; RMSEA=0.037; SRMR=0.017; China: $\chi^2_{59}=203.9$, $P<.001$; CFI=0.96; TLI=0.97; RMSEA=0.055; SRMR=0.041).

The US sample scored lower than the Chinese sample on menopausal knowledge, perceived severity, perceived benefits, self-efficacy, and motivation, but higher on perceived susceptibility and perceived barriers (Table 3). All $t$ statistics were statistically significant ($P<.001$).

Several demographic variables were associated with motivation to seek help. In the US sample, higher motivation was related to lower age ($r=-0.79$, $P=.02$), having a full-time occupation ($t_{914}=4.59$, $P<.001$), being affiliated to a religion ($t_{898}=6.87$, $P<.001$), and being married ($t_{913}=8.42$, $P=.03$). Help-seeking motivation was also related to menopausal state ($F_{2,916}=7.37$, $P=.001$), with significantly higher motivation to seek help among women in premenopause rather than those in postmenopause ($P<.001$), and with those in perimenopause in between the other 2 states. Motivation to seek help was not significantly related to education, income, ethnicity, or the number of chronic illnesses.

In the Chinese sample, higher motivation was related to higher income ($r=0.181$, $P<0.001$), having a full-time occupation ($t_{833}=3.758$, $P<.001$), and being married ($t_{833}=2.168$, $P=.03$). Motivation to seek help was also associated with menopausal state ($F_{2,832}=3.096$, $P=.046$), with the highest motivation among women in perimenopause, followed by those in postmenopause ($P=.04$), and then those in premenopause ($P=.03$). Motivation was not significantly related to age, education, the presence of a religious affiliation, or the number of chronic illnesses.

SEM was used to explore the pathways between knowledge, health beliefs, and the motivation to seek help. The overall fit was satisfactory for the US model ($\chi^2_{95}=320.48$, $P<.001$; CFI=0.96; TLI=0.95; RMSEA=0.053, $P=.24$; SRMR=0.037) despite the significant chi-square test result (Figure 1). The model accounted for 53% of the variance of motivation. Surprisingly, higher knowledge was related to lower perceived susceptibility, perceived benefits, perceived COVID-19– and non–COVID-19–related barriers, and even self-efficacy to seek help. Higher perceived benefits, perceived COVID-19–related barriers, and self-efficacy were related to higher motivation to seek help in American participants.

The overall fit was also satisfactory for the Chinese model ($\chi^2_{95}=320.90$, $P<.001$; CFI=0.96; TLI=0.95; RMSEA=0.053, $P=0.24$; SRMR=0.037) despite the significant chi-square test result (Figure 2). Inspection of the regression paths revealed that menopausal knowledge was associated with higher perceived COVID-19–related barriers and lower perceived severity. Higher perceived severity, perceived benefits, perceived COVID-19–related barriers, self-efficacy, and lower perceived non–COVID-19–related barriers were related to higher motivation to seek help. The model explained 45.3% of the variance of motivation. In both the United States and China models, knowledge was not significantly associated with motivation to seek help. Table 4 illustrates the results of the SEM with the United States and China data.
Table 2. Sample characteristics.

<table>
<thead>
<tr>
<th>Variables</th>
<th>United States (n=1002)</th>
<th>China (n=860)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>48.0 (6.3)</td>
<td>48.6 (5.5)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary or below</td>
<td>7 (0.7)</td>
<td>16 (1.9)</td>
</tr>
<tr>
<td>Secondary</td>
<td>86 (8.6)</td>
<td>277 (32.2)</td>
</tr>
<tr>
<td>Tertiary or above</td>
<td>909 (90.7)</td>
<td>567 (65.9)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>852 (85.4)</td>
<td>835 (97)</td>
</tr>
<tr>
<td>Single</td>
<td>81 (8.1)</td>
<td>10 (1.2)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>65 (6.5)</td>
<td>15 (1.8)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time occupation</td>
<td>805 (80.8)</td>
<td>662 (77)</td>
</tr>
<tr>
<td>Others</td>
<td>191 (19.2)</td>
<td>198 (23)</td>
</tr>
<tr>
<td><strong>Religious affiliation, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>906 (92.4)</td>
<td>101 (11.7)</td>
</tr>
<tr>
<td>No</td>
<td>75 (7.7)</td>
<td>759 (88.3)</td>
</tr>
<tr>
<td><strong>Income per annum (US $), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-24,999</td>
<td>61 (6.1)</td>
<td>N/A(^a)</td>
</tr>
<tr>
<td>25,000-49,999</td>
<td>162 (16.2)</td>
<td>N/A</td>
</tr>
<tr>
<td>50,000-74,999</td>
<td>382 (38.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>75,000-99,999</td>
<td>170 (17.0)</td>
<td>N/A</td>
</tr>
<tr>
<td>100,000-124,999</td>
<td>107 (10.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>125,000-149,999</td>
<td>57 (5.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>150,000-174,999</td>
<td>30 (3.0)</td>
<td>N/A</td>
</tr>
<tr>
<td>180,000-199,999</td>
<td>17 (1.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>≥200,000</td>
<td>10 (1.0)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Income per annum (US $), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-8,353</td>
<td>N/A</td>
<td>55 (6.4)</td>
</tr>
<tr>
<td>8,355-16,707</td>
<td>N/A</td>
<td>158 (18.4)</td>
</tr>
<tr>
<td>16,708-25,060</td>
<td>N/A</td>
<td>122 (14.2)</td>
</tr>
<tr>
<td>25,061-33,414</td>
<td>N/A</td>
<td>217 (25.2)</td>
</tr>
<tr>
<td>33,415-41,767</td>
<td>N/A</td>
<td>119 (13.8)</td>
</tr>
<tr>
<td>41,768-50,120</td>
<td>N/A</td>
<td>61 (7.1)</td>
</tr>
<tr>
<td>50,122-58,474</td>
<td>N/A</td>
<td>47 (5.5)</td>
</tr>
<tr>
<td>58,475-66,827</td>
<td>N/A</td>
<td>34 (4.0)</td>
</tr>
<tr>
<td>66,828-75,181</td>
<td>N/A</td>
<td>17 (2.0)</td>
</tr>
<tr>
<td>75,182-83,534</td>
<td>N/A</td>
<td>11 (1.3)</td>
</tr>
<tr>
<td>≥83,535</td>
<td>N/A</td>
<td>19 (2.2)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>835 (83.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Black</td>
<td>49 (4.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>Asian</td>
<td>97 (9.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>Variables</td>
<td>United States (n=1002)</td>
<td>China (n=860)</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Others</td>
<td>21 (2.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Menopausal state, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular menses in last 12 months (premenopausal)</td>
<td>533 (53.2)</td>
<td>321 (37.3)</td>
</tr>
<tr>
<td>Menses stopped in last 12 months (postmenopausal)</td>
<td>359 (35.8)</td>
<td>267 (31.1)</td>
</tr>
<tr>
<td>Less than 3 menses or irregular periods in the last 12 months (perimenopausal)</td>
<td>110 (11)</td>
<td>272 (31.6)</td>
</tr>
<tr>
<td>Number of chronic illness, mean (SD)</td>
<td>1.6 (1.3)</td>
<td>1.24 (1.9)</td>
</tr>
<tr>
<td>Health care used in last 3 months, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General practitioner</td>
<td>811 (80.9)</td>
<td>537 (62.4)</td>
</tr>
<tr>
<td>Gynecologist</td>
<td>677 (67.6)</td>
<td>504 (58.6)</td>
</tr>
<tr>
<td>Traditional Chinese medicine</td>
<td>422 (42.1)</td>
<td>470 (54.7)</td>
</tr>
<tr>
<td>Individual psychotherapy</td>
<td>566 (56.5)</td>
<td>109 (12.7)</td>
</tr>
<tr>
<td>Family counselling</td>
<td>503 (50.2)</td>
<td>70 (8.1)</td>
</tr>
<tr>
<td>Complementary and alternative therapy</td>
<td>657 (65.6)</td>
<td>585 (68)</td>
</tr>
<tr>
<td>Sex therapy</td>
<td>377 (39.4)</td>
<td>36 (4.2)</td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.
\(^b\)Race data were not collected for the Chinese sample.

Table 3. Descriptive statistics.

<table>
<thead>
<tr>
<th>Variables</th>
<th>United States</th>
<th>China</th>
<th>(t) test (df)(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value, mean (SD)</td>
<td>Participants, n</td>
<td>Value, mean (SD)</td>
</tr>
<tr>
<td>Knowledge</td>
<td>14.58 (3.30)</td>
<td>1002</td>
<td>16.86 (2.49)</td>
</tr>
<tr>
<td>Perceived severity</td>
<td>14.84 (9.93)</td>
<td>1002</td>
<td>21.96 (12.38)</td>
</tr>
<tr>
<td>Perceived susceptibility</td>
<td>5.24 (1.69)</td>
<td>1002</td>
<td>4.56 (1.58)</td>
</tr>
<tr>
<td>Perceived benefits</td>
<td>5.54 (1.41)</td>
<td>1002</td>
<td>5.74 (1.12)</td>
</tr>
<tr>
<td>Perceived COVID-19–related barriers</td>
<td>5.01 (1.97)</td>
<td>1002</td>
<td>4.15 (1.94)</td>
</tr>
<tr>
<td>Perceived non–COVID-19–related barriers</td>
<td>4.80 (1.72)</td>
<td>997</td>
<td>2.94 (1.14)</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>5.45 (1.40)</td>
<td>905</td>
<td>5.91 (1.09)</td>
</tr>
<tr>
<td>Motivation</td>
<td>5.58 (1.43)</td>
<td>919</td>
<td>6.03 (1.10)</td>
</tr>
</tbody>
</table>

\(^a\)All \(t\) statistics were statistically significant (\(P<.001\)).
Figure 1. Significant paths for the structural equations model with the US sample (N=1002). The numbers denote the unstandardized estimates and the standard errors are in brackets.

Figure 2. Significant paths for the structural equations model with the Chinese sample (N=860). The numbers denote the unstandardized estimates and the standard errors are in brackets.
Discussion

Principal Findings

Midlife women in the United States and China revealed rather different patterns of factors associated with menopausal help-seeking. For the US participants, greater knowledge conferred less perceived severity (indicated by bother about symptoms) but a lower self-efficacy to seek help. Furthermore, knowledge was related to lower perceived susceptibility, perceived benefits, and perceived barriers. In turn, motivation was only related to perceiving more benefits, COVID-19–related barriers, and self-efficacy, as was also the case among Chinese participants. For Chinese participants, higher knowledge was related to lower perceived severity and more concerns over contracting COVID-19 from attending clinics. In turn, greater perceived severity, perceiving more benefits, self-efficacy, and concerns for COVID-19, as well as lower non–COVID-19–related barriers, were related to greater motivation to seek help. Paradoxically for this sample, knowledge appeared to suppress the motivation to seek help by reducing perceived benefits, COVID-19–related barriers, and self-efficacy.

We put forward the following reasons that may explain why knowledge did not impose a direct positive effect on help-seeking motivation. First, a good proportion of menopausal women tend to rely on self-initiated lifestyle modifications rather than formal care for their menopausal symptoms [10,31]; thus, higher knowledge may have helped women tackle some menopausal symptoms on their own and effectively reduce help-seeking motivation. Second, Pimenta et al [47] found greater perceived control was related to fewer symptoms. Likewise, a systematic review by Ayers et al [48] revealed more reported symptoms in women with more negative attitudes toward menopause. If knowledge confers perceived control, it may reduce the experience of distress associated with menopause (which is supported by the negative relationships between knowledge and perceived severity in both of our samples), and eventually lead to lower help-seeking motivation. Third, greater menopausal knowledge may dilute menopause as something that should be tolerated, thereby reducing one’s propensity to seek professional help. In a qualitative study by Steffan [9], working midlife women in Britain exhibited strong self-reliance and personal responsibility either through putting up with or handling their menopausal symptoms and were silently “wanting yet not accepting help.” Knowledge may not only serve to normalize help-seeking behavior but may also dilute the intensity of the menopause experience. However, whether neoliberal individualization, or even trivialization, may have resulted in diminished service-seeking warrants more research into the lived experience of midlife women in different societies. In conclusion, whether knowledge imposes a direct effect on tackling the menopausal concerns or acts as an indirect effect via perceived control on palliating or diluting the distressing menopausal experience, which all in turn may reduce engaging in help-seeking, should be subject to further research. Moreover, the role of knowledge on perceived barriers against help-seeking may be socioculturally dependent. In our Chinese sample, knowledge was related to perceiving more barriers against help-seeking, while the reverse was observed among the American sample. In a culture that emphasizes the importance of “face” and regards the discussion of the female body and sexuality a taboo [49], knowing more about menopause may mean learning more about the stigmatizing details of the health condition, which may in turn hinder help-seeking. Also, in China, since gynecological and menopausal well-being support as well as psychotherapies are not as common as in the United States, knowing more about menopause may be concomitant with large investments of time, money, and effort. Instead, Chinese medicine regards menopause as a manifestation of weakened “kidney” function that may be

<table>
<thead>
<tr>
<th>Regression paths</th>
<th>United States (n=1002)</th>
<th></th>
<th>China (n=860)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unstandardized estimate (SE)</td>
<td>$P$ value</td>
<td>Unstandardized estimate (SE)</td>
<td>$P$ value</td>
</tr>
<tr>
<td>Knowledge $\rightarrow$ perceived severity</td>
<td>-0.69 (0.09)</td>
<td>&lt;.001</td>
<td>-0.50 (0.17)</td>
<td>.003</td>
</tr>
<tr>
<td>Knowledge $\rightarrow$ perceived susceptibility</td>
<td>-0.09 (0.02)</td>
<td>&lt;.001</td>
<td>0.01 (0.02)</td>
<td>.83</td>
</tr>
<tr>
<td>Knowledge $\rightarrow$ perceived benefits</td>
<td>-0.05 (0.01)</td>
<td>&lt;.001</td>
<td>0.00 (0.02)</td>
<td>.86</td>
</tr>
<tr>
<td>Knowledge $\rightarrow$ perceived COVID-19–related barriers</td>
<td>-0.17 (0.02)</td>
<td>&lt;.001</td>
<td>0.12 (0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Knowledge $\rightarrow$ perceived non–COVID-19–related barriers</td>
<td>-0.17 (0.02)</td>
<td>&lt;.001</td>
<td>0.02 (0.01)</td>
<td>.09</td>
</tr>
<tr>
<td>Knowledge $\rightarrow$ self-efficacy</td>
<td>-0.05 (0.01)</td>
<td>&lt;.001</td>
<td>-0.02 (0.02)</td>
<td>.13</td>
</tr>
<tr>
<td>Knowledge $\rightarrow$ motivation</td>
<td>0.00 (0.01)</td>
<td>&gt;.99</td>
<td>0.00 (0.01)</td>
<td>.93</td>
</tr>
<tr>
<td>Perceived severity $\rightarrow$ motivation</td>
<td>-0.01 (0.01)</td>
<td>.38</td>
<td>0.01 (0.00)</td>
<td>.03</td>
</tr>
<tr>
<td>Perceived susceptibility $\rightarrow$ motivation</td>
<td>0.03 (0.07)</td>
<td>.65</td>
<td>-0.03 (0.04)</td>
<td>.40</td>
</tr>
<tr>
<td>Perceived benefits $\rightarrow$ motivation</td>
<td>0.41 (0.06)</td>
<td>&lt;.001</td>
<td>0.13 (0.03)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Perceived COVID-related barriers $\rightarrow$ motivation</td>
<td>0.08 (0.04)</td>
<td>.04</td>
<td>0.07 (0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Perceived non–COVID-related barriers $\rightarrow$ motivation</td>
<td>0.02 (0.06)</td>
<td>.75</td>
<td>-0.35 (0.05)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Self-efficacy $\rightarrow$ motivation</td>
<td>0.40 (0.04)</td>
<td>&lt;.001</td>
<td>0.43 (0.03)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
enhanced by herbal medicines and acupuncture [50], which serves as an explanatory model of a socially cohesive and culturally accepted mode of intervention. The barriers to seeking formal, “Westernized” or medical help could be amplified when self-care means are relatively accessible. Greater knowledge, however, may infer lower embarrassment and fewer financial, information, and time barriers in a society, like the United States, in which health systems have developed to embrace positive aging, femininity, and sexuality coupled with greater access to established menopausal care.

In terms of demographic factors, only being employed full-time and being married were associated with higher motivation to seek help in both samples. Menopausal state was also related to motivation to seek help, yet the directions of association were different in the 2 samples. While a full-time occupation and being married may confer financial and family support for help-seeking in both societies, other variables did not demonstrate cross-cultural consistency. An in-depth exploration of the demographic impetus of help-seeking is beyond the scope of this paper considering the limitations of our web-based survey panels. Social determinants of health, such as discriminatory and disabling beliefs about menopause, the accessibility of health care systems, the availability of treatments, social support, occupational characteristics, and general literacy, could be strongly related to the menopausal experience and may affect formal help-seeking [11]. Differences in cultural representations of the menopausal experience may also affect practitioners’ understanding of their patients’ distress and needs, especially with women from ethnic minorities [51]. Future studies may adopt a person-in-context approach to thoroughly explore microlevel, mesolevel, and macrolevel factors that influence menopausal help-seeking in midlife women.

Limitations
In developing our survey, we needed to balance the need for a succinct measurement tool that engages the attention of our participants against the requirement for comprehensive coverage of the important constructs of the Health Belief Model. Therefore, most items were constructed specifically for this study. The widely used Champion’s Health Belief Model Scale has been adopted to examining the motivation for breast cancer self-examination, Papanicolaou tests, and other women’s cancer screening programs for women in menopause [45,52-54], but we found no adoption to the context of menopausal care to date. Also, as our survey relied on self-reported measures, the motivation to seek help can hardly be counter-checked with the actual help-seeking behavior.

Although the web-based survey panels, unlike face-to-face survey panels, have efficiently compressed the duration of data collection, which minimized the impacts of unexpected, time-sensitive factors (eg, sudden outbreaks), there are several limitations. First, we have no access to the number of potential participants our advertisements have reached, nor the exact randomization procedures of the panels with their eligible members. Thus, it is not possible to estimate the response rate. Also, although smart devices have become ubiquitous in the United States and China, web-based surveys have often resulted in the overrepresentation of individuals with higher socioeconomic status, those belonging to an ethnic majority, and those with a greater likelihood to have convenient and stable internet access. Our findings, however, could be particularly informative for telehealth development as we have engaged a group of women with good potential of early patronage [55]. Other factors, including distraction through concurrent apps on the smart device, self-selection bias, excessive nonnaive, and social desirability bias, may have also affected the data quality of the web-based survey panels [56].

Implications
Although our study was conducted during the height of the COVID-19 pandemic, our findings could be helpful for developing telehealth services targeting midlife women post the pandemic. Perceived COVID-19–related barriers were positively correlated with perceived severity in both of our samples. Higher levels of COVID-19–related concern may reflect either higher health consciousness [57,58] or an underlying health vulnerability that warrants an enhanced motivation to seek formal help. In this light, we call for expanding the use of telehealth to replace physical visits in menopausal care, especially for women who may be concerned about iatrogenic infection or during peaks of seasonal infectious diseases (eg, flu seasons). In fact, even prior to the COVID-19 pandemic, telehealth had been developed to reduce the direct and indirect costs of clinic attendance (eg, travel time and time off from work), as well as enhance privacy and autonomy by allowing consultation with practitioners at a location comfortable to the patient [59,60]. It is expected that these benefits will continue to prevail for a good proportion of patients even though some remote consultations may have reverted to face-to-face visits after the pandemic. Further research is needed to gather empirical evidence to indicate the proportion of remote consultations that have reverted to physical ones and which populations are likely to continue with remote consultations and telehealth services at-large post the pandemic.

Moreover, recent studies have reported that interventions including cognitive-behavioral-, behavioral-, or mindfulness-based therapies [61,62], walking programs supported by a virtual platform [63], and multidisciplinary clinics [64] may be useful for tackling menopausal symptoms as well as fostering the overall quality of life of women. As our study demonstrates, knowledge alone does not necessarily lead women to seek formal support. We suggest telehealth services focus on relieving psychosocial distress, connecting women to build a supportive and empathetic community, and enabling healthy lifestyles. ALSwayied et al [65] found early evidence that mobile-based physical activity programs are helpful for enhancing moderate to vigorous physical activities among midlife women. Another systematic review by Sediva et al [66] on 13 digital health interventions found that most interventions for midlife women targeted weight loss, lifestyles, and menopausal symptom management and primarily relied on the provision of instructions on healthy behaviors, fostering capability, and motivation. Although the heterogeneity of the studies and their outcomes forbids a conclusion about the general effectiveness of these digital health interventions, it is hoped that these contents may build confidence and familiarity in women with their health care providers, thus connecting women...
in need to more specialized support. Telehealth may be used to facilitate shared decision-making in the treatment of menopausal symptoms. The North American Menopause Society has developed MenoPro, an app that relies on a menopause decision-support algorithm, which can be accessed by both patients and health care providers to make decisions for pharmacologic or nonpharmacologic treatments [67]. The app may support a more transparent and personalized approach for risk stratification and shared decision-making that in turn enhances perceived benefits and self-efficacy of the support, which are 2 facilitators for help-seeking revealed in our findings.

A multistakeholder, participatory design study conducted by Backonja et al [21] found that women with menopause had distinctive concerns compared to their health care providers when it came to their desired Menotech functionality, with the former wanting Menotech to “understand, prevent, and positively reframe their menopause experience” and the latter placing greater emphasis on “tracking and patient-provider communication.” Future Menotech is expected to go beyond symptom tracking and assemble the collected data to predict the onset of bothersome experiences and prevent such experiences by offering personalized, quotidian suggestions, which, in turn, integrate and celebrate this life transition under the expanding occupational lifespan of midlife women. Our study points to the intricate cross-cultural differences in the role of knowledge between American and Chinese midlife women. In this light, we call for the development and testing of culturally targeted Menotech services that may enhance perceived benefits and self-efficacy or even capitalize on the health consciousness of midlife women to foster help-seeking. These culturally sensitive considerations may span from content (eg, inclusion of indigenous therapies, such as Chinese medicine and Qigong, among others) and service mode (eg, collaboration with established telehealth platforms of general practice vs a standalone system) to technologies (eg, apps or devices) and policies about data protection and privacy.

**Conclusion**

This study is the first to compare the factors affecting help-seeking tendencies for menopausal care among midlife women in the United States and China. Our findings reveal disparate pathways of help-seeking. Higher self-efficacy, perceived COVID-19–related barriers, and perceived benefits, but not knowledge, were related to higher help-seeking motivation in both samples. In the Chinese sample, motivation to seek help was associated with higher perceived severity and lower perceived non–COVID-19–related barriers. In the US sample, knowledge had negative associations with all factors. Implications for post–COVID-19 telehealth services for menopausal support have been discussed.

**Acknowledgments**

This project was financially supported by a matching fund for the Center for Interdisciplinary Evidence-based Practice and Research, Hong Kong Shue Yan University, Hong Kong. The authors would like to express their gratitude to the research assistance by Ms Febie Lin, Ms Lian Pat, Mr Eric Shum, and Ms Tiffany Sio. Generative AI was not used in producing any part of the manuscript.

**Conflicts of Interest**

None declared.

**References**


43. paramsothy P, Harlow S, Nan B. Duration of the menopausal transition is longer in women with young age at onset: the multietnic Study of Women's Health Across the Nation. Menopause 2017;24(2):142-149. [doi: 10.1097/gme.0000000000000736]


Abbreviations

CFI: comparative fit index
MTurk: Amazon Mechanical Turk
RMSEA: root-mean-square error of approximation
SEM: structural equations model
SRMR: standardized root-mean-square residual
TLI: Tucker-Lewis index
The Framing Effect of Digital Textual Messages on Uptake Rates of Medical Checkups: Field Study

Amnon Maltz¹, PhD; Stella Rashkovich², MPA; Adi Sarid³, PhD; Yafit Cohen⁴, BA; Tamar Landau⁵, MA; Elina Saifer⁴, MA; Neta Amorai Belkin⁴, MA; Tamar Alcalay², MPH

¹Department of Economics, University of Haifa, Haifa, Israel
²Nursing Division, Maccabi Healthcare Services, Tel Aviv, Israel
³Sarid Research Services, Haifa, Israel
⁴Marketing Automation Department, Maccabi Healthcare Services, Tel Aviv, Israel
⁵AI & Big Data Department, Maccabi Healthcare Services, Tel Aviv, Israel

Corresponding Author:
Amnon Maltz, PhD
Department of Economics
University of Haifa
Abba Khoushy Ave 199
Haifa, 3103301
Israel
Phone: 972 48240023
Email: amaltz@econ.haifa.ac.il

Abstract

Background: Health care authorities often use text messages to enhance compliance with medical recommendations. The effectiveness of different message framings has been studied extensively over the past 3 decades. Recently, health care providers have begun using digital media platforms to disseminate health-related messages.

Objective: This study aimed to examine the effectiveness of some of the most widely used message framings on the uptake rates of medical checkups.

Methods: This study used a large-scale digital outreach campaign conducted by Maccabi Healthcare Services (MHS) during 2020-2021, involving a total of 113,048 participants. MHS members aged 50-74 years were invited to take their recommended medical actions from the following list: human papillomavirus (HPV), mammography, abdominal aortic aneurysm, fecal occult blood test (FOBT), and pneumococcal vaccination. Each member was randomly assigned to receive 1 of 6 message framings: control (neutrally framed; n=20,959, 18.5%), gains (benefits of compliance; n=20,393, 18%), losses (negative consequences of noncompliance; n=15,165, 13.4%), recommendation (a recommendation by an authoritative figure, in this context by a physician; n=20,584, 18.2%), implementation intentions (linking potential outcomes to future reactions; n=20,701, 18.3%), and empowerment (emphasizing personal responsibility for maintaining good health; n=15,246, 13.5%). The time frames for measuring a successful intervention were 14 days for scheduling screenings (ie, HPV, mammography, or abdominal aortic aneurysm), 30 days for performing the FOBT, and 60 days for receiving pneumococcal vaccination. We also examined the effectiveness of media channels (text message or email) on uptake rates and whether the subject-line length is correlated with message-opening rates.

Results: No significant effect of message framing on uptake rates of medical checkups was observed. The rates of appointments for screening ranged from 12.9% to 14.1% across treatments. Based on a chi-square test, there was no evidence to reject the null hypothesis that these compliance rates are independent of the treatments (P=.35). The uptake rates for the FOBT and pneumococcal vaccination ranged from 23.3% to 23.8% across treatments, and we could not reject the hypothesis that they are independent of the treatments (P=.88). We also found that emails are more effective than text messages (P<.001) and that the subject-line length is negatively correlated with message-opening rates.

Conclusions: No evidence was found for an effect of the 5 message framings on uptake rates of medical checkups. To enhance compliance rates, public health officials may consider alternative framings. Furthermore, media channels and the subject-line length should be given careful consideration in the planning stages of health care campaigns.

Trial Registration: AEA RCT Registry AEARCTR-0006317; https://www.socialscienceregistry.org/trials/6317/history/201365

https://publichealth.jmir.org/2024/1/e45379
**Introduction**

Successful preventive care requires high uptake rates of early detection tests. Increasing these rates at relatively low costs is one of the more pressing challenges faced by public health officials. Over the past 3 decades, behavioral insights have been used to nudge individuals and increase their compliance rates with recommendations of their national health care authorities. One common tool used in this line of research is message framing, that is, the content of textual messages sent out by health care providers. Message framing has been used to encourage healthy behavior in various public health contexts, such as smoking cessation [1-3], early-detection cancer screenings [4,5], influenza vaccinations [6-8], and, recently, COVID-19 vaccinations [9-11].

In this large-scale field study, we reexamined some of the most influential and commonly used message framings to shed light on their effectiveness. Specifically, we investigated prospect theory, physician’s recommendations, implementation intentions, and empowerment.

Prospect theory [12-14] incorporates the idea that losses loom larger than gains. Over the past 3 decades, dozens of studies have examined the effect of these types of framings on different medical outcomes. Results have been rather mixed and have led to an important ongoing discussion regarding the usefulness of prospect theory–based frames in medical contexts (see Refs. [5,15,16] for systematic reviews).

Physician’s recommendations have also been examined in the literature, especially with respect to cancer screenings, and have been found to increase uptake rates [17-19].

Implementation intentions [20,21] link potential outcomes (eg, results of medical tests) and reactions (eg, steps to be taken after results are received). Such links in critical situations have been shown to facilitate individuals in reaching their health goals [22]. Implementation intentions have been used in different medical contexts, such as influenza vaccination [7,23], colorectal cancer screenings, and the fecal occult blood test (FOBT), for which some studies have reported positive effects [24,25], while others have found no effect at all [26].

Empowerment messaging is meant to emphasize the individual’s responsibility to take care of their own health. Thus far, it has shown potential to increase breast and cervical cancer screening rates [27-29].

Our study contributes to this literature by examining the large-scale effectiveness of these types of framings. To do so, we made use of a massive digital outreach campaign that was held during 2020-2021 by Maccabi Healthcare Services (MHS), the second-largest health maintenance organization (HMO) in Israel.

In addition to the examination of message framing, we also investigated 2 other topics: the effect of the digital media channel on message-opening rates and the influence of the subject-line length on responsiveness. As health care organizations turn to digital media as their main means of communication with the public, there is growing interest in the role of different media channels in generating positive health outcomes [30-36] and in the impact of the subject-line length on digital marketing campaigns [37-39]. These topics are essential for enhancing the public’s responsiveness to health-related digital communication, and they were addressed in this study.

Although the digital setup of this study allowed us to shed light on these 2 topics of emerging interest in public health, the main aim of our work was to test whether specific types of framings can increase adherence to medical recommendations. More specifically, we tested the null hypothesis: the well-known types of framings of digital messages used in this study have no effect on the uptake rates of medical checkups.

**Methods**

**Study Design**

MHS contacted members aged 50-74 years and invited them to take preventive medical actions according to their age, medical history, and guidelines of the Israeli Ministry of Health. The campaign ran from July 2020 to December 2021 and targeted the following medical procedures: mammography (for women), human papillomavirus (HPV; for women), abdominal aortic aneurysm screening (for men), and the FOBT and pneumococcal vaccination (for men and women).

MHS members included in this campaign were randomly assigned into 1 of 6 groups:

- Control: received an informative message, which was identical to the one used by MHS prior to this study
- Gains: highlighting the benefits of compliance
- Losses: highlighting the potential negative effects of noncompliance
- Recommendation: citing a recommendation by an MHS physician
- Implementation intentions: stating a plan that links outcomes to reactions
- Empowerment: emphasizing the personal responsibility of the members for their own health

We included all members who received a digital message from MHS for the first time. Those who received a digital message prior to the initiation of the study were excluded. (MHS started sending out digital messages about 1 year before the initiation of the study as part of a large digitization process within the organization. These messages used the neutral frame that later became the control group in our study.)

---

https://publichealth.jmir.org/2024/1/e45379

JMIR Public Health Surveill 2024 | vol. 10 | e45379 | p.719

(page number not for citation purposes)
Messages were communicated to members via email unless they lacked a valid email address in the MHS registry, in which case, a text message (SMS) was sent with a link redirecting to a landing page. The email subject line and contents (or the text message and the corresponding landing page) had 6 variations according to the assignment to control or treatment groups. Exact message contents are available in Multimedia Appendix 1.

Members received an invitation to perform 1 (or more) of 5 medical checkups that were recommended for them. These recommendations were based on the following criteria that combine the guidelines of the Israeli Ministry of Health with MHS’s designated target groups of the digital campaign:

- Mammmography: women above the age of 50 years for whom at least 2 years had passed since their last mammography test
- HPV: women aged 50-54 years for whom at least 3 years had passed since their last HPV test
- FOBT: women and men above the age of 50 years for whom at least 1 year had passed since their last FOBT and at least 5-10 years had passed since their last colonoscopy test (the exact number of years since the last colonoscopy that triggers an invitation to perform the FOBT determined by the individual’s specific health risks)
- Pneumococcal vaccination: women and men above the age of 65 years who never received the vaccination
- Abdominal aortic aneurysm screening: men above the age of 65 years with a history of smoking who had never performed the screening

Note that members could have been eligible for more than 1 checkup. In these cases, the invitation message included all the relevant checkups.

**Procedure**

Members who met 1 or more of the aforementioned criteria received a message from MHS in which they were invited to perform all the medical checkups that were recommended for them by the Israeli Ministry of Health. (Members who had already been diagnosed with a specific condition that 1 of the 5 medical procedures was meant to detect did not receive an invitation to test for that specific condition. They were contacted by MHS separately to follow up on their earlier diagnosis and to assess their overall medical well-being.) Following the initial contact, if the member scheduled an appointment or performed the recommended checkups, they did not receive any further messages. If they did not comply with the recommendations for all checkups that appeared on their initial message, they received reminders that followed the same framing theme that appeared in their initial message and included the remaining recommended checkups that the member did not perform or did not schedule an appointment for since the initial message was sent out. The first reminder was sent out 2 days after the first message but only in case the first message was sent via email and was not opened. The second and third reminders were sent out 2 and 4 weeks after the initial message, respectively (+/- a day or two if the reminder was supposed to be sent out on the weekend).

For mammography, HPV, and abdominal aortic aneurysm (which we call “screenings”), we collected data regarding scheduled appointments (we did not have access to the actual performance data of these screenings). For the FOBT and pneumococcal vaccination (which we call “lab tests”), we collected data regarding the actual performance (these procedures do not require appointments, and one can perform them simply by showing up at the clinic). Next to every medical procedure that appeared on a member’s invitation was a link that, once clicked, redirected the member to receive more information about the specific procedure on the website of MHS. Messages with screening recommendations included a link through which members could make the appointment (lab tests require no appointment). A total of 129,070 MHS members participated in this study, of which 113,048 (87.6%) were included in our analysis (the data-cleaning process is described in detail in Multimedia Appendix 2). Of these, 71,140 (62.9%) members received messages that only included lab tests, 30,878 (27.3%) received messages that only included screenings, and 11,030 (9.8%) received messages that included at least 1 lab test and at least 1 screening. Looking at the number of checkups that were included in the first invitation message of members who participated in the analysis, we found that 100,125 (88.6%) received a message that referred to only 1 recommended checkup, 10,416 (9.2%) received a message that included 2 recommended checkups, and 2507 (2.2%) received a message with 3 recommended checkups.

**Outcome Variables**

We analyzed the data separately for screenings and for lab tests since each of the 2 has a different outcome measure (appointment scheduling vs actual performance, respectively). In addition, since each message may have included more than 1 recommended medical action, we defined a “screening success” as a case in which a member scheduled an appointment to at least 1 of their recommended screenings. Similarly, a “lab test success” was a case in which a member performed at least 1 of their recommended lab tests. Members who received a message that included a recommendation for a screening and a lab test were included in both groups for analysis purposes. Such members could count as a success in screenings or in lab tests or in both, depending on their actions. Following the criteria set by MHS, the time frame considered for measuring a success was 14 days from the last contact date for screenings, 30 days for the FOBT, and 60 days for pneumococcal vaccination (setting different time frames had no effect on the qualitative results and conclusions).

The time frame that we considered for opening a message was 45 days from the initial contact date. (Since the final reminder was sent roughly 4 weeks after the first message, this time frame ensured that we essentially examined all members who made the minimal effort to respond to the messages by opening at least 1 of them.) A text message was considered opened if a member clicked the link that appeared in the text message that redirected the member to the landing page with the content of the invitation message (an email was considered opened if the member viewed the body of the email).
Covariates

In addition to the type of treatment (control, and 5 more treatment arms), we used the following covariates in our analysis:

- Gender: male or female.
- Age group: 50-54, 55-59, 60-64, 65-69, and 70-74 years.
- Socioeconomic status (SES): an index that is constructed by the Israel Central Bureau of Statistics (CBS), which incorporates information at the household level (education, income, employment, etc), alongside information regarding the geographical place of residence of the household and its local municipality. The index ranges from 1 to 10, where 1 reflects the weakest status and 10 the highest.
- Existing medical conditions (3 distinct covariates): indicating a previous history of (1) cardiac illness, (2) diabetes, or (3) a chronic condition of other type, such as being diagnosed with other chronic illnesses or receiving medication over a long duration of time.
- History of previous checkups: indicating whether the patient underwent any of the screenings or lab tests in the past.
- Media type: indicating whether the patient received an email message or an SMS text message.
- Periphery: indicating whether the patient resided in a geographically peripheral area (ie, based on the patient’s residence area relative to the central district of Israel).

Statistical Analysis

We examined the effectiveness of the different framings for members who opened 1 of their messages, including reminders. To test the significance and effect of the treatment groups on either screenings or lab tests, we used 2 logistic regression models, respectively. The dependent variable was a dummy variable that received 1 if that observation was a success and 0 otherwise. The independent variables of interest were the treatment dummies. Our full set of covariates was added to the logistic regressions as control variables. We also tested for independence of the compliance rates versus association with each of the groups using a chi-square test (ie, independence test). In addition, we reported 95% CIs (via the proportion test) for the compliance rates with screenings and lab tests for the control and treatment groups.

We explored the relationship between the SES and message-opening rates, using visualization and computation of correlation coefficients. Logistic regression was then run to examine whether the effect of sending a message by email on the message-opening rate was positive even when the SES was controlled for. The dependent variable was a dummy variable that received 1 if the member opened the message and 0 otherwise. The independent variables of interest were media channels (email received 1, and text message received 0) and the SES.

We compared the opening rates of messages (proportion of email messages read vs proportion of text message link clicks) using a 2-sided proportion test. We also compared the proportion of test compliance, given message opening (again, using a 2-sided proportion test). These comparisons were performed for various subgroups of the population (ie, controlling for demographic characteristics).

We used 2-sided proportion tests (95% CI) to compare the opening rates of short-subject-line messages versus long-subject-line messages.

Ethical Considerations

This research was approved by the Helsinki Committee of MHS (study number 0099-20-MHS) and by the Ethical Research and the Protection of Human Participants Committee, Faculty of Social Sciences, University of Haifa (approval number 369/21). Approvals included a waiver of consent. Identifying information was not shared with the researchers (ie, the researchers received anonymized data). No compensation was paid to participants.

Results

Sample Characteristics

Table 1 shows the demographic characteristics of our sample across treatments. (The table includes only members for whom we had all demographic variables and the SES, for a total of 112,802 members.) The treatment arms were well balanced on all observable characteristics, which included gender, age group, the SES, and existing medical conditions. Note that the age of the target group in the study and the broad definition of the “existing medical conditions” variable explain the high percentages of members who were classified as having a chronic condition.
Table 1. Demographics of the study sample by version.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Control (n=20,912, 18.5%), %</th>
<th>Gains (n=20,351, 18.0%), %</th>
<th>Losses (n=15,129, 13.4%), %</th>
<th>Recommendation (n=20,544, 18.2%), %</th>
<th>Implementation intentions (n=20,651, 18.3%), %</th>
<th>Empowerment (n=15,215, 13.5%), %</th>
<th>Total members (N=112,802), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>41.2</td>
<td>42.6</td>
<td>42.3</td>
<td>42.7</td>
<td>42.1</td>
<td>42.3</td>
<td>47,571 (42.2)</td>
</tr>
<tr>
<td>Female</td>
<td>58.8</td>
<td>57.4</td>
<td>57.7</td>
<td>57.3</td>
<td>57.9</td>
<td>57.7</td>
<td>65,231 (57.8)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-55</td>
<td>45.7</td>
<td>44.9</td>
<td>40.8</td>
<td>43.8</td>
<td>45.7</td>
<td>41.2</td>
<td>49,559 (43.9)</td>
</tr>
<tr>
<td>55-60</td>
<td>15.2</td>
<td>15.0</td>
<td>16.6</td>
<td>15.3</td>
<td>14.8</td>
<td>16.1</td>
<td>17,364 (15.4)</td>
</tr>
<tr>
<td>60-65</td>
<td>16.6</td>
<td>16.3</td>
<td>16.7</td>
<td>17.1</td>
<td>17.0</td>
<td>17.1</td>
<td>18,947 (16.8)</td>
</tr>
<tr>
<td>65-70</td>
<td>12.5</td>
<td>13.0</td>
<td>13.8</td>
<td>12.8</td>
<td>12.2</td>
<td>13.8</td>
<td>14,605 (12.9)</td>
</tr>
<tr>
<td>70-74</td>
<td>10.0</td>
<td>10.8</td>
<td>12.1</td>
<td>11.0</td>
<td>10.3</td>
<td>11.8</td>
<td>12,327 (10.9)</td>
</tr>
<tr>
<td>SES*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>313 (0.3)</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>1.7</td>
<td>2.3</td>
<td>1.7</td>
<td>1.7</td>
<td>2.0</td>
<td>2020 (1.8)</td>
</tr>
<tr>
<td>3</td>
<td>7.4</td>
<td>7.5</td>
<td>8.1</td>
<td>7.4</td>
<td>7.8</td>
<td>8.1</td>
<td>8657 (7.7)</td>
</tr>
<tr>
<td>4</td>
<td>9.9</td>
<td>10.2</td>
<td>10.5</td>
<td>9.7</td>
<td>9.8</td>
<td>11.2</td>
<td>11,465 (10.2)</td>
</tr>
<tr>
<td>5</td>
<td>15.5</td>
<td>15.9</td>
<td>16.9</td>
<td>16.3</td>
<td>15.9</td>
<td>16.7</td>
<td>18,189 (16.1)</td>
</tr>
<tr>
<td>6</td>
<td>19.6</td>
<td>18.7</td>
<td>19.4</td>
<td>19.3</td>
<td>18.9</td>
<td>18.8</td>
<td>21,551 (19.1)</td>
</tr>
<tr>
<td>7</td>
<td>17.0</td>
<td>17.5</td>
<td>16.1</td>
<td>16.9</td>
<td>17.1</td>
<td>16.0</td>
<td>18,987 (16.8)</td>
</tr>
<tr>
<td>8</td>
<td>14.6</td>
<td>14.8</td>
<td>13.5</td>
<td>14.9</td>
<td>14.9</td>
<td>13.7</td>
<td>16,329 (14.5)</td>
</tr>
<tr>
<td>9</td>
<td>11.0</td>
<td>10.5</td>
<td>9.9</td>
<td>10.5</td>
<td>10.6</td>
<td>10.4</td>
<td>11,841 (10.5)</td>
</tr>
<tr>
<td>10</td>
<td>3.2</td>
<td>3.0</td>
<td>3.0</td>
<td>3.1</td>
<td>3.1</td>
<td>2.9</td>
<td>3450 (3.1)</td>
</tr>
<tr>
<td>Medical condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>14.5</td>
<td>15.1</td>
<td>16.3</td>
<td>15.6</td>
<td>15.1</td>
<td>16.5</td>
<td>17,383 (15.4)</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>6.8</td>
<td>7.3</td>
<td>7.6</td>
<td>7.2</td>
<td>7.2</td>
<td>7.8</td>
<td>8201 (7.3)</td>
</tr>
<tr>
<td>Chronic illness</td>
<td>62.2</td>
<td>61.6</td>
<td>63.9</td>
<td>61.8</td>
<td>61.6</td>
<td>63.8</td>
<td>70,343 (62.4)</td>
</tr>
</tbody>
</table>

*SES: socioeconomic status. This is an index defined by the Central Bureau of Statistics (CBS), ranging from 1 to 10, where 1 reflects the weakest status and 10 the highest.

Main Findings
Approximately 60% of the 113,048 participants in the study, totaling 67,772 MHS members, opened 1 of their messages within the 45-day time frame.

The first 2 columns in Table 2 report success rates (conditional on opening 1 of the messages) by treatment arm. There were no significant differences in success rates across treatments for screenings and for lab tests. Chi-squared tests did not allow us to reject the null hypothesis that compliance rates are independent of treatments ($P=.37$ for screenings and $P=.88$ for lab tests). The last 2 columns of the table relate to the message’s subject-line length (ie, number of words) and the respective message’s opening rate. Table 3 includes logistic regressions with our full set of covariates added as controls (members who had missing values for any of the covariates were omitted from the regressions). Our null result is reflected in the table, as none of the framing versions had a significant effect on success rates. This null result held when we ran the same type of logistic regressions with interactions between the treatment variables and each of our controls (given the large number of interactions, we did not report the results of these regressions in the paper, but they are available upon request).
Table 2. Compliance rates by group (treatment or control).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Compliance rate, total n (%); 95% CI</th>
<th>Subject-line length (words)</th>
<th>Message-opening rate, total n (%); 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Screenings</td>
<td>Lab tests</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>5075 (14.1); 13.1%-15.1%</td>
<td>9581 (23.6); 22.7%-24.4%</td>
<td>20,959 (63.4); 62.7%-64.1%</td>
</tr>
<tr>
<td>Gains</td>
<td>4416 (13.9); 12.9%-15.0%</td>
<td>8687 (23.3); 22.4%-24.2%</td>
<td>20,393 (58.6); 57.9%-59.2%</td>
</tr>
<tr>
<td>Losses</td>
<td>3122 (12.9); 11.8%-14.1%</td>
<td>6027 (23.5); 22.4%-24.6%</td>
<td>15,165 (55.3); 54.5%-56.1%</td>
</tr>
<tr>
<td>Recommendation</td>
<td>4890 (13.1); 12.1%-14.1%</td>
<td>9357 (24); 23.2%-24.9%</td>
<td>20,584 (63.2); 62.6%-63.9%</td>
</tr>
<tr>
<td>Implementation intentions</td>
<td>4854 (14.1); 13.2%-15.2%</td>
<td>9015 (23.4); 22.6%-24.3%</td>
<td>20,701 (61.1); 60.5%-61.8%</td>
</tr>
<tr>
<td>Empowerment</td>
<td>3153 (14.1); 13.0%-15.4%</td>
<td>6133 (23.8); 22.7%-24.9%</td>
<td>15,246 (55.7); 54.9%-56.5%</td>
</tr>
</tbody>
</table>

Table 3. Two logistic regression models (for scheduling screenings and performing lab tests). For brevity, the following, statistically insignificant variables were omitted from the table (but were included in the model): gender, cardiovascular disease history, and diabetes history.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Dependent variable (success)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Screenings</td>
</tr>
<tr>
<td><strong>Independent variables, coefficient (SE)</strong></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>–2.325(^a) (0.099)</td>
</tr>
<tr>
<td>Gains</td>
<td>0.005 (0.063)</td>
</tr>
<tr>
<td>Losses</td>
<td>–0.078 (0.07)</td>
</tr>
<tr>
<td>Recommendation</td>
<td>–0.073 (0.062)</td>
</tr>
<tr>
<td>Implementation intentions</td>
<td>0.019 (0.061)</td>
</tr>
<tr>
<td>Empowerment</td>
<td>0.024 (0.069)</td>
</tr>
<tr>
<td>Conducted past checkup(s)</td>
<td>0.433(^b) (0.057)</td>
</tr>
<tr>
<td>Media channel (text message)</td>
<td>0.145(^b) (0.042)</td>
</tr>
<tr>
<td>Age 55-60 years</td>
<td>–0.128(^b) (0.058)</td>
</tr>
<tr>
<td>Age 60-65 years</td>
<td>–0.326(^a) (0.059)</td>
</tr>
<tr>
<td>Age 65-70 years</td>
<td>–0.324(^a) (0.065)</td>
</tr>
<tr>
<td>Age 70-74 years</td>
<td>–0.252(^a) (0.062)</td>
</tr>
<tr>
<td>SES(^c)</td>
<td>0.026(^b) (0.011)</td>
</tr>
<tr>
<td>Periphery</td>
<td>0.082 (0.072)</td>
</tr>
<tr>
<td>Chronic illness</td>
<td>0.105(^b) (0.048)</td>
</tr>
<tr>
<td>Observations, n</td>
<td>22,992</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>–9162</td>
</tr>
<tr>
<td>Akaike information criteria</td>
<td>18,361</td>
</tr>
</tbody>
</table>

\(^a\)P<.01.  
\(^b\)P<.05.  
\(^c\)SES: socioeconomic status. This is an index defined by the Central Bureau of Statistics (CBS), ranging from 1 to 10, where 1 reflects the weakest status and 10 the highest.

**Effectiveness by Media Channel**

We compared the effectiveness of the media channels through which messages were sent (69,777, 61.7%, members received emails, and 43,271, 38.3%, members received text messages). Table 4 shows that these groups differed in terms of their demographic characteristics: Members who received text messages were older on average, had a lower SES, and were more likely to be males. More details are provided in the Limitations section. Therefore, we compared the performance of the different media channels within subgroups of the population with similar demographic characteristics. Table 5 provides the opening rates of email messages, text messages, sample sizes, and the P value comparing the 2 rates.
for subgroups of the population with similar demographics. It is evident that text messages were disregarded more often than emails for every subgroup.

Table 6 provides for the same subgroups a comparison of the compliance rates of emails and text messages, given that a message was opened (rates, sample sizes, and $P$ values). In most subgroups examined, and overall, text messages exhibited a slightly higher (but significant) conditional-compliance rate.

Table 7 shows the overall effectiveness of media channels (ie, the overall unconditional success rates of emails vs text messages). Email messages outperform text messages in all subgroups and overall.

Table 4. Demographics of the sample by media channel allocation.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Email (n=69,777), %</th>
<th>SMS text message (n=43,271), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>40.8</td>
<td>44.4</td>
</tr>
<tr>
<td>Female</td>
<td>59.2</td>
<td>55.6</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-55</td>
<td>47.1</td>
<td>38.8</td>
</tr>
<tr>
<td>55-60</td>
<td>15.1</td>
<td>15.8</td>
</tr>
<tr>
<td>60-65</td>
<td>16.1</td>
<td>17.9</td>
</tr>
<tr>
<td>65-70</td>
<td>12.0</td>
<td>14.5</td>
</tr>
<tr>
<td>70-74</td>
<td>9.7</td>
<td>12.9</td>
</tr>
<tr>
<td>SES*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>0.7</td>
<td>3.6</td>
</tr>
<tr>
<td>3</td>
<td>4.9</td>
<td>12.2</td>
</tr>
<tr>
<td>4</td>
<td>7.2</td>
<td>15.0</td>
</tr>
<tr>
<td>5</td>
<td>13.5</td>
<td>20.3</td>
</tr>
<tr>
<td>6</td>
<td>18.9</td>
<td>19.4</td>
</tr>
<tr>
<td>7</td>
<td>19.1</td>
<td>13.1</td>
</tr>
<tr>
<td>8</td>
<td>18.0</td>
<td>8.8</td>
</tr>
<tr>
<td>9</td>
<td>13.6</td>
<td>5.4</td>
</tr>
<tr>
<td>10</td>
<td>4.1</td>
<td>1.5</td>
</tr>
</tbody>
</table>

*SES: socioeconomic status. This is an index defined by the Central Bureau of Statistics (CBS), ranging from 1 to 10, where 1 reflects the weakest status and 10 the highest.
Table 5. Opening rates, sample size, and P values by media channel for demographic subgroups.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Message opening, %</th>
<th>Sample size, n</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Email</td>
<td>SMS</td>
<td>Email</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-55</td>
<td>65.3</td>
<td>50.2</td>
<td>32,836</td>
</tr>
<tr>
<td>55-60</td>
<td>64.5</td>
<td>50.9</td>
<td>10,551</td>
</tr>
<tr>
<td>60-65</td>
<td>66.5</td>
<td>51.2</td>
<td>11,220</td>
</tr>
<tr>
<td>65-70</td>
<td>67.6</td>
<td>51.3</td>
<td>8356</td>
</tr>
<tr>
<td>70-74</td>
<td>67.7</td>
<td>48.3</td>
<td>6773</td>
</tr>
<tr>
<td>SES&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>59.5</td>
<td>39.6</td>
<td>3897</td>
</tr>
<tr>
<td>4-6</td>
<td>62.6</td>
<td>49.3</td>
<td>27,563</td>
</tr>
<tr>
<td>7-10</td>
<td>68.9</td>
<td>58.5</td>
<td>38,174</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>65.5</td>
<td>53.2</td>
<td>41,267</td>
</tr>
<tr>
<td>Male</td>
<td>66.4</td>
<td>46.9</td>
<td>28,469</td>
</tr>
<tr>
<td>Overall</td>
<td>65.9</td>
<td>50.4</td>
<td>69,777</td>
</tr>
</tbody>
</table>

<sup>a</sup>SES: socioeconomic status. This is an index defined by the Central Bureau of Statistics (CBS), ranging from 1 to 10, where 1 reflects the weakest status and 10 the highest.

Table 6. Effectiveness of the media channel conditional on message opening for demographic subgroups.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Compliance rates, given message opening, %</th>
<th>Sample size, n</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Email</td>
<td>SMS</td>
<td>Email</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-55</td>
<td>16.9</td>
<td>16.8</td>
<td>21,431</td>
</tr>
<tr>
<td>55-60</td>
<td>19.7</td>
<td>21.7</td>
<td>6803</td>
</tr>
<tr>
<td>60-65</td>
<td>18.1</td>
<td>21.5</td>
<td>7459</td>
</tr>
<tr>
<td>65-70</td>
<td>19.1</td>
<td>22.0</td>
<td>5646</td>
</tr>
<tr>
<td>70-74</td>
<td>21.0</td>
<td>24.5</td>
<td>4587</td>
</tr>
<tr>
<td>SES&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>18.7</td>
<td>20.5</td>
<td>2319</td>
</tr>
<tr>
<td>4-6</td>
<td>18.9</td>
<td>20.3</td>
<td>17,257</td>
</tr>
<tr>
<td>7-10</td>
<td>17.7</td>
<td>19.7</td>
<td>26,284</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>15.8</td>
<td>17.5</td>
<td>27,023</td>
</tr>
<tr>
<td>Male</td>
<td>21.6</td>
<td>23.9</td>
<td>18,903</td>
</tr>
<tr>
<td>Overall</td>
<td>18.2</td>
<td>20.2</td>
<td>45,961</td>
</tr>
</tbody>
</table>

<sup>a</sup>SES: socioeconomic status. This is an index defined by the Central Bureau of Statistics (CBS), ranging from 1 to 10, where 1 reflects the weakest status and 10 the highest.
Table 7. Overall effectiveness (unconditional) of media channels for demographic subgroups.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Success rate, %</th>
<th>Sample size, n</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Email</td>
<td>SMS</td>
<td>Email</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-55</td>
<td>11.0</td>
<td>8.4</td>
<td>32,836</td>
</tr>
<tr>
<td>55-60</td>
<td>12.7</td>
<td>11.1</td>
<td>10,551</td>
</tr>
<tr>
<td>60-65</td>
<td>12.0</td>
<td>11.0</td>
<td>11,220</td>
</tr>
<tr>
<td>65-70</td>
<td>12.9</td>
<td>11.3</td>
<td>8356</td>
</tr>
<tr>
<td>70-74</td>
<td>14.2</td>
<td>11.8</td>
<td>6773</td>
</tr>
<tr>
<td>SES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3</td>
<td>11.1</td>
<td>8.1</td>
<td>3897</td>
</tr>
<tr>
<td>4-6</td>
<td>11.8</td>
<td>10.0</td>
<td>27,563</td>
</tr>
<tr>
<td>7-10</td>
<td>12.2</td>
<td>11.5</td>
<td>38,174</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10.4</td>
<td>9.3</td>
<td>41,267</td>
</tr>
<tr>
<td>Male</td>
<td>14.3</td>
<td>11.2</td>
<td>28,469</td>
</tr>
<tr>
<td>Overall</td>
<td>12.0</td>
<td>10.2</td>
<td>69,777</td>
</tr>
</tbody>
</table>

SES: socioeconomic status. This is an index defined by the Central Bureau of Statistics (CBS), ranging from 1 to 10, where 1 reflects the weakest status and 10 the highest.

We now provide results concerning the robustness of the findings reported in Table 5, according to which emails were opened more frequently than text messages. First, Figure 1 shows a strong positive relationship between email availability and the SES (correlation coefficient=0.298, $P<.001$). Next, Figure 2 visualizes a positive relationship between the message-opening rate and the SES for both media channels (correlation coefficient=0.079 and 0.13 for emails and text messages, respectively; $P<.001$ in both cases). Moreover, Figure 2 shows that emails had higher message-opening rates compared to text messages for each SES.

In Table 8, we report the results of regressing, in a logistic model, the message-opening rates on the media channel being email and the SES. The email media channel positively affected the message-opening rates even when the SES (which had a positive effect of its own) was controlled for. Table 8 reports this result without controls and while controlling for the entire set of covariates.
Figure 1. Email availability as a function of socioeconomic status (SES).

Figure 2. Message-opening rates as a function of the socioeconomic status (SES) for email and text messages (SMS).
Table 8. Two logistic regression models for message-opening rates, with no controls and controlling for all available demographic characteristics and chronic conditions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Dependent variable (message opening)</th>
<th>Without controls</th>
<th>Full set of controls</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent variable, coefficient (SE)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>–0.596$^a$ (0.021)</td>
<td>–0.642$^a$ (0.028)</td>
<td></td>
</tr>
<tr>
<td>SES$^b$</td>
<td>0.112$^a$ (0.003)</td>
<td>0.112$^a$ (0.004)</td>
<td></td>
</tr>
<tr>
<td>Media channel (email)</td>
<td>0.516$^a$ (0.013)</td>
<td>0.513$^a$ (0.014)</td>
<td></td>
</tr>
<tr>
<td>Gender (male)</td>
<td>__$^c$</td>
<td>__$^c$</td>
<td>–0.034$^d$ (0.014)</td>
</tr>
<tr>
<td>Age 55-60 years</td>
<td>—</td>
<td>0.004 (0.02)</td>
<td></td>
</tr>
<tr>
<td>Age 60-65 years</td>
<td>—</td>
<td>0.059$^a$ (0.019)</td>
<td></td>
</tr>
<tr>
<td>Age 65-70 years</td>
<td>—</td>
<td>0.089$^b$ (0.021)</td>
<td></td>
</tr>
<tr>
<td>Age 70-74 years</td>
<td>—</td>
<td>0.024 (0.023)</td>
<td></td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>—</td>
<td>–0.019 (0.025)</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>—</td>
<td>–0.111$^a$ (0.018)</td>
<td></td>
</tr>
<tr>
<td>Chronic illness</td>
<td>—</td>
<td>0.037$^d$ (0.016)</td>
<td></td>
</tr>
<tr>
<td>Periphery</td>
<td>—</td>
<td>–0.018 (0.022)</td>
<td></td>
</tr>
<tr>
<td>Conducted past checkup(s)</td>
<td>—</td>
<td>0.058$^a$ (0.017)</td>
<td></td>
</tr>
<tr>
<td>Observations, n</td>
<td>112,802</td>
<td>101,675</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>–74,063</td>
<td>–66,604</td>
<td></td>
</tr>
<tr>
<td>Akaike information criteria</td>
<td>148,133</td>
<td>133,234</td>
<td></td>
</tr>
</tbody>
</table>

$^a P<.01$.

$^b$SES: socioeconomic status. This is an index defined by the Central Bureau of Statistics (CBS), ranging from 1 to 10, where 1 reflects the weakest status and 10 the highest.

$^c$These variables were not included in this regression model.

$^d P<.05$.

**Message-Opening Rates and Subject-Line Length**

Figure 3 shows the relationship between message-opening rates and the subject-line length for the different media channels and overall (see Table 2 for 95% CIs). As the length of the subject line increased, the message-opening rate decreased. For emails, the average opening rate of the 2 messages with the shortest subject lines (8 and 11 words) was 68.2% and the average opening rate of the 2 messages with the longest subject line (15 and 16 words) was 66.3%. For text messages, these rates were 54.1% and 48%, respectively (both differences were significant, with $P<.001$).
**Figure 3.** Message-opening rates as a function of the subject-line length, with a linear regression fit for email (dashed line), text messages (dotted line), and overall (solid line).

**Discussion**

**Principal Findings**

Our main findings do not allow rejecting the null hypothesis according to which uptake rates of medical checkups are independent of the message framing. Compliance rates (whether appointment scheduling or actual performance) are relatively similar across treatments, and this finding holds even after controlling for different demographic and health-related covariates.

We identified several variables that have a significant effect on compliance: conducting a medical procedure in the past increases the chances of compliance, as expected. Text messages seem to increase uptake rates, conditional on opening the message—a finding that is also robust when examining subgroups with similar demographic characteristics. It is interesting to note that age has a significant effect on success rates but the direction of the effect changes depending on the type of checkups (the reference age group in the regressions was 50-55 years).

The reason for these directional effects is likely the following: Older members may find it more difficult to schedule appointments through the online interface—hence the negative coefficients for the older age groups when we measured screening success. By contrast, lab tests only require a visit to the clinic and no appointments. Older members tend to visit their local clinic more regularly than younger members, which leads to the observed positive effect of age on lab test success. The variable for the SES also showed opposing effects, depending on the type of checkup. As one might expect, appointments made through the online interface increased with higher SES. By contrast, a higher SES leads to lower levels of performance of lab tests. Although we cannot delineate a clear explanation for this finding, it may be that, similar to older adults, individuals of a lower SES tend to visit their local clinics more often than those with a higher SES and therefore have higher performance rates of lab tests.

Regarding the effectiveness of media channels, we conducted this analysis for subgroups of the population with similar demographics, since there was no random assignment to media channels and the 2 groups apparently had different characteristics. For message-opening rates, we found highly significant differences favoring emails for every subgroup (Table 5).

Within each media method, we found that shorter subject lines increase the likelihood of opening messages. Given our main finding of no effect of framing on appointment scheduling/actual performance, we believe that this pattern suggests that shorter subject lines, rather than the subject line’s content, increase the likelihood of opening a message.
Conditional on message opening, we observed that text messages had a significant edge for 8 of 10 subgroups, as shown in Table 6. The overall effectiveness (ie, unconditional success rate) of each channel showed that emails outperform text messages and significantly so for all subgroups (Table 7). We concluded that the high message-opening rates for emails compared to text messages is the driving force behind this media channel’s increased success in our study.

The observation that text messages are disregarded more often than emails may be because text messages are considered more intrusive than emails. However, once opened, members are more likely to actually make an appointment or take the recommended medical action in the case of a text message than when reading through an email (Table 6). It may be that those who are willing to open a text message rather than disregard it have a more positive attitude toward health care recommendations than those willing to read through an email.

There was a positive relationship between the message-opening rate and the SES for both media channels. Moreover, emails had higher message-opening rates than text messages for each SES. These observations suggest that message-opening rates are positively affected by both the SES and the media channel being email.

These findings may be valuable for future public health campaigns as they provide information about the potential of different media channels to fulfill their designated purpose. However, the lack of random assignment into the different media channels prevents us from drawing firm conclusions on this matter. More research needs to be conducted (specifically with random assignment to media channels) in order to substantiate these findings.

**Relation to the Literature and Contribution**

Two of the framings used in our study, gains and losses, which are based on prospect theory [12-14], are the most commonly studied in the literature on framing of health-related communication. Dozens of studies have examined their effectiveness in nudging individuals toward desired healthy behaviors, including early-detection screening tests for different types of cancer. In the past 3 decades, evidence has amounted and has shown mixed results with respect to the empirical validity of these findings (see Refs. [5,15,40,41] for systematic reviews). As a result, conceptual concerns regarding their use in health-related behavior change have been raised [16]. Furthermore, it has recently been suggested that there is not much difference between the effectiveness of the 2 types of messages as far as persuasiveness is concerned [42], even when accounting for different moderating factors. Our study fits right into this discussion and allows testing this paradigm, and the other types of message framings that we have taken from the literature (implementation intentions, recommendation, and empowerment), in a large-scale field study. As mentioned before, we found no evidence that these types of framings are more effective in increasing compliance rates with medical checkups than a standard neutrally framed invitation message. These results suggest that new framings may need to be devised to attract individuals’ attention to important health recommendations.

Our 2 additional findings indicate that more research should be put into the exploration of the effect of the method of delivery of digital messages on uptake rates of medical checkups. The literature on the effect of the delivery medium on preventive health care outcomes has, thus far, been unable to clearly point toward a specific medium that is superior to the rest [43]. Findings to date suggest that all types of mediums are generally successful at encouraging behavior change, but there is no clear emerging pattern linking the type of medium and the type of message to adherence with specific health behaviors [33,34,36,44]. Moreover, a recent review of the potential effects of text messages in promoting health behavior change has called for large-scale implementations of text messages in order to better assess their pros and cons [45]. Our study contributes to this literature as we found that each communication medium in our study surpasses the other along a different dimension:

- **Message-opening rates:** The percentage of email recipients who opened their emails surpassed the percentage of SMS text message recipients who clicked the enclosed link in the text message.
- **Compliance conditional on message opening:** Among those who opened their invitation message (email or SMS text message), we observed higher compliance rates (ie, higher success rates) in SMS text message recipients compared to email recipients.

Given the unique strengths of each communication channel, our goal was to evaluate and compare their overall effectiveness (specifically, to measure compliance rates among recipients of each communication type). Our findings showed higher compliance rates among email recipients in comparison to the compliance rates of those receiving SMS text messages. Consequently, we concluded that the increased rates of email-opening rates elevate the overall compliance beyond what is observed for SMS text messages.

We think that examining these 2 media channels more carefully could be an exciting topic for future research. It also entails clear policy implications regarding tailoring the preferred media channel and combining the channels in specific campaigns. Messages that call for action and are expected (in advance) to have high opening rates due to their high relevance (eg, COVID 19–related messages during the recent pandemic) may benefit from text messages. However, messages that refer to routine checkups and ongoing recommendations, as those that appeared in our study, may benefit from being sent by email, as they may be opened more frequently and consequently will receive more attention than if they are delivered by SMS.

Regarding the subject-line length, it has received increasing attention in recent marketing literature [37-39], but we are unaware of any studies that examine this topic in relation to health behavior. Our suggestive finding that the subject-line length is negatively correlated with message-opening rates may trigger future research on the topic. Designing a field experiment that examines the effect of additional words/characters in messages’ subject lines on message-opening rates and the overall compliance of health-related messages may have important implications for the success of future health behavior campaigns in reaching their target audience.
Limitations

One limitation of our study is that there was no random assignment to media channels. Members with no valid email address in the registry of MHS received a text message. To deal with this limitation, we conducted the relevant analysis for subgroups with similar demographic characteristics. However, it is possible that there are also unobserved characteristics that differ across the 2 groups. This prevented us from drawing firm conclusions regarding the effectiveness of the different media channels. More research needs to be conducted (specifically with random assignment to media channels) in order to substantiate these findings. Another potential limitation, as in many text message interventions, is the lack of control over whether members who opened their message actually read it. It is possible that some skim through their messages without paying much attention to the content (see Ref. [46] for a recent pilot study regarding patients’ reading rates of emails sent by their physicians). The random assignment to different groups should attenuate the potential effects of this limitation. To further address this issue and strengthen the effect of the framing, we included the main theme of each framing, not only in the body of the message, but also in the subject line. However, one should acknowledge the fact that in the era of digital messaging, limited attention of recipients is likely and should be considered. In this respect, our field study realistically mirrors the potential difficulties that future digital campaigns are likely to encounter. Specifically, it may be more difficult to generate a specific state of mind by framing in the digital era compared to the traditional communication methods (eg, brochures, fliers, post).

Policy Implications

Our findings have important policy implications since message framings are frequently used by health care providers to increase individuals’ compliance rates with health recommendations. Taking a broader perspective, it raises questions regarding the ability of specific textual nudges to increase the public’s adherence and compliance rates with different types of government regulations and professional recommendations. It has been recently argued that accounting for publication bias, these nudges may have a small or no effect in different contexts [47,48]. Our findings add to this ongoing discussion and may suggest that there is room for novel framings or other types of behavioral nudges in order to increase compliance rates with preventive care recommendations. It also sheds light on the seemingly large potential effect of the choice of media through which messages are delivered and the subject-line length. These choices may seem of low consequence at first glance, but our suggestive findings imply otherwise. Future research that carefully examines these topics on a large scale may prove useful for the success of future digital health care campaigns.

Conclusion

This study provides evidence regarding the usefulness of well-known framings from health communication research in a large-scale digital field experiment held in Israel. Our main finding is that compliance with recommendations is not affected by the types of framings we used. We believe that this evidence, coming from a wide outreach campaign, shows that digital messaging may require new framings in order to enhance compliance in the digital era. We also report suggestive evidence on significant differences between the effectiveness of the media channels used in our study (email and SMS) and on the effect of subject-line length on message-opening rates.

Data Availability

The code will be made available on a public GitHub repository. The pre-registration is available through the American Economic Association’s (AEA) randomized controlled trial (RCT) registry [49]. The individual-level data sets used in this study are not publicly available, as we do not have legal permission to publish them due to Maccabi Healthcare Services (MHS) policy. MHS considers these to be sensitive health data that contain individual-level information regarding health. These type of data are not made available even if they are deidentified (as it may still be possible to reidentify patients from such data). To keep transparency as high as possible, given this limitation, we have posted our code for the analysis to a public GitHub repository. In addition, some of the tables in the paper provide aggregated data for different variables (eg, Tables 1, 4-7). In special cases of interest, a request may be sent to the corresponding author, who will submit a special inquiry to MHS for partial or full access to the data. Access may still be denied as the decision will be made by MHS in compliance with its data sharing policy.

Authors’ Contributions

AM was responsible for conceptualization, formal analysis, methodology, supervision, validation, visualization, writing—original draft, and revising the manuscript. AS was responsible for formal analysis, methodology, software, visualization, and writing and revising the manuscript. SR was responsible for project administration, resources, supervision, and validation. YC was responsible for data curating. TL was responsible for data curating, and writing and revising the manuscript. ES was responsible for software and project administration. NAB was responsible for resources, project administration, and supervision. TA was responsible for supervision. AM, AS, YC, and TL directly accessed and verified the data.

Conflicts of Interest

None declared.
References


34. Gram IT, Larbi D, Wangberg SC. Comparing the efficacy of an identical, tailored smoking cessation intervention delivered by mobile text messaging versus email: randomized controlled trial. JMIIR Public Health Surveillance 2019 Sep 27;7(9):e12137 [FREE Full text] [doi: 10.2196/12137] [Medline: 31573935]


42. O’Keefe DJ, Hoeken H. Message design choices don’t make much difference to persuasiveness and can’t be counted on—not even when moderating conditions are specified. Front Psychol 2021 Jun 29;12:664160 [FREE Full text] [doi: 10.3389/fpsyg.2021.664160] [Medline: 34267703]


49. Increasing cancer screening rates and vaccine uptake using digital messages based on insights from behavioral economics. AEA RCT Registry. URL: https://doi.org/10.1257/rct.6317 [accessed 2024-01-17]

Abbreviations
- CBS: Central Bureau of Statistics
- FOBT: fecal occult blood test
- HPV: human papillomavirus
- MHS: Maccabi Healthcare Services
- SES: socioeconomic status

©Amnon Maltz, Stella Rashkovich, Adi Sarid, Yafit Cohen, Tamar Landau, Elina Saifer, Neta Amorai Belkin, Tamar Alcalay. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 06.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Association Between Physical Activity and the Risk of Burnout in Health Care Workers: Systematic Review

Pierpaolo Mincarone¹,²*, PhD; Antonella Bodini³*, PhD; Maria Rosaria Tumolo⁴, MSc; Saverio Sabina²,⁵, MSc; Riccardo Colella⁵,⁶, PhD; Linda Mannini⁵, PhD; Eugenio Sabato¹,⁷, MD; Carlo Giacomo Leo²,⁵, PhD

¹Research Unit of Brindisi, Institute for Research on Population and Social Policies, National Research Council, Brindisi, Italy
²MOVE-Mentis s.r.l, Cesena, Italy
³Institute for Applied Mathematics and Information Technologies “Enrico Magenes”, National Research Council, Milan, Italy
⁴Biological and Environmental Sciences and Technology Department, University of Salento, Lecce, Italy
⁵Institute of Clinical Physiology, National Research Council, Lecce, Italy
⁶Innovation Engineering Department, University of Salento, Lecce, Italy
⁷Respiratory Diseases Unit, “Antonio Perrino” P.O., Local Health Unit “ASL Brindisi”, Brindisi, Italy

*these authors contributed equally

Corresponding Author:
Maria Rosaria Tumolo, MSc
Biological and Environmental Sciences and Technology Department, University of Salento
Strada per Monteroni
Lecce, 73100
Italy
Phone: 1 3890118275
Email: mariarosaria.tumolo@unisalento.it

Abstract

Background: Burnout is a multidimensional psychological syndrome that arises from chronic workplace stress. Health care workers (HCWs), who operate in physically and emotionally exhausting work contexts, constitute a vulnerable group. This, coupled with its subsequent impact on patients and public economic resources, makes burnout a significant public health concern. Various self-care practices have been suggested to have a positive effect on burnout among HCWs. Of these, physical activity stands out for its ability to combine psychological, physiological, and biochemical mechanisms. In fact, it promotes psychological detachment from work and increases self-efficacy by inhibiting neurotransmitters and neuromodulators, increasing endorphin levels, enhancing mitochondrial function, and attenuating the hypothalamic pituitary-adrenal axis response to stress.

Objective: Our objective was to conduct a systematic review of the evidence on the association between physical activity and burnout among HCWs.

Methods: We considered HCWs, physical activity, and burnout, framing them as population, exposure, and outcome, respectively. We searched APA PsycArticles, MEDLINE, and Scopus until July 2022. We extracted relevant data on study design, methods to measure exposure and outcome, and statistical approaches.

Results: Our analysis encompassed 21 independent studies. Although 10% (2/21) of the studies explicitly focused on physical activity, the remaining investigations were exploratory in nature and examined various predictors, including physical activity. The most commonly used questionnaire was the Maslach Burnout Inventory. Owing to the heterogeneity in definitions and cutoffs used, the reported prevalence of burnout varied widely, ranging from 7% to 83%. Heterogeneity was also observed in the measurement tools used to assess physical activity, with objective measures rarely used. In total, 14% (3/21) of the studies used structured questionnaires to assess different types of exercise, whereas most studies (18/21, 86%) only recorded the attainment of a benchmark or reported the frequency, intensity, or duration of exercise. The reported prevalence of physically active HCWs ranged from 44% to 87%. The analyses, through a variety of inferential approaches, indicated that physical activity is often associated with a reduced risk of burnout, particularly in the domains of emotional exhaustion and depersonalization. Furthermore, we compiled and classified a list of factors associated with burnout.

Conclusions: Our comprehensive overview of studies investigating the association between physical activity and burnout in HCWs revealed significant heterogeneity in definitions, measurements, and analyses adopted in the literature. To address this issue, it is crucial to adopt a clear definition of physical activity and make thoughtful choices regarding measurement tools and...
methodologies for data analysis. Our considerations regarding the measurement of burnout and the comprehensive list of associated factors have the potential to improve future studies aimed at informing decision-makers, thus laying the foundation for more effective management measures to address burnout.

*KEYWORDS*
burnout; Maslach Burnout Inventory; MBI; Copenhagen Burnout Inventory; CBI; Professional Fulfillment Index; PFI; physical activity; health care workers; public health policy

### Introduction

#### Background

Burnout is a multidimensional psychological syndrome resulting from chronic workplace stress. It is characterized by feelings of energy depletion or exhaustion, increased mental distance from one’s own job or cynicism, and reduced professional efficacy [1].

Burnout is gaining attention as a major public health concern for the mental health challenges of health care workers (HCWs) [2]. It poses a threat to the quality of care delivery, especially in terms of patient safety [3], and results in high resource consumption to face its consequences [4,5]. This became even more evident during the COVID-19 pandemic [6], which tightened physicians’ work conditions [7,8], which highlighted the frailty of health care systems in emergency management [11]. Although COVID-19 no longer constitutes a public health emergency [12], HCWs continue to be particularly vulnerable to burnout [13,14], with weakened abilities to manage the typical difficulties of care work and increased exposure to emotionally challenging situations [15]. This is especially true for those directly engaged in patient care given the demanding nature of their roles [16]. They often work long hours, experience sleep deprivation, contend with irregular schedules, and are exposed to emotionally challenging situations. In addition, they face the pressure to master a vast body of clinical knowledge [17]. Furthermore, HCWs are at high risk of workplace injuries [18] and assaults [19,20]. These challenges are compounded by common life stressors such as work-home conflicts, educational debts, relationship status, the age of their children, and the employment status of their partners [21]. As reported in a literature analysis [6], burnout can result in various negative outcomes and criticalities for HCWs, including anxiety, depressive disorders, alcohol abuse or dependence, and suicidal ideation. Burned-out workers experience poor mental health even without a clinically diagnosable disorder, as also proved by the fact that, within this group, there are more mental health problems than in most other occupational groups [22].

Within the broad spectrum of possible public health options to address burnout in HCWs, prioritizing measures to promote mental well-being has become paramount [23,24]. This entails addressing cultural factors, particularly those related to stigma; ensuring protected access to mental health care services; and implementing active policies to encourage healthy lifestyles [25,26]. Among these, physical activity is recognized worldwide as a key strategy for promoting mental well-being [27,28] in addition to helping prevent and manage noncommunicable diseases. However, unlike physical health [29], determining the “optimal dose” of physical activity—considering the combination of intensity, duration, and frequency [30]—for mental well-being remains uncertain and dependent on the specific domain of investigation [27].

A preliminary search of previous systematic reviews on the effects of physical activity on burnout was conducted in Scopus, PubMed, and PROSPERO and resulted in 2 different studies. Naczenski et al [31] found that physical activity was an effective strategy to reduce burnout among workers from various sectors. Regarding health professionals specifically, Bischoff et al [32] identified a potential beneficial effect in health professionals of mind-body practices such as yoga or qigong on occupational stress, one of the conditions for burnout [1].

#### Objectives

Our study systematically investigated how physical activity was incorporated into studies on burnout among HCWs. Recognizing that strategies created without evidence lead to ineffective programs, wasted resources, and persistently poor health outcomes [33], our ambition was to contribute to the development of evidence-based public health policies.

First, we aimed to provide insights into the reported level of participation in physical activity and the extent of participants at high risk of burnout. We also intended to verify whether a correlation emerges between levels of physical activity and burnout and whether a dose-response association exists.

Our second purpose was to thoroughly assess the quality of the available evidence in terms of collecting, compiling, managing, analyzing, and using health data. This assessment could also offer indications to generate hypotheses for further research to strengthen the body of evidence.

These are essential steps along the road that leads to shaping public health strategies and resource allocation for HCWs’ well-being.

### Methods

As recommended for systematic reviews of association studies [34,35], we adopted the population, exposure, and outcome (PEO) approach by considering physical activity habits as the exposure factor and burnout as the outcome. To ensure accuracy and transparency, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [36] and the PRISMA-S (PRISMA literature search extension) [37] guidelines.

---

https://publichealth.jmir.org/2024/1/e49772

JMIR Public Health Surveill 2024 | vol. 10 | e49772 | p.736

(page number not for citation purposes)
Eligibility Criteria

The eligibility criteria are detailed in Multimedia Appendix 1. We included studies that examined the association between physical activity and burnout in health personnel directly involved in the provision of care services, such as physicians, nurses, and technicians, with no restrictions on demographics or workplace context. We considered both qualitative and quantitative measures of physical activity as eligible for inclusion. As we were interested in the association between physical activity and burnout, interventional studies were included regardless of the presence of a control group. A validated assessment tool for burnout (a general work-related stress outcome was not of interest) was mandatory for inclusion.

Our search was not limited by geographic context, funding source, or time horizon. We excluded studies in which physical activity was not distinct from other supportive strategies aimed at managing stress or building resilience or from other mindfulness practices. In addition, studies that focused on yoga or qigong were excluded as these are classified as meditation practices with only a light-intensity component of physical activity when performing respiration and poses [38-40].

Following the recommendation by Munn et al [35] to clearly report the exposure or risk factor and how it was measured or identified, we excluded studies that proposed only a simple question on physical activity without any reference to its frequency (eg, a generic “Yes/No” question on practicing “exercise” or even “regular exercise”). Furthermore, we excluded literature reviews, gray literature, conference proceedings, and unpublished material. Finally, we excluded studies that lacked a quantitative evaluation of the association.

Information Sources

We searched the following electronic databases up to July 2022: MEDLINE through the Ovid platform, APA PsycArticles, and Scopus. Using a “snowballing” approach, we manually screened the reference lists of included articles and conducted systematic citation tracking in Scopus, PubMed, and Google Scholar.

Search Strategy

In total, 2 authors, CGL and PM, developed search strings using the PEO framework to structure the research question. We selected search terms to identify HCWs, physical activity interventions, and burnout. Regarding burnout, to increase search sensitivity, we also considered terms related to the assessment instruments indicated by Rotenstein et al [41] and those identified in the more recent works by Edú-Valsanía et al [42] and Shoman et al [43]. Moreover, to further increase the sensitivity of our search, we used the National Library of Medicine–controlled vocabulary thesaurus (Medical Subject Headings) with entry terms and synonyms. We limited the search to articles published in English. The complete search strategy can be found in Multimedia Appendix 2 [41-43].

Duplicates were removed by PM using an automatic procedure based on PubMed ID and digital object identifier, and this was performed in Microsoft Excel (Microsoft Corp).

Selection Process

Search results were retrieved from the databases and double screened independently by all the authors. The initial screening was based on the title and abstract using Rayyan (Rayyan Systems Inc) [44] for support. To refine and clarify the eligibility criteria, ensure consistency when applied by different reviewers, and train the team, a pilot phase was carried out on 500 works. The potentially relevant articles were retrieved for full-text screening, and their eligibility was determined as described in the previous step. Any disagreement was resolved through plenary discussion among all the authors until a consensus was reached.

Data Collection Process

Each selected study was randomly assigned to and independently evaluated by 2 authors to extract relevant data. Any disagreement was addressed as mentioned previously.

Data Items

We extracted 3 different types of information. First, we recorded the general characteristics, such as aim, context (country, workplace, and period), and population (type of HCW, number, and gender mix). Second, we summarized the methods used to measure burnout, the criteria used to quantitatively summarize the phenomenon in the study population, and a possible definition of severity. When possible, we included the frequency distributions of categorical outcomes and means of total scores with SDs. Third, we noted the methods adopted to measure physical activity, including the assessment of frequency, time, or intensity of activity and the percentage of participants practicing physical activity. Finally, we reported the adopted measures of the association between physical activity and burnout along with its strength (P values, odds ratios, and their 95% CIs), eventually derived by AB from published data, and a synthesis of the reported evidence. In addition, factors other than physical activity that are significantly associated with burnout were listed. A first list was drawn from bivariate analyses and reported only variables that were significantly associated (P<.05) with burnout. The consideration of even weak associations was hampered by the often incomplete presentation of these analyses. A second list came from multivariable analyses and included all the variables considered in the final (or presented) models regardless of whether physical activity was included in the model itself.

Synthesis of Evidence

The data are presented in a tabular form. A narrative approach was adopted to provide an overall summary of the findings of the included studies and their biases, strengths, and limitations, with an in-depth discussion of the causes of heterogeneity.

Quality

The quality of the studies included in the analysis was assessed using the Joanna Briggs Institute Critical Appraisal Checklist for Cross-Sectional Studies [45]. This checklist consists of 8 items that are rated as “yes,” “no,” “unclear,” or “not applicable.” To further refine our judgment, we added “partial” as a fifth option.
Owing to the eligibility criteria adopted, 2 issues were always rated positively: item 1 (“Were the criteria for inclusion in the sample clearly defined?”), as we only considered studies that clearly reported associations in health personnel directly involved in the provision of care services, and item 4 (“Were objective, standard criteria used for measurement of the condition?”), as we only included studies that used validated questionnaires for burnout.

Item 2 allowed us to register whether a complete description of the participants and setting was provided. We assigned a score of “partial” or “no” if some or all the information was missing.

To assess whether the outcomes were measured in a valid and reliable way (item 7), we verified that the burnout measurement tools were used consistently with the dictates expressed by their developers.

We evaluated the validity and reliability of the exposure measurement (item 3) by positively appraising the use of structured questionnaires (eg, the International Physical Activity Questionnaire–Short Form [IPAQ-SF]) or automatic measurement devices such as pedometers, whereas the use of ad hoc questions to register the frequency or intensity of physical activity was considered a partial achievement.

Concerning the confounding factors, we did not assess how they were considered (item 5) and addressed in the statistical analyses (item 6) in studies exploring multiple wide-scope associations with burnout. On the other hand, we applied these items in studies that focused on a specific exposure factor (either physical activity or another variable of interest). In such cases, the use of statistical modeling, such as multivariable regression analysis, was considered a partial achievement in the absence of an advisable dedicated discussion on confounding factors.

For the appropriateness of the statistical analysis (item 8), a partial quality level was assigned if multivariable regression analysis, though possible, was not conducted or when it was conducted without an adequate method of model selection. Moreover, we assigned a partial level of quality if it was not possible to clearly understand all the details of the analyses because of omissions or results not being clearly reported.

Each study was evaluated for quality by 2 authors, and AB, a statistician, also reviewed items pertaining to statistics.

Discrepancies were resolved through plenary discussion among all authors until a consensus was reached. The quality assessment was not taken into account for eligibility purposes.

Results

Selected Studies

A total of 8937 records were identified, and after removing 2422 (27.1%) duplicates (Figure 1), 6515 (72.9%) publications remained following the initial screening. Of these 6515 publications, 6427 (98.65%) were excluded based on the title and abstract. These records were disregarded as they did not address physical activity, burnout, or health care personnel or because they were conference proceedings, reviews, or nonoriginal research (eg, letters or commentaries). Of the 88 studies selected for potential inclusion, 3 (3%) were not retrieved as they were published in journals not accessible through our organizations even after writing to the authors to request the accepted versions of their manuscripts [46-48]. Following full-text screening, 75% (64/85) of the records were excluded. Consequently, 21 independent studies involving 15,782 HCWs were included in this review (see Table 1 for details).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram. BO: burnout.
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Aim</th>
<th>Country</th>
<th>Period of the survey</th>
<th>Type of HCW</th>
<th>Sample size, N</th>
<th>Sex (female participants), %</th>
<th>Age (y)</th>
</tr>
</thead>
</table>
| Ajab et al [49], 2021 | To understand the availability of personal protective equipment and the levels of anxiety, depression, and burnout of HCWs | Asia; United Arab Emirates | From July 2020 to August 2020 | Physicians, nurses, allied health care professionals, and laboratory technicians; not specified occupation | 1290 | 78 | Mean 38.7 (SD 8.73)
| Alvares et al [50], 2020 | To assess the prevalence of and factors associated with burnout syndrome | America; Brazil | From November 2011 to June 2013 | Nurses and physicians in the ICU | 125 nurses and 116 physicians | Nurses: 87.2%, physicians: 59.5% | |
| Bin Dahn mash et al [51], 2020 | To explore the prevalence of burnout and its predictors in radiology residents to minimize burnout rates and improve radiology residents’ well-being | Asia; Saudi Arabia | February 2019 | Radiology residents | 108 | 46.3 | Mean 27.2 (SD 1.83)
| Chokri et al [52], 2021 | To determine the prevalence of burnout among health care professionals and investigate the relationship between burnout and sociodemographic characteristics, working systems, and the level of physical activity in health care professionals | Africa; Morocco | February 2021 | Nurses, physicians, and health technicans | 145 | 70 | Mean 38.3 (SD 9.44)
| Portero de la Cruz et al [53], 2020 | To estimate the burnout, perceived stress, job satisfaction, coping, and general health levels experienced by nurses working in Spanish emergency departments and analyze the relationship between the sociodemographic, occupational, and psychological variables and the occurrence of burnout syndrome among these professionals | Europe; Spain | From March 2016 to December 2016 | Nurses | 171 | 73.1 | Mean 47.85 (SD 8.11)
| Eckstein et al [54], 2022 | To examine how burnout is related to mindfulness, fulfillment, specialty choice, and other lifestyle factors | America; United States | From December 2019 to February 2020 | Physicians from different specialties (residents and attending physicians at an academic institution) | 60 | _e | Median 31 (range 25-70)
| Feng et al [55], 2018 | To determine the prevalence of burnout among ophthalmology residents through a national survey and associate burnout with demographic factors, year in training, practice setting, self-reported workload physical activity, and sleep; in addition, this survey sought to solicit comments from ophthalmology residents regarding factors that they personally felt positively and negatively affected wellness and quality of life | America; United States | From January 17, 2017, to March 18, 2017 | Ophthalmology residents from PGY-2 to PGY-4 | 267 | 46.1 | Mean 29.7 (SD 2.3)
<p>| Ghoraiashian et al [56], 2022 | To evaluate the frequency and factors associated with occupational burnout in orthopedic specialists and residents | Asia; Iran | 2019 | Orthopedic surgeons and residents | 180 | 5.6 | Mean 42.8 (SD 11.17) |</p>
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Aim</th>
<th>Country</th>
<th>Period of the survey</th>
<th>Type of HCW</th>
<th>Sample size, N</th>
<th>Sex (female participants), %</th>
<th>Age (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldberg et al [57], 1996</td>
<td>To measure the degree of burnout among emergency physicians and identify and rank predictive factors</td>
<td>America; United States</td>
<td>During the ACEP&lt;sup&gt;4&lt;/sup&gt; An- nual Scientif- ic Assembly years 1992-1995</td>
<td>Emergency physi- cians</td>
<td>1272</td>
<td>25.7</td>
<td>Mean 39</td>
</tr>
<tr>
<td>Hu et al [58], 2021</td>
<td>To investigate the severity of burnout and its associated factors among physicians and nurses in ICUs</td>
<td>Asia; China</td>
<td>From July 25, 2019, to July 30, 2019</td>
<td>Physicians and nurses in ICUs</td>
<td>2411</td>
<td>68.7</td>
<td>Mean 33.5 (SD 5.95)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Lebensohn et al [59], 2013</td>
<td>To study the associations between commonly used indicators of well-being (perceived stress, depression, burnout, and satisfaction with life) and wellness behaviors at the start of family medicine residency</td>
<td>America; United States</td>
<td>At the begin- ning of PGY-1 in the graduat- ing classes of 2012 and 2013</td>
<td>PGY-1 family medicine (residents)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>168</td>
<td>59.9</td>
<td>Median 29</td>
</tr>
<tr>
<td>McClafferty et al [60], 2021</td>
<td>To document the concerning state of burnout in early pediatric trainees and examine the potential of the University of Arizona Center for Integrative Medicine PIMR&lt;sup&gt;c&lt;/sup&gt; curriculum to provide interventions that address gaps in lifestyle behaviors with recognized association with burnout and how they might be introduced into residency training</td>
<td>America; United States</td>
<td>First trimester of residency in 4 consecutive years from 2012 to 2015</td>
<td>First-year pediatric residents</td>
<td>203</td>
<td>76</td>
<td>Mean 28</td>
</tr>
<tr>
<td>Ng et al [61], 2020</td>
<td>To examine the prevalence and severity of burnout and explore the factors (sociodemographic, lifestyle behaviors, and career satisfaction) associated with burnout among medical graduates up to 20 y after graduation</td>
<td>Asia; Hong Kong</td>
<td>From January 29, 2016, to April 15, 2016</td>
<td>Physicians entering medical school</td>
<td>447</td>
<td>43.6</td>
<td>Mean 34.1 (SD 6.0)</td>
</tr>
<tr>
<td>Olson et al [62], 2014</td>
<td>To determine the association between achievement of national physical activity guidelines and burnout in internal medicine resident physicians</td>
<td>America; United States</td>
<td>From September 2012 to October 2012</td>
<td>Internal medicine physicians (resi- dents)</td>
<td>76</td>
<td>47.3</td>
<td>Mean 29.2 (SD 2.9)</td>
</tr>
<tr>
<td>Panse et al [63], 2020</td>
<td>To assess burnout in plastic surgery residents</td>
<td>Asia; India</td>
<td>From March 2019 to April 2019</td>
<td>Plastic surgery resi- dents</td>
<td>185</td>
<td>28.4</td>
<td>—</td>
</tr>
<tr>
<td>Reed et al [64], 2020</td>
<td>To explore the correlations between resident burnout and procedure volume, nonclinical responsibilities, and mindfulness practices along with gathering updated work hour data</td>
<td>America; United States</td>
<td>Not reported</td>
<td>Otolaryngology resi- dents</td>
<td>182</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Shanafelt et al [65], 2012</td>
<td>To evaluate the health habits, routine medical care practices, and personal wellness strategies of American surgeons and explore associations with burnout and quality of life</td>
<td>America; United States</td>
<td>October 2010</td>
<td>Surgeons</td>
<td>7197</td>
<td>14.6</td>
<td>Median 53</td>
</tr>
<tr>
<td>Tiwari et al [66], 2020</td>
<td>To investigate the prevalence of burnout among rheumatology practitioners and its associations</td>
<td>America; United States&lt;sup&gt;2&lt;/sup&gt;</td>
<td>February 2019</td>
<td>Rheumatology prac- titioners</td>
<td>128</td>
<td>53.7</td>
<td>Mean 49.9 (SD 12.0)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Study, year</td>
<td>Aim</td>
<td>Country</td>
<td>Period of the survey</td>
<td>Type of HCWa</td>
<td>Sample size, N</td>
<td>Sex (female participants), %</td>
<td>Age (y)</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------</td>
<td>-------------------------------</td>
<td>-----------------------------------</td>
<td>----------------</td>
<td>-------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Vinnikov et al [67], 2019</td>
<td>To assess burnout prevalence in physicians and nurses of a cardiological hospital and ascertain whether smoking, alcohol, and physical activity may predict job-associated burnout</td>
<td>Asia; Kazakhstan</td>
<td>September 2018</td>
<td>Cardiology personnel: physicians, nurses, and technical personnel</td>
<td>259</td>
<td>82</td>
<td>Median 34</td>
</tr>
<tr>
<td>Vinnikov et al [68], 2021</td>
<td>To verify the prevalence of occupational burnout in oncology physicians and nurses 2020 (before the breakout of COVID-19)</td>
<td>Asia; Kazakhstan</td>
<td>2020 (before the breakout of COVID-19)</td>
<td>Oncology physicians and nurses</td>
<td>256</td>
<td>62</td>
<td>Median 37.5</td>
</tr>
<tr>
<td>Yang et al [69], 2018</td>
<td>To measure the prevalence of burnout and resilience levels in transplant nurses, identify any relationships between the 2 concepts, and determine whether demographic factors were associated with burnout in this group</td>
<td>Asia; China</td>
<td>From July 2015 to November 2015</td>
<td>Transplant nurses</td>
<td>536</td>
<td>96.3</td>
<td>Mean 28.40 (SD 4.80)</td>
</tr>
</tbody>
</table>

aHCW: health care worker.
bEstimated by the authors from grouped data.
cICU: intensive care unit.
dCalculated by the authors as, in the paper, the reported percentages were weighted for the unequal probabilities of participant selection.
eNot available.
fPGY-2: postgraduate at year 2.
gPGY-4: postgraduate at year 4.
hACEP: American College of Emergency Physicians.
iPGY-1: postgraduate at year 1.
jPIMR: Pediatric Integrated Medicine in Residency.
kThe 10-hour PIMR curriculum is designed in part to help pediatric programs meet new resident well-being requirements. The topics covered include self-nutrition and physical activity, mind-body therapies, dietary supplements, whole systems of medicine, and clinical applications.
lThe location of the conference where the survey was administered was considered to determine the country.

Although 10% (2/21) of the studies explicitly focused on exploring the association between physical activity and burnout [62,67], the remaining studies investigated the association among several factors, including physical activity. Most of the studies were conducted in America (10/21, 48%) and Asia (9/21, 43%), with only 10% (2/21) of the studies [49,52] being conducted after the spread of COVID-19. In total, 43% (9/21) of the studies [51,54-56,59,60,62-64] focused on residents. All the studies included in this review (21/21, 100%) had a cross-sectional design.

Demography of Study Populations
The study populations consisted of diverse groups of HCWs with varying proportions of women individuals, ranging from 5.6% to 96.3%. The professional categories represented included physicians, nurses, and technicians from different specialties, which were present in 90% (19/21), 38% (8/21), and 14% (3/21) of the studies, respectively. The sample sizes ranged from 60 to 7197 participants. The mean age varied from 27.2 (SD 1.83) years to 49.9 (SD 12.0) years, with a few studies (5/21, 24%) reporting the median ages, ranging from 29 to 53 years.

Measurement of Burnout
In terms of burnout measurement, the 22-item Maslach Burnout Inventory (MBI) was the most commonly used questionnaire. Some studies (3/21, 14%) used MBI-related measures based on fewer items [63-65]. In total, 10% (2/21) of the studies used different questionnaires: the Professional Fulfillment Index [54], which is designed to measure physician well-being [70], and the Copenhagen Burnout Inventory [61], which was proposed as an alternative tool to the MBI for measuring burnout [71] (see Multimedia Appendix 3 [49,51-53,55,56,58-63,66-69,72,73] for details).

The severity of the condition in the investigated population was reported in all the studies (21/21, 100%), albeit with differences in terms of the dimensions considered and the criteria used for deriving a summary assessment of burnout across the dimensions and, in some cases, by providing measures of location and scale parameters. Further details are provided in Table 2.
Table 2. Burnout levels in the study populations.a

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Measure of burnout (MBI)b</th>
<th>Frequency distribution</th>
<th>Criteria for burnout</th>
<th>Location and scale parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chokri et al [52], 2021</td>
<td>MBI; normative data</td>
<td>65%, 15%, and 21%</td>
<td>EEc, DPd, PAe</td>
<td>4 classes—high: EE+, DP+, and PA--; moderate: 2 dimensions criticalb; low: 1 dimension critical; absent: no critical dimension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48%, 28%, and 23%</td>
<td></td>
<td>Burnout</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32%, 25%, and 36%</td>
<td></td>
<td>EE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11%, 41%, and 17%</td>
<td></td>
<td>DP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not considered</td>
<td></td>
<td>PA</td>
</tr>
<tr>
<td>Porter de la Cruz et al [53], 2020</td>
<td>MBI; normative data</td>
<td>21%, 19%, and 60%</td>
<td></td>
<td>3 classes—high: EE+, high DP+, and PA--; moderate: if not high and not absent; absent: EE−, DP−, and PA−</td>
</tr>
<tr>
<td></td>
<td></td>
<td>43%, 28%, and 29%</td>
<td></td>
<td>Burnout score defined as the sum of the scores in the 3 dimensions1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53%, 20%, and 26%</td>
<td></td>
<td>No threshold defined</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not considered</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Tiwari et al [66], 2020</td>
<td>MBI; normative data</td>
<td>38%, NR, and NR</td>
<td>Binary classification: positive if EE+ or DP+ and PA−</td>
<td>Mean 24.38 (SD 9.38)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>31%, NR, and NR</td>
<td></td>
<td>Mean 7.83 (SD 6.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR, NR, and 21%</td>
<td></td>
<td>Mean 29.38 (SD 8.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51%</td>
<td></td>
<td>Mean 61.59 (SD 17.72)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Lebensohn et al [59], 2013</td>
<td>MBI; normative data</td>
<td>14%, 28%, and 58%</td>
<td>Not considered</td>
<td>3 classes—high: EE+ and DP+; moderate: if not high and not low; low: EE− and DP−</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24%, 26%, and 51%</td>
<td></td>
<td>9%, 51%, and 41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not considered</td>
<td></td>
<td>Mean 6.38 (SD 4.7)</td>
</tr>
<tr>
<td>McClafferty et al [60], 2021</td>
<td>MBI; normative data</td>
<td>20%, 25%, and 55%</td>
<td></td>
<td>Binary classification: high if EE+ and DP+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32%, 27%, and 42%</td>
<td></td>
<td>Mean 18.1 (SD 9.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR</td>
<td></td>
<td>Mean 7.4 (SD 4.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean 29.5 (SD 6.3)</td>
</tr>
<tr>
<td>Olson et al [62], 2014</td>
<td>MBI; normative data</td>
<td>NR, NR, and NR</td>
<td>Not considered</td>
<td>Binary classification: high if EE+ or DP+ and PA−</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR, NR, and NR</td>
<td></td>
<td>Mean 24.0 (SD 11.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR, NR, and 65%</td>
<td></td>
<td>Mean 29.0 (SD 6.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Binary classification: high if EE+ or DP+ and PA−</td>
<td></td>
<td>Mean 27.3 (SD 8.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24%</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Bin Dahmash et al [51], 2020</td>
<td>MBI; scoring key</td>
<td>57%, NR, and NR</td>
<td></td>
<td>Binary classification: high if EE+ and DP+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32%, NR, and NR</td>
<td></td>
<td>Mean 6.38 (SD 4.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR, NR, and 65%</td>
<td></td>
<td>Mean 18.1 (SD 9.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Binary classification: high if EE+ or DP+ and PA−</td>
<td></td>
<td>Mean 7.4 (SD 4.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>63%</td>
<td></td>
<td>Mean 29.5 (SD 6.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Feng et al [55], 2018</td>
<td>MBI; scoring key</td>
<td>55%, 29%, and 17%</td>
<td>Binary classification: high if EE+ or DP+ and PA−</td>
<td>63%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>46%, 25%, and 28%</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71%, 18%, and 12%</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Binary classification: high if EE+ or DP+ and PA−</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>54%</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Hu et al [58], 2021</td>
<td>MBI; scoring key</td>
<td>Physicians: 61%, 33%, and 6%; nurses: 57%, 37%, and 7%</td>
<td>Binary classification: high if EE+ or DP+ and PA−</td>
<td>Physicians: 71%; nurses: 68%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physicians: 37%, 36%, and 27%; nurses: 31%, 37%, and 32%</td>
<td></td>
<td>Physicians: 64%, 16%, and 20%; nurses: 66%, 14%, and 20%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physicians: 64%, 16%, and 20%; nurses: 66%, 14%, and 20%</td>
<td></td>
<td>Physicians: 64%, 16%, and 20%; nurses: 66%, 14%, and 20%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physicians: 64%, 16%, and 20%; nurses: 66%, 14%, and 20%</td>
<td></td>
<td>Physicians: median 19 (IQR 15.8); nurses: median 18 (IQR 17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physicians: median 19 (IQR 15.8); nurses: median 18 (IQR 17)</td>
<td></td>
<td>Physicians: median 41 (IQR 9); nurses: median 38 (IQR 17)</td>
</tr>
</tbody>
</table>

a Numbers represent percentages. b MBI, Maslach Burnout Inventory. c EE, emotional exhaustion. d DP, depersonalization. e PA, personal accomplishment.
<table>
<thead>
<tr>
<th>Study, year</th>
<th>Measure of burnout (MBI&lt;sup&gt;a&lt;/sup&gt;)</th>
<th>Frequency distribution</th>
<th>Location and scale parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vin-nikov et al [68], 2021</td>
<td>MBI; scoring key</td>
<td>Reported data were not consistent with the reported categorization</td>
<td>Median 29 (IQR 15.8)</td>
</tr>
<tr>
<td>Ajab et al [49], 2021</td>
<td>MBI; Ajab et al [49]</td>
<td>Reported data were not consistent with the reported categorization</td>
<td>Median 15 (IQR 10)</td>
</tr>
<tr>
<td>Gho-raishian et al [56], 2022</td>
<td>MBI; Gho-raishian et al [56]</td>
<td>Reported data were not consistent with the reported categorization</td>
<td>Median 26 (IQR 15.8)</td>
</tr>
<tr>
<td>Panse et al [63], 2020</td>
<td>9-item MBI (aMBI&lt;sup&gt;m&lt;/sup&gt;) [74]</td>
<td>Not considered</td>
<td>Not considered</td>
</tr>
<tr>
<td>Gold-berg et al [57], 1996</td>
<td>MBI</td>
<td>Not considered</td>
<td>Not considered</td>
</tr>
<tr>
<td>Alvares et al [50], 2020</td>
<td>MBI</td>
<td>Not considered</td>
<td>Not considered</td>
</tr>
<tr>
<td>Shanafelt et al [65], 2012</td>
<td>2 single-item measures adapted from the MBI [75]&lt;sup&gt;a&lt;/sup&gt;</td>
<td>EE reported at least weekly: 23%</td>
<td>Binary classification: high if EE+ or DP+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria for burnout</th>
<th>Burnout</th>
<th>EE</th>
<th>DP</th>
<th>PA</th>
<th>Burnout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not considered</td>
<td>N/A</td>
<td>Median 26 (IQR 19)</td>
<td>Median 15 (IQR 10)</td>
<td>Median 29 (IQR 15.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Not considered</td>
<td>N/A</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>N/A</td>
</tr>
<tr>
<td>Not considered</td>
<td>N/A</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>N/A</td>
</tr>
<tr>
<td>Not considered</td>
<td>N/A</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>N/A</td>
</tr>
<tr>
<td>Not considered</td>
<td>N/A</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>N/A</td>
</tr>
<tr>
<td>Not considered</td>
<td>N/A</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>N/A</td>
</tr>
<tr>
<td>4 classes—severe: EE+, DP+, and PA; moderate: 2 dimensions critical; mild: 1 dimension critical; absent: no critical dimension</td>
<td>49%</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Burnout score defined as the sum of the scores in EE and DP and classified as binary (cutoff of ≥19; no reference provided for the validation of this cutoff)</td>
<td>7%, 16%, 27%, and 50%</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>N/A</td>
</tr>
<tr>
<td>Binary classification: moderate to high levels of burnout according to the Golembiewski classification&lt;sup&gt;b&lt;/sup&gt;</td>
<td>61%</td>
<td>Mean 23.31 (SD 8.55)</td>
<td>Mean 20.70 (SD 8.49)</td>
<td>Mean 24.72 (SD 9.17)</td>
<td>NR</td>
</tr>
<tr>
<td>2 binary classifications: (1) critical values in all the dimensions and (2) critical values in at least one dimension</td>
<td>Physicians: (1) 1% and (2) 34%; nurses: (1) 0% and (2) 39%</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>N/A</td>
</tr>
<tr>
<td>EE reported at least weekly: 23%</td>
<td>N/A</td>
<td>Binary classification: high if EE+ or DP+</td>
<td>27%</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Study, year | Measure of burnout (MBI\(^{a}\)) | Frequency distribution | Location and scale parameters
---|---|---|---
Reed et al [64], 2020 | 1 single question [76] | EE\(^{c}\) | N/A | N/A | N/A | N/A | 50% | N/A | N/A | N/A | N/A | N/A

\(^{a}\)Frequency distribution from high to low. Independent of the way the authors reported the frequency distribution for PA, in accordance with Maslach et al [72], we considered low PA as a condition characterizing burnout. The numbers in the frequency distribution are rounded to percentage units.

\(^{b}\)MBI: Maslach Burnout Inventory. In total, 3 categories were defined: emotional exhaustion, depersonalization, and PA. Most recent edition: Maslach et al [77].

\(^{c}\)EE: emotional exhaustion.

\(^{d}\)DP: depersonalization.

\(^{e}\)PA: personal accomplishment.

\(^{f}\)Italics indicate critical values for burnout.

\(^{g}\)The burnout dimension is in the high range (+) or low range (−).

\(^{h}\)When referring to MBI and not otherwise specified, critical levels in burnout dimensions mean high in EE and DP and low in PA.

\(^{i}\)Authors’ interpretation.

\(^{j}\)N/A: not applicable.

\(^{k}\)NR: not reported.

\(^{l}\)The proposed burnout score does not consider that PA should be interpreted in the opposite direction from EE and DP as a low degree of burnout is reflected in high scores on PA [72].

\(^{m}\)aMBI: abbreviated Maslach Burnout Inventory.

\(^{n}\)Raw data from the MBI used to divide the burnout process into 8 phases, with phases I-III representing a low degree and phases IV-V and VI-VIII representing moderate and high degrees, respectively [78].

\(^{o}\)MBI adapted.

\(^{p}\)MiniZ (adapted).

In two studies, authors adopted tools not referable to BMI. Ng et al [61] adopted the Copenhagen Burnout Inventory [71], which comprises 3 categories: personal, physical, and psychological exhaustion, work-related physical and psychological exhaustion, and patient-related physical and psychological exhaustion. The reported frequency distribution among study participants is 63%, 56%, and 35% respectively. No specific criteria were defined for burnout.

Eckstein et al [54] used the Professional Fulfillment Index [70], which consists of three response categories: Professional Fulfillment (PF), Interpersonal Disengagement, and Work Exhaustion (WE), employing 5-point Likert scales (“not at all true” to “completely true” for PF items and “not at all” to “extremely” for WE and Interpersonal Disengagement items). All responses are scored from 0 to 4. However, in the considered study, the PF domain was not taken into account. A binary classification (presence vs absence of burnout) is adopted according to the following criterion: the average score of Interpersonal Disengagement and WE ≥1.33 [70]. Participants meeting this criterion for burnout comprise 38%. Regarding Location and scale parameters, the median score for WE, Interpersonal Disengagement, and burnout is respectively 1.50, 0.83, and 1.00.

With due caution regarding the aforementioned heterogeneity of definitions and cutoffs, the percentage of participants classified as burned out varied from 7% to 83%.

**Measurement of Physical Activity**

None of the included studies used objective measurement tools, such as pedometers, to assess physical activity (Table 3).

Structured questionnaires that distinguished among different typologies of physical activity or ad hoc items were considered to investigate the habits of HCWs. The IPAQ-SF, the Ricci-Gagnon scale, and the Arizona Lifestyle Inventory were used in the studies by Olson et al [62], Chokri et al [52], and McClafferty et al [60], respectively. The IPAQ-SF was the only validated tool used, although it tends to overestimate actual physical activity levels [84]. A total of 33% (7/21) of the studies [50,53,61,63,67-69] used an ad hoc question to assess the achievement of a given benchmark, with a threshold frequency ranging from 1 day every 2 weeks [63] to every day [53]. In the remaining studies, physical exercise was measured in terms of frequency and intensity [65], frequency and time [57], only frequency [49,51,54-56,58,59,66], or only time [64]. When it was possible to report a distribution of frequencies or times, we considered the lowest category that did not indicate “no physical activity at all.” Whenever possible, we adopted the most widely used benchmark of physical exercise of “at least 1 day per week” to define physically active participants. Using this criterion, we found that the percentage of active workers ranged from 44% to 82% for residents and from 46% to 87% for the other categories.
### Table 3. Level of physical activity in the study populations.

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Method used to measure physical activity</th>
<th>Percentage of HCWs&lt;sup&gt;a&lt;/sup&gt; practicing physical activity</th>
<th>Criterion for prevalence</th>
<th>Prevalence, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olson et al [62], 2014</td>
<td>IPAQ-SF&lt;sup&gt;b&lt;/sup&gt; [79]</td>
<td>Compliance with the DHHS&lt;sup&gt;c&lt;/sup&gt; benchmark: ≥150 min/wk [80]</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Chokri et al [52], 2021</td>
<td>Ricci-Gagnon scale [81]</td>
<td>Not inactive</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>McClafferty et al [60], 2021</td>
<td>Arizona Lifestyle Inventory [82]</td>
<td>≥1 d/wk for ≥30 min/session of moderate physical activity</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>Porter de la Cruz et al [53], 2020</td>
<td>Question assessing a benchmark achievement</td>
<td>7 d/wk</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Ng et al [61], 2020</td>
<td>Question assessing a benchmark achievement</td>
<td>Performing regular exercise: ≥5 d/wk for ≥10 min/session of any vigorous or moderate physical activities</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Vinnikov et al [67], 2019</td>
<td>Question assessing a benchmark achievement</td>
<td>Performing regular exercise: ≥3 d/wk for ≥40 min/session of any off-work physical activity</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Vinnikov et al [68], 2021</td>
<td>Question assessing a benchmark achievement</td>
<td>Performing regular exercise: ≥3 d/wk of any physical activity</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Alvares et al [50], 2020</td>
<td>Question assessing a benchmark achievement</td>
<td>Performing regular exercise: ≥3 d/wk of any physical activity</td>
<td>Nurses: 16&lt;sup&gt;d&lt;/sup&gt;, physicians: 51.7&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Yang et al [69], 2018</td>
<td>Question assessing a benchmark achievement</td>
<td>≥1 d/wk</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Panse et al [63], 2020</td>
<td>Question assessing a benchmark achievement</td>
<td>≥1 d for 2 wk</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Shanafelt et al [65], 2012</td>
<td>Frequency and intensity</td>
<td>&gt;30 min/wk (moderately intense aerobic exercise)</td>
<td>≥75</td>
<td></td>
</tr>
<tr>
<td>Shanafelt et al [65], 2012</td>
<td>Frequency and intensity</td>
<td>&gt;30 min/wk (vigorously intense aerobic exercise)</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Shanafelt et al [65], 2012</td>
<td>Frequency and intensity</td>
<td>Compliance with the CDC&lt;sup&gt;e&lt;/sup&gt; recommendation for aerobic exercise and muscle strength training [83]</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Goldberg et al [57], 1996</td>
<td>Frequency and time</td>
<td>≥1 d/wk, ≥10 min/session</td>
<td>≥78&lt;sup&gt;f&lt;/sup&gt;, ≥79&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Feng et al [55], 2018</td>
<td>Frequency</td>
<td>Compliance with the DHHS benchmark: ≥150 min/wk [80]</td>
<td>35.2</td>
<td></td>
</tr>
<tr>
<td>Ajab et al [49], 2021</td>
<td>Frequency</td>
<td>≥1 d/wk</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Bin Dahmash et al [51], 2020</td>
<td>Frequency</td>
<td>≥1 d/wk</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Eckstein et al [54], 2022</td>
<td>Frequency</td>
<td>≥1 d/wk</td>
<td>≥78&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Hu et al [58], 2021</td>
<td>Frequency</td>
<td>≥1 d/wk</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Lebensohn et al [59], 2013</td>
<td>Frequency</td>
<td>≥1 d/wk</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Tiwari et al [66], 2020</td>
<td>Frequency</td>
<td>≥1 d/wk</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Ghoraiashian et al [56], 2022</td>
<td>Frequency</td>
<td>A clear cutoff was not indicated</td>
<td>N/A&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Reed et al [64], 2020</td>
<td>Time</td>
<td>A clear cutoff was not indicated</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>HCW: health care worker.

<sup>b</sup>IPAQ-SF: International Physical Activity Questionnaire–Short Form.

<sup>c</sup>DHHS: US Department of Health and Human Services.

<sup>d</sup>Calculated by the authors as the percentages reported in the paper were weighted for the unequal probabilities of participant selection.

<sup>e</sup>CDC: Centers for Disease Control and Prevention.

<sup>f</sup>Authors’ interpretation.

<sup>g</sup>N/A: not applicable.

### Methods of Association Assessment

Bivariate associations between physical activity and burnout were assessed using usual methods. Table 4 provides further details. All studies except those by Olson et al [62], Vinnikov et al [67], Ghoraiashian et al [56], Feng et al [55], and Tiwari et al [66] also conducted a multivariable regression analysis. In all 5 cases, a multivariable analysis could have been conducted to obtain adjusted odds ratios. Logistic and linear regression were equally used in 43% (9/21) and 33% (7/21) of the studies, respectively.
### Table 4. Association between physical activity and burnout.

<table>
<thead>
<tr>
<th>Study, year</th>
<th>Data analysis method</th>
<th>Bivariate analysis</th>
<th>Multivariable regression analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ajab et al [49], 2021</td>
<td>Bivariate analysis and multivariable linear regression</td>
<td>Low levels of EE$^a$ and DP$^b$ were significantly more frequent among HCWs$^c$ who performed physical activity almost every day than among HCWs who were not physically active the previous week ($P&lt;.001$). High levels of personal accomplishment were significantly more frequent among HCWs who performed physical activity almost every day or every day than among HCWs not physically active the previous week ($P&lt;.001$)</td>
<td>Not clearly reported and not coherent with the statistical methods declared</td>
</tr>
<tr>
<td>Alvares et al [50], 2020</td>
<td>Chi-square and Fisher exact tests and multivariable logistic regression</td>
<td>Nurses: participants who did not exercise $&gt;3$ d/wk$^d$ were at higher risk of high levels$^e$ of EE (OR $^{f}$ 7.36, 95% CI 1.14-47.32) and at lower risk of high levels$^e$ of DP (OR 0.05, 95% CI 0.004-0.61); physicians: crude ORs were not significant</td>
<td>Nurses: adjusted for other covariates, participants who did not exercise $&gt;3$ d/wk$^d$ were at higher risk of high levels of EE (OR 11.01, 95% CI 2.73-44.39) and at lower risk of high levels of DP (OR 0.07, 95% CI 0.007-0.79)</td>
</tr>
<tr>
<td>Bin Dah-mash et al [51], 2020</td>
<td>Univariable and multivariable logistic regression</td>
<td>Participants who exercised $\geq 1$ d/wk were significantly less at risk of having high DP (OR 0.33, 95% CI 0.13-0.78), low personal accomplishment (OR 0.43, 95% CI 0.19-0.97), or high burnout (OR 0.29, 95% CI 0.10-0.77; $P=0.01$) than those who never exercised</td>
<td>Participants who exercised $\geq 1$ d/wk were significantly less at risk of having high DP (aOR $^g$ 0.38, 95% CI 0.15-1; $P=.04$) than those who never exercised</td>
</tr>
<tr>
<td>Chokri et al [52], 2021</td>
<td>Chi-square test and multivariable linear regression</td>
<td>Degree of physical activity was not significantly associated with degree of EE ($P=.86$); it was weakly associated with DP and personal accomplishment ($P=.09$ and $P=.08$, respectively)</td>
<td>Physical activity not included in the model</td>
</tr>
<tr>
<td>Porter de la Cruz et al [53], 2020</td>
<td>Bivariate analysis and univariable and multivariable linear regression</td>
<td>Those who did not take part in daily physical exercise had higher mean DP ($P=.005$) scores. There were no statistically significant differences in mean EE ($P=.09$) and mean personal accomplishment ($P=.48$) according to daily physical exercise</td>
<td>Adjusted for other covariates, the lack of daily physical activity was a significant predictor of higher DP values</td>
</tr>
<tr>
<td>Eckstein et al [54], 2022</td>
<td>Univariable and multivariable logistic regression</td>
<td>Frequency of exercise not significantly associated with burnout</td>
<td>Physical activity not included in the model</td>
</tr>
<tr>
<td>Feng et al [55], 2018</td>
<td>Bivariate analyses</td>
<td>The probability of low EE was significantly higher in participants who engaged in physical activity $\geq 150$ min/wk ($P=.02$). No association was found with levels of DP ($P=.32$), personal accomplishment ($P=.29$), and burnout ($P=.13$)</td>
<td>NP$^h$</td>
</tr>
<tr>
<td>Ghoraishian et al [56], 2022</td>
<td>Chi-square test and univariable logistic regression</td>
<td>Participants who exercised $\leq 1$ h/wk were at higher risk of burnout than those who exercised $&gt;1$ h/wk (OR 2.3, 95% CI 1.24-4.48)</td>
<td>NP</td>
</tr>
<tr>
<td>Goldberg et al [57], 1996</td>
<td>Chi-square test and multivariable logistic regression</td>
<td>Results of association analysis were inconsistent with the data.</td>
<td>Low levels of exercise were significantly associated with burnout</td>
</tr>
<tr>
<td>Hu et al [58], 2021</td>
<td>Chi-square test and multivariable logistic regression</td>
<td>Not reported in the paper</td>
<td>Participants who exercised $\geq 1$ d/wk were significantly at lower risk of burnout (once a week: OR 0.66, 95% CI 0.45-0.95; every 2 or 3 d: OR 0.56, 95% CI 0.39-0.80; every day: OR 0.52, 95% CI 0.36-0.75) than those who exercised less frequently or never</td>
</tr>
<tr>
<td>Lebensohn et al [59], 2013</td>
<td>ANOVA and multivariable linear regression</td>
<td>Physical activity was not significantly associated with burnout</td>
<td>More frequent physical activity was a significant adjusted predictor of lower values of both EE and DP</td>
</tr>
<tr>
<td>McClafferty et al [60], 2021</td>
<td>ANOVA and multivariable linear regression</td>
<td>Among individuals at high risk of burnout, the frequency of physical activity was lower than the group mean, whereas among individuals at low or moderate risk, the frequency was higher ($P=0.10$)</td>
<td>Adjusted for other covariates, a higher frequency of exercise was a significant predictor of a higher score on personal accomplishment</td>
</tr>
<tr>
<td>Ng et al [61], 2020</td>
<td>Univariable and multivariable linear regression</td>
<td>Practicing regular exercise significantly reduced CBI-PeE$^{l}$ (slope: $-9.882; P&lt;.001$) and CBI-PaE$^{l}$ (slope: $-6.932; P=.004$). It was not correlated with CBI-WrE$^{l}$</td>
<td>Practicing regular exercise was a significant adjusted predictor of lower values of CBI-PeE and CBI-PaE</td>
</tr>
</tbody>
</table>

---

https://publichealth.jmir.org/2024/1/e49772

JMIR Public Health Surveill 2024 | vol. 10 | e49772 | p.746

*(page number not for citation purposes)*
Table 4 also presents the evidence of the association between physical activity and burnout. Owing to the high heterogeneity of the studies, a direct comparison of the results was not feasible. As previously mentioned, the 2 sources of heterogeneity were the diverse and sometimes vague definitions of physical activity used in the studies and the variations in the definitions and cutoffs for measuring burnout. A total of 14% (3/21) of the studies [49,57,64] reported a few results that were not consistent with the data, and in one case [58], the authors referred to supplementary material that was not available on the journal web page. In total, 80% (4/5) [55,56,62,66] of the studies that presented only bivariate analysis were consistent with the data. Compliance with CDC aerobic exercise and muscle strength training recommendations was not independently associated with burnout. Physical activity was not significantly associated with burnout. The definition of burnout adopted in the logistic regression was not clear. It was also a significant predictor of lower values of personal accomplishment. Participants with moderate burnout in the dimension were excluded from the analysis. The result was reported for completeness even though we believe that the burnout score was meaningless as it did not consider that personal accomplishment should be interpreted in the opposite direction from EE and DP.

**Existence and Degree of Association**

Participants conforming with DHHS guidelines were significantly less at risk of having high subscores in burnout (OR 0.38, 95% CI 0.147-0.99) than noncompliant HCWs. Physical activity was not significantly associated with burnout. Participants who performed any physical activity for fitness at least once in 2 weeks were significantly less at risk of high values of burnout (OR 0.41, 95% CI 0.22-0.77) than those who did not perform physical activity. Compliance with CDC aerobic exercise and muscle strength training recommendations was not independently associated with burnout. Physical activity was not significantly associated with burnout. Participants who did not regularly exercise ≥3 d/wk were significantly more at risk of having high EE (aOR 9.91, 95% CI 2.92-27.2) than the other participants. Participants who did not regularly exercise were at higher risk of high EE (OR 5.02, 95% CI 2.25-12.42) and high DP (OR 2.37, 95% CI 1.20-4.74) than those who regularly exercised. Adjusted for other covariates, exercising every week was a significant predictor of lower values of EE and burnout. It was also a significant predictor of lower values of personal accomplishment.

**Study, year** | **Data analysis method** | **Bivariate analysis** | **Multivariate regression analysis**
--- | --- | --- | ---
Olson et al [62], 2014 | Chi-square test and univariate logistic regression | Participants compliant with DHHS guidelines were significantly less at risk of having high subscores in burnout (OR 0.38, 95% CI 0.147-0.99) than noncompliant HCWs | NP
Panse et al [63], 2020 | Univariable and multivariable logistic regression | Participants who performed any physical activity for fitness at least once in 2 weeks were significantly less at risk of high values of burnout (OR 0.41, 95% CI 0.22-0.77) than those who did not perform physical activity | Physical activity was not significantly associated with burnout
Reed et al [64], 2020 | Chi-square test, ANOVA, and multivariable logistic regression | Results of association analysis were inconsistent with the data | Physical activity was not significantly associated with burnout
Shanafelt et al [65], 2012 | Multivariable logistic regression | Participants who lacked exercise at least 1 d/wk were at higher risk of burnout (OR 5.00, 95% CI 1.3-18.5) than participants who exercised | NP
Tiwari et al [66], 2020 | Univariable logistic regression | Participants who lacked exercise at least 1 d/wk were at higher risk of burnout (OR 5.00, 95% CI 1.3-18.5) than participants who exercised | NP
Vinnikov et al [67], 2019 | Bivariate analysis | Physical activity was not significantly associated with any MBI dimension | NP
Vinnikov et al [68], 2021 | Bivariate analysis and multivariable logistic regression | Participants who did not regularly exercise were at higher risk of high EE (OR 5.02, 95% CI 2.25-12.42) and high DP (OR 2.37, 95% CI 1.20-4.74) than those who regularly exercised | Participants who did not regularly exercise ≥3 d/wk were significantly more at risk of having high EE (aOR 9.91, 95% CI 2.92-27.2) than the other participants
Yang et al [69], 2018 | Multivariable linear regression | Participants who lacked exercise at least 1 d/wk were at higher risk of burnout (OR 5.00, 95% CI 1.3-18.5) than participants who exercised | Adjusted for other covariates, exercising every week was a significant predictor of lower values of EE and burnout. It was also a significant predictor of lower values of personal accomplishment

**Notes**

- **EE**: emotional exhaustion.
- **DP**: depersonalization.
- **HCW**: health care worker.
- **Derived by the authors.**
- **Participants with moderate burnout in the dimension were excluded from the analysis.**
- **OR**: odds ratio.
- **aOR**: adjusted OR.
- **NP**: not performed.
- **CBI**: Copenhagen Burnout Inventory. In total, 3 categories were defined: personal physical and psychological exhaustion, work-related physical and psychological exhaustion, and patient-related physical and psychological exhaustion.
- **PeE**: personal physical and psychological exhaustion.
- **PaE**: patient-related physical and psychological exhaustion.
- **WrE**: work-related physical and psychological exhaustion.
- **DHHS**: US Department of Health and Human Services. Its guidelines [80] set a benchmark of 150 minutes per week of physical activity.
- **CDC**: Centers for Disease Control and Prevention. It set recommendations for aerobic exercise and muscle strength training [83].
- **The definition of burnout adopted in the logistic regression was not clear.**
- **MBI**: Maslach Burnout Inventory.
- **The result was reported for completeness even though we believe that the burnout score was meaningless as it did not consider that personal accomplishment should be interpreted in the opposite direction from EE and DP.**
analyses indicated that a lack of physical activity was associated with high values of at least one component of the MBI. Similar results were reported by half (8/16, 50%) of the other studies that presented bivariate analyses regardless of the measure of burnout adopted.

In studies in which bivariate analysis was preparatory to multivariable regression analysis [50-54,58,61,63,68], physical activity may not have emerged as a predictor in the multivariable models [51,52,54,63,68]. This is sometimes due to the lack of a significant (albeit weak) association already in the bivariate analyses [52,54]. However, this result is also influenced by the different approaches to variable inclusion and selection. When conducting a bivariate analysis as a preliminary step for variable selection in the multivariable model, a less stringent significance criterion than P<.05 should be considered. Variables that show weak individual associations can become important predictors when considered jointly (eg, the study by Hosmer and Lemeshow [85]). Unfortunately, the selected studies used very different criteria, ranging from P<.05 for each variable to no preselection at all. In addition, methods for model selection were not always applied or clearly stated.

In cases in which physical activity was found to be a significant predictor in multivariate analyses, a protective effect was observed, especially against emotional exhaustion [50,59,61,68,69] and depersonalization [50,51,53,59]. This held true across a range of activity frequencies, from as little as 1 session per week to daily engagement.

For the sake of completeness, factors other than physical activity that showed a statistically significant association with burnout have been listed in 2 separate tables: one for bivariate analyses and the other for multivariable analysis (Tables S1 and S2 in Multimedia Appendix 4). The predictors were grouped by topic (demography, health conditions, lifestyle, personal attitude, work-life balance, work organization and environment, work profile, and self-perception at work) and ordered within each class according to their frequency.

Quality Evaluation
The risk of bias was assessed using the Joanna Briggs Institute tool, and the details are presented in Multimedia Appendix 3 [45,49-69].

All the studies provided detailed descriptions of participants and settings, with the exception of the studies by Eckstein et al [54], in which gender specifications were missing, and Reed et al [64], in which the period of the survey, gender mix, and age were not reported.

Regarding the assessment of physical activity, only a minority of studies (3/21, 14%) [52,60,62] adopted structured questionnaires and, therefore, received a positive evaluation. Most studies (18/21, 86%) only partially fulfilled this criterion, either relying on simple ad hoc questions to assess the activity frequency, time, or intensity or referring to meeting a threshold of activity. This limitation was observed even in studies that claimed to have a specific focus on physical activity [55,67]. Confounding was never explicitly addressed. As mentioned in the Methods section, we applied the dedicated items to the 14% (3/21) of studies that focused on a specific exposure [62,67,69]. All these papers touched on this issue through their conducted analyses and, therefore, obtained a “partial” rating for item 5. Of these studies, 33% (1/3) conducted only a bivariate analysis and, thus, did not comply with item 6 [62].

As for the use of the MBI, only a few studies (9/21, 43%) received a positive evaluation for item 7. Contrary to the explicit indications of the MBI developers [86], the authors of most studies (11/21, 52%) [50,51,55-58,62,63,66-68] adopted a categorical classification of participants based on burnout dimensions for descriptive as well as inferential purposes. Moreover, Yang et al [69], although correctly adopting the score to assess each MBI dimension, defined and adopted an overall burnout score as the raw sum of the scores in each dimension (as reported in Table 4, even without considering that personal accomplishment should be interpreted in the opposite direction from emotional exhaustion and depersonalization). In all these cases except the studies by Ghoraishian et al [56] and Panse et al [63], we assigned a partial achievement of the item. Indeed, these 2 studies were ranked with a “no” as they did not even provide the rationale behind their choice of ranges of scores in each dimension. Except for 14% (3/21) of the studies—Ajab et al [49] and Vinnikov et al [67,68]—all the other studies referring to the MBI or related multi-item measures introduced a definition of overall burnout. Despite this, studies that also adopted the total score of the burnout dimensions in their inferential analyses [52,53,59,60] were ranked with a “yes.” Ng et al [61] and Eckstein et al [54] were classified as “yes” as they correctly adopted the measurement tools—Copenhagen Burnout Inventory [73] and Professional Fulfillment Index [54], respectively.

Regarding item 8, statistical analysis was not considered appropriate in 10% (2/21) of the studies [50,58], in which the significant differences between physicians and nurses in the dimensions of the MBI were not adequately handled in the multivariable analysis. The analyses conducted in another 14% (3/21) of the studies [49,56,64] were rated as unclear either because the presented results were not coherent with the declared statistical methods [64] or because of unexplained methods [49,56]. Goldberg et al [57], Bin Dahmash et al [51], Shanafeh et al [65], Vinnikov et al [68], Porterio de la Cruz et al [53], and Chokri et al [52] clearly explained their statistical methods and conducted comprehensive analyses, receiving a positive quality assessment. In the other cases, a better statistical methodology could have been applied by conducting a multivariable analysis, considering more suitable criteria to include the explanatory factors in the multivariable analysis, or again by applying model selection methods to obtain more parsimonious and general models.

Discussion
Principal Findings
The COVID-19 pandemic has taken a heavy toll on HCWs in terms of physical and mental health, and the long-term effects of the pandemic will further increase the burden of HCWs’ mental health disorders and burnout [87,88]. It has been argued that addressing burnout could serve as a nonstigmatized and
systemic approach to address a long-standing issue in medicine through mental health initiatives, whether prevention oriented or treatment focused [89].

In addition to the personal consequences for HCWs experiencing burnout, it is important to consider its impact on patient care and resource consumption. Emotionally and physically healthy HCWs are among the most relevant factors influencing health care service quality [90]. All these considerations emphasize the importance of implementing strategies to prevent and manage burnout at an individual, organizational, and cultural level. The literature on burnout management primarily presents fragmented solutions that are infrequently tested in practice. These solutions often align with 1 of 2 predominant lines of intervention: one emphasizes strengthening individual capabilities to navigate the inherent challenges of health care work, whereas the other acknowledges work organization as a contributing factor and attempts to intervene at that level. However, there is a growing realization that individual and organizational well-being are intricately interconnected, thus necessitating systemic solutions. These comprehensive approaches encompass organizational interventions, instilling a culture of well-being in the workplace, and integrating well-being into health care education for true efficacy [24,91].

A recent meta-analysis of 20 controlled trials found that the most effective existing interventions for reducing burnout were those targeting multiple facets of well-being [92]. Regrettably, these systemic solutions are complex and often come at a significant cost [93-95]; therefore, it is particularly relevant for research to precisely identify the characteristics that specific interventions must have to succeed. This review focused on physical activity as an effective factor in fostering a culture of well-being among HCWs, which is crucial for tackling the physical and mental consequences of work-related stress [24]. Physical activity has enormous potential to mitigate the physical and mental impacts of work-related stress [96,97]. Indeed, it is suggested that it facilitates psychological detachment from work and enhance self-efficacy [53], providing an opportunity to divert attention from stressful thoughts [50]. It has also been shown that moderate-intensity exercise training programs improve feelings of vigor, energy, and vitality [98]. In particular, greater effects occurred when combining resistance exercise with aerobic exercise compared with aerobic exercise alone [98]. This finding is consistent with the World Health Organization recommendations for physical activity, which underline the importance for adults of regularly performing both aerobic and muscle-strengthening activities to support health, including mental health outcomes [96]. With this review, we aimed to systematically assess the strength of the evidence and, eventually, the dose-response association between physical activity and burnout in a physically and emotionally exhausting work environment. However, it is essential to note that all studies included in our review had a cross-sectional design. As is well known, randomized controlled trials with a longitudinal perspective are the gold standard to highlight any potential cause-effect relationship between an exposure and an outcome. Nevertheless, the cross-sectional approach is useful in highlighting potential relationships between burnout and related factors, aiding in the identification of a multiplicity of risk factors and mitigation strategies, as the complexity of the phenomenon requires. This is particularly true in our case as most of the included studies (19/21, 90%) were exploratory investigations that examined various aspects potentially related to burnout rather than focusing specifically on physical activity.

As shown in Tables S1 and S2 in Multimedia Appendix 4, a wide range of variables were considered in bivariate analyses (74 issues) and included in multivariable models (58 issues), reflecting the interplay between burnout and demographic characteristics, health, lifestyle, personal attitude, self-perception at work, work organization and environment, job profile, and work-life balance.

As mentioned previously, physical activity is considered a potential tool to cope with distress, and the PEO approach adopted in this study was specifically structured in this direction. However, it is important to recognize the possibility that high levels of burnout negatively influence the level of physical activity [99]. According to the study by Olson et al [62], the “lack of energy” among burned-out residents suggests that high levels of burnout lead HCWs to reduce the time dedicated to leisure activities, including physical exercise. Stults-Kolehmainen and Sinha [99] found a similar result in the literature, stating that stress hinders individuals from being more physically active and has a negative influence on other health behaviors, including smoking, alcohol, and drug use. We suggest that future studies include sections dedicated to exploring the reasons for individuals’ inability to meet their desired level of physical activity. Factors such as time constraints, lack of interest, or underlying health conditions should also be accounted for. The Barriers to Being Active Quiz developed by the US Centers for Disease Control and Prevention [100] is an example of a tool that may support the exploration of the dynamic interactions among personal, socioenvironmental, and behavioral factors, serving as a knowledge base for promoting more active and healthy lifestyles.

This study confirms that the dose-response relationship between physical activity and psychological well-being and health-related quality of life is far from being established. Significant mental health benefits could be achieved even at physical activity levels below the public health recommendations [101,102]. This also emerged from our study and has direct implications on healthy lifestyle recommendations, especially for inactive HCWs for whom incorporating brief bouts of physical activity into daily activities may be a more realistically achievable goal than meeting the guideline-recommended physical activity levels. This can be crucial to promote physical activity and, therefore, trigger a virtuous circle with benefits for burnout. In addition, it may have broader implications for the general population as there are indications that physicians’ involvement in physical activity is linked to their propensity to advise patients on the advantages of exercise [103,104]. Recognizing that any physical activity is better than none and considering engagement in physical activity as a modifiable behavior, adopting a strategy of gradually increasing activity through small habit changes is deemed effective for establishing a consistent exercise routine. This approach can be facilitated by the use of activity trackers [105,106].

This systematic review highlighted another general challenge in quantifying the strength of the physical activity–burnout phenomenon.
association and establishing the dose-response curve, primarily as various methods were used to detect burnout. Different self-reporting tools are indeed available grounded on different theoretical bases [42]. Most of the studies in our review (19/21, 90%) proposed the 22-item MBI or MBI-related questionnaires (Table 2), confirming the substantial monopoly of this tool in burnout research. All these studies fell into some form of misuse of the index. The 3-factor structure of the MBI implies, on the one hand, that each dimension must be treated separately and, on the other hand, that none of them should be ignored. In contrast, in some studies where it would have made sense (3/17, 18%) [59,60,62], personal accomplishment was excluded from the analysis. This practice is not uncommon as several authors consider personal accomplishment not as a reaction to stressful situations but rather as a personality trait or coping resource and, therefore, as not contributing to the comprehensive concept of burnout [107]. Another misuse of the MBI is its diagnostic application, which erroneously considers the MBI dimensions as symptoms of burnout [71]. The 7-point scale (from Never to Every day) used to report the feelings experienced by the respondents was intended by Maslach and the coauthors of the tool as an operational simplification of the measurement of a dynamic phenomenon evolving continuously over time rather than as the assessment of a dichotomous condition (absent or present) defined through an arbitrary cutoff. Considering MBI scores for diagnostic purposes would inevitably invoke the wrong concept of burnout as a disease or disability, ignoring decades of research and the recent statement from the World Health Organization [1]. The MBI was not designed as a diagnostic tool [86], and the cutoff scores established to classify people at low, moderate, and high levels of burnout were “intended primarily as feedback for individual respondents.” These scores were published up to the third edition of the MBI Manual [72] accompanied, however, by a strong recommendation to use the original total scores for any statistical analysis. In the fourth edition released in 2018 [77], the categorization was finally removed. Therefore, it is surprising that, despite the extraordinary diffusion of the MBI, there was a failure to implement the correct instructions for its use. In fact, all studies except the one by Goldberg et al [57] considered some classification of the severity (sometimes referred to as risk) of burnout in each of the 3 MBI dimensions also for inferential purposes. Furthermore, a variety of classification criteria, even when taken from the same reference [72], and algorithms for combining dimensions into a single overall burnout indicator, typically a high score in at least one dimension and sometimes in all dimensions (reversing personal accomplishment, if considered), further increased the heterogeneity of the analyses. These methodological considerations become even more crucial in future research, particularly now that apps providing burnout self-diagnosis are available and transparency is needed in the adopted assessing algorithms [108].

Moving on to the evidence on the association between physical activity and burnout that emerged from this review, the most compelling results came from multivariable analyses that considered emotional exhaustion and depersonalization as the outcome. In these analyses, physical activity appeared to be associated with a reduction in critical conditions. However, it is still unclear whether this association depends on the type, intensity, duration, or frequency of physical activity, as previously mentioned. Some of the included studies (2/21, 10%) suggested that exercising for at least one day per week is sufficient to see a positive effect [58,69], whereas others (2/21, 10%) suggested a frequency of 3 days per week [50,68]. Lebensohn et al [59] observed that, the more frequently HCWs engaged in physical activity, the greater the positive effect.

It is worth noting that, in a few of the studies that considered a multivariable analysis (2/15, 13%), physical activity was not included among the predictors. However, this can sometimes be related to strict variable selection methods.

One limitation of most of the included studies (18/21, 86%) was related to the modality of physical activity measurement. Future studies should consider using objective measurement tools such as pedometers or validated questionnaires such as the International Physical Activity Questionnaire [79] even in its short version [84]. These approaches enhance the accuracy and reliability of data collection. On the other hand, relying on simple questions about regular physical activity or adherence to recommendations for a healthy lifestyle, although suitable for preliminary exploratory investigations, limits the comparability of studies.

The use of precise and detailed definitions to collect measures of intensity and frequency can prove to be a valuable strategy to delve deeper into the dose-response relationship.

Limitations

Our systematic review has some limitations that warrant consideration. First, our search was confined to 3 databases and only considered English-language articles. Moreover, we did not perform a rerun of the search shortly before submission. However, to mitigate the risk of overlooking relevant papers, we used forward and backward citation tracking, including the use of Google Scholar.

Another potential limitation arises from our eligibility criteria as we considered only HCWs in direct contact with patients and excluded practices with only a light component of physical exercise.

Finally, the heterogeneity in measurement methods and statistical analyses, which we have extensively covered in the Results section, made a meta-analysis inappropriate and precluded the determination of any pooled effect size.

Conclusions

Our comprehensive overview of studies exploring the association between physical activity and burnout in HCWs revealed a significant level of heterogeneity in definitions, measurements, and analyses adopted in the literature. Our work aimed to advance effective public health practices by addressing this critical issue in the existing evidence. It is important to adopt a clear definition of burnout and physical activity and make thoughtful choices regarding measurement tools and methodologies for data analysis. This becomes particularly crucial when considering that burnout is not a diagnosable disease but rather a multifaceted psychological syndrome that
emerges in response to chronic interpersonal stressors in the workplace.

Our findings strongly emphasize the beneficial connection between physical activity and burnout when a statistically significant association is present in the analyses. However, they also highlight the importance of a more in-depth investigation of the specific dependencies on exercise type, intensity, duration, and frequency, knowledge that currently represents a research gap in the field of burnout studies. Moreover, our considerations regarding the measurement of burnout and the comprehensive list of associated factors have the potential to enhance the quality of future studies. Our findings have significant implications for policy makers and health care professionals, underlining the importance of promoting physical activity as an easily accessible mitigation strategy for the well-being of the workforce and the overall effectiveness of the health care system.

Acknowledgments

The authors would like to thank Roberto Guarino (National Research Council of Italy—Institute of Clinical Physiology, Lecce) for his technical support (informatics tools for document management and title and abstract screening). This study was conducted within the research agreement between the National Research Council of Italy—Institute of Clinical Physiology, National Research Council of Italy—Institute for Applied Mathematics and Information Technologies, “E. Magenes,” and National Research Council of Italy—Institute for Research on Population and Social Policies and the Brindisi Local Health Unit (CNR-IRPPS-547/2022; deliberazione ASL-BR-1027/2022). The agreement was executed for free.

Data Availability

All data generated or analyzed during this study are included in this published paper (and its supplementary information files).

Authors' Contributions

Conceptualization involved contributions from CGL and PM. Methodology development was the responsibility of AB, PM, and CGL. All the authors participated in conducting the literature search and extracting data. AB, CGL, and PM were involved in preparing the original draft of the manuscript. Reviewing and editing the manuscript was a collaborative effort among all the authors. Throughout the research process, supervision was provided by CGL and PM. All the authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Eligibility criteria.
[DOCX File , 15 KB - publichealth_v10i1e49772_app1.docx ]

Multimedia Appendix 2
Search strings.
[DOCX File , 26 KB - publichealth_v10i1e49772_app2.docx ]

Multimedia Appendix 3
Classification and interpretation of burnout scores.
[DOCX File , 17 KB - publichealth_v10i1e49772_app3.docx ]

Multimedia Appendix 4
Predictors adopted in bivariate analyses and in multivariable regression analyses (other than physical activity).
[XLSX File (Microsoft Excel File), 44 KB - publichealth_v10i1e49772_app4.xlsx ]

Multimedia Appendix 5
Quality assessment.
[DOCX File , 18 KB - publichealth_v10i1e49772_app5.docx ]

Multimedia Appendix 6
PRISMA checklist.
[PDF File (Adobe PDF File), 60 KB - publichealth_v10i1e49772_app6.pdf ]
References


70. Trockel M, Bohman B, Leslie E, Hamidi MS, Dolezal D, Roberts L, et al. A brief instrument to assess both burnout and professional fulfillment in physicians: reliability and validity, including correlation with self-reported medical errors, in a sample of resident and practicing physicians. Acad Psychiatry 2018 Feb;42(1):11-24 [FREE Full text] [doi: 10.1007/s11606-017-0849-3]


**Abbreviations**

HCW: health care worker

IPAQ-SF: International Physical Activity Questionnaire–Short Form
MBI: Maslach Burnout Inventory
PEO: population, exposure, and outcome
PF: Professional Fulfillment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
WE: Work Exhaustion

Edited by A Mavragani; submitted 09.06.23; peer-reviewed by S Liu, E Lindsäter; comments to author 28.08.23; revised version received 18.09.23; accepted 12.12.23; published 18.03.24.

Please cite as:
Pierpaolo Mincarone, Antonella Bodini, Maria Rosaria Tumolo, Saverio Sabina, Riccardo Colella, Linda Mannini, Eugenio Sabato, Carlo Giacomo Leo. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 18.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Association Between Adherence to 24-Hour Movement Guidelines and Noncommunicable Disease Risk in Chinese Adults: Prospective Cohort Study

Siyi Huang1*, PhD; Yuxuan Gu2*, PhD; Shahmir H Ali3, PhD; Jingjing Xue4, PhD; Ronghua Zhang5*, MPH; Xu Wen1*, PhD

1Department of Sports Science, College of Education, Zhejiang University, Hangzhou, China
2Nanjing Normal University, Nanjing, China
3Department of Social and Behavioral Sciences, School of Global Public Health, New York University, New York, NY, United States
4School of Humanities, Beijing Dance Academy, Beijing, China
5Department of Nutrition and Food Safety, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
*these authors contributed equally

Corresponding Author:
Ronghua Zhang, MPH
Department of Nutrition and Food Safety
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Road
Binjiang District
Hangzhou
China
Phone: 86 13958109105
Email: rhzhang@cdc.zj.cn

Abstract

Background: The increasing annual global deaths are attributable to noncommunicable chronic diseases (NCDs). Adhering to healthy lifestyle behaviors is associated with lower NCD risk, particularly among individuals with ample movement, enough sleep, and reduced sedentariness. Nevertheless, there are only few prospective assessments on the association of interactions between daily activities with NCD prevention, while the associations between adhering to Canadian 24-Hour Movement Guidelines (24HGs) for adults and NCD risks are still unknown. Compared to the general population, obese and overweight populations are at a higher risk of developing NCDs. Currently, it is unclear whether the health benefits of adhering to 24HGs differ between the general population and the obese population.

Objective: This study explores prospective associations between adherence to 24HGs and NCD risks by weight status among overweight and obese adults in China.

Methods: This decadal study consists of 9227 adults aged 35 years and older without any major NCDs at enrolment in the China Health and Nutrition Survey (2004-2011) and followed up until 2015. The exposure of interest was the overall score of compliance with 24HGs measured by participants’ self-report, wherein 1 point was assigned for compliance to each component, resulting in an aggregated score ranging from 0 to 3. The primary outcome was the first occurrence of major NCDs (high blood pressure, stroke, diabetes, cancer, and acute myocardial infarction). Log-binomial regression models were used to evaluate the associations.

Results: Overall, 4315 males and 4912 females, with 25,175 person-years of follow-up, were included in our analyses. The average baseline age was 50.21 (SD 11.04) years. Among the overweight and obese groups, those adhering to 1 (risk ratio [RR] 0.37, 95% CI 0.19-0.74; P=0.04), 2 (RR 0.37, 95% CI 0.19-0.72; P=0.03), and 3 (RR 0.32, 95% CI 0.14-0.73; P=0.06) recommendations of 24HGs had a significantly lower NCD risk than those not adhering to any of the activity guidelines. Among the normal or underweight groups, those adhering to 1 (RR 0.49, 95% CI 0.27-0.96; P=0.03) and 3 (RR 0.40, 95% CI 0.17-0.94; P=0.03) components had a significantly lower NCD risk than those not adhering to any of the activity guidelines.

Conclusions: In this prospective study, we found that active adherence to recommendations of 24HGs was associated with lower risks of NCDs, especially among overweight and obese participants. Additionally, overweight and obese individuals who...
met at least 1 component of 24HGs were at a significantly lower risk for NCDs, but this protective effect was not found among individuals in the normal and underweight groups. Individuals with excess body weight who tend to be more susceptible to health risks may gain greater health benefits than the general population by adhering to the recommendations of 24HGs.

(JMIR Public Health Surveill 2024;10:e47517) doi: 10.2196/47517

KEYWORDS
chronic diseases; 24-hour movement guidelines; obesity; noncommunicable disease; sleep; risk; overweight

Introduction

In 2021, 71% of deaths worldwide (41 million) were attributable to noncommunicable chronic diseases (NCDs) [1]. Premature deaths due to NCDs were due to cardiovascular disease (17.9 million), cancer (9.3 million), respiratory diseases (4.1 million), and diabetes (1.5 million). Individuals with NCDs have a shorter life expectancy, struggle with disability, and experience a lower quality of life [2]. Several studies have shown that adhering to healthy lifestyle behaviors contributes to a reduced NCD risk. For example, 2 recent longitudinal studies [2,3] revealed that adhering to healthy lifestyle behaviors at midlife is associated with longer life expectancy and living free from NCDs, particularly if individuals practice at least 3 healthy lifestyle behaviors (ie, ample movement, enough sleep, reduced sedentari ness). Among modifiable lifestyle factors, physical activity, limited leisure screen time, and adequate sleep duration have been consistently shown to have a protective effect against NCDs [4-7]. A meta-analysis evaluating the association between the levels of physical activity and the incidence of breast cancer showed that the relative risk of developing breast cancer was only 0.88 among women with the highest level of daily physical activity compared to those with the lowest level of daily physical activity [4]. Regarding sleep, either short or long sleep duration have been correlated with an increased risk of cancer, cardiovascular disease, and diabetes [5-7]. The available evidence from the UK Biobank has suggested that daily screen viewing, particularly television viewing and computer use, is associated with a modest, but significant, increase in the risk of esophageal and colon cancers [8]. Although the association between NCD risk and physical activity, sleep traits, or leisure screen time has been well-documented, only few longitudinal investigations have assessed the onset of NCDs in relation to long-term patterns of daily movements.

The global population has more than 19 billion overweight and 650 million obese individuals as of 2016 [9]. Overweight and obese individuals are at a higher risk of NCDs, losing at least 3.9 years of life expectancy compared with the general population [10]. Additionally, obesity is linked to immunologic dysfunction, which is detrimental to both innate and adaptive immune responses [11,12]. As such, strategies to reduce overweight and obesity could have a positive impact on the health of billions of individuals at risk of NCDs.

There are daily movement patterns associated with obesity and NCDs. For instance, exercise may help prevent NCDs by controlling blood pressure, preventing impaired glucose tolerance, and improving C-reactive protein levels [13]. Although some studies have investigated the association between the time spent on daily movement and risks of NCDs in overweight and obese populations, it remains unclear as to what extent adhering to recommendations on daily movement facilitates long-term benefits in preventing NCDs. Moreover, it is currently unclear whether the health benefits of adhering to recommended behaviors differ between the general population and obese populations.

To address these research gaps, this study prospectively assesses the association between the reports of overweight and obese patients on their behavioral patterns and their risk of NCDs by using the data of Chinese adults enrolled in the China Health and Nutrition Survey (CHNS). As a metric of behavioral patterns, we used the Canadian 24-Hour Movement Guidelines (24HGs) for adults, which measures participant compliance to physical activity, screen time, and sleep duration.

Methods

Study Population

Data were collected from CHNS, which is an ongoing, longitudinal, nationwide survey initiated in 1989 among a sample of Chinese residents aged ≥22 years from 9 provinces, with follow-ups every 2-4 years. Structured questionnaires were used to obtain data, including demographic, lifestyle, and health factors, from study participants by trained staff. A physical examination was conducted to collect anthropometric data. Details on the full questionnaire and methodology of data collection have been described elsewhere [14]. The survey methodology and data collection procedures were approved by the institutional review board of the University of North Carolina and the National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention (2015017). A longitudinal cohort analysis of the 2004-2015 CHNS data with complete publicly available daily movement and health data was chosen because it was after 2004 that the CHNS was used to collect sedentary time data in front of the screen. All NCD-free adults (age≥35 years) with complete data on height, weight, physical activity, screen time, chronic disease data, and sleep duration in a survey year were considered eligible participants. Only 1 baseline timepoint was used for the exposure variables to maximize the sample sizes and time intervals between exposure and potential outcomes. The final analyses included 9227 eligible participants.

Measurement of Adherence to 24HGs

The exposure of interest was the overall scores of compliance with the Canadian 24HGs for adults (age 18-64 years) and older adults (65+ years), which were measured at baseline for all participants. Compliance was measured through adherence to 3 components of 24HGs: (1) adequate moderate to vigorous physical activity (MVPA) (≥150 min/week), (2) limited screen
time (≤3 hours/day), and (3) recommended sleep duration (7-9 hours for adults aged 18-64 years and 7-8 hours for those older than 65 years). Using these baseline 24HGs data, we calculated an aggregated adherence score for each participant, ranging from 0-3 depending on how many of the aforementioned 24HGs components were met. Participants’ average daily time on MVPA was calculated using the reported time they spent on physical activities, commuting, and housework per day. We used the Compendium of Energy Expenditures for Adults to describe the absolute intensity of physical activities and estimate the metabolic equivalents for different kinds of exercises. The compendium defines activities involving energy expenditure >3 metabolic equivalents as MVPA. Participants in this study were classified as meeting physical activity guidelines when they accumulated an average of at least 150 minutes of MVPA per week, and 1 point was assigned to them. For sleep duration, participants reported the whole time spent on bed, covering daytime and nighttime. The total hours of all-day sleeping were recorded on average per day. According to 24HGs, uninterrupted sleep of 7-9 hours (age 18-64 years) and 7-8 hours per night (age ≥65 years) is considered as meeting the sleep duration recommendation, and 1 point was assigned to them. To measure leisure screen time, participants reported the average time spent on different recreational screen activities on a typical weekend and on a weekday. Participants were classified within the “recommended screen time” group if they spent less than 3 hours/day both on weekends and weekdays.

**Measurement of Weights and BMI**

Weight and height were measured by trained staff following standardized protocols and performed at the same location as well as by following the same protocol at each survey visit. Weight was measured to the nearest 0.1 kg by using a calibrated beam scale while wearing lightweight clothing. Height was measured to the nearest 0.1 cm without wearing shoes by using a portable stadiometer. BMI was calculated as weight (kg) divided by the square of height (meter). The Working Group on Obesity in China proposed a threshold value for Chinese adults to judge the degree of overweight and obesity in 2003. This overweight and obesity metric is based on large-scale survey data of the Chinese population, wherein overweight was defined as BMI≥24 kg/m² and obesity was defined as BMI≥28 kg/m² [15].

**Assessment of Covariates**

Demographic characteristics consisted of age, sex (male/female), household register (rural/urban), region (northern, southern, western, eastern, and, central China), baseline year (2004/2006, 2009/2011), energy intake (kcal), regular intake of sugar-sweetened beverage (more than 3 times per week or not), regular intake of tea (more than 3 times per week or not), regular intake of coffee (more than 3 times per week or not), smoking status (yes/ever/never), drinking alcohol status (yes/ever/never), light physical activities, family net income (in yuan; ¥1= US $0.15), and highest education level (graduation from lower middle school degree or lower/upper middle school degree, technical or vocational degree/university or college degree or higher). According to the level of economic development and regional characteristics, we divided the regions into 5 groups: eastern region (Shanghai, Jiangsu, and Shandong), western region (Chongqing and Guizhou), southern region (Guangxi), northern region (Beijing, Liaoning, and Heilongjiang), and central region (Henan, Hubei, and Hunan). Family income was inflated to 2015 and adjusted for regional differences.

**Ascertainment of NCDs**

The primary outcome was the first occurrence of major NCDs, including high blood pressure, stroke, diabetes, cancer, and acute myocardial infarction. In each wave, participants were asked, “Have you ever been diagnosed with high blood pressure/stroke/diabetes/cancer/acute myocardial infarction?” Thus, in each survey wave, participants self-reported their health status, drawing from their recollection of hospital diagnoses and medical records. We obtained information pertaining to the presence of chronic diseases among individuals in the current wave. We considered participants to have NCDs as long as they had at least one of these NCDs.

**Ethics Approval**

Written informed consent was obtained from all the study participants, and the CHNS received institutional review board approval from the University of North Carolina (No.2015017). The CHNS data sets are openly accessible, and the data are freely available to scholars in the CHNS official websites after submitting applications. Our research was based on approved deidentified data, and this study was approved by the medical ethics committee of the Department of Psychological and Behavioral Sciences at Zhejiang University (approval 2022060).

**Patient and Public Involvement Statement**

Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our investigations. The CHNS data sets and published papers are available to the public online [16].

**Statistical Methods**

Baseline characteristics were described using frequencies (percentages) and mean (SD). The characteristic differences of the study population were compared across meeting 24HGs by using Poisson regression, log-binomial regression models were used, which offer better convergence and directly provides us with the risk ratio (RR). RRs and 95% CIs for NCDs associated with 24HGs were estimated with the use of log-binomial regression models. We evaluated the associations of adhering to recommendations with the incidence of NCDs by using log-binomial regression models with stepwise adjustment of confounding variables. Multivariate models were adjusted for age, age*age, and sex in model 1. Additionally, these were adjusted for the type of household register, province, survey year, smoking status, alcohol status, frequency of drinking sugar-sweetened beverages/tea/coffee, family net income, and light physical activity in model 2, and adjusted for the whole movement time in model 3. We imputed missing data by using multiple
imputations [17] for continuous variables and the mode for the missing data of categorical variables. Adding the square of the variable allows us to model more accurately the effect of age, which may have a nonlinear relationship with our independent variable [18]. We applied component analysis—the ratio of the time spent to the recommended time for each activity based on the Physical Activity CoDa Regression Model and code [19].

In the sensitivity analysis considering the influence of specific recommendations, we assessed the association between adhering to a specific combination of 24HGs and the onset of chronic diseases among overweight/obese participants. In addition, we repeated the analysis after excluding underweight individuals at baseline. We also conducted subgroup analyses by using negative binomial regression analysis to explore whether the corresponding associations varied by age (35-49 years/≥50 years), sex (male/female), education level (low/high), and family income (low/high). We also tested the moderation effect of adhering to guidelines and the group variables. Analyses were performed with the R statistical package (version 4.2.2) [20]. All $P$ values were 2-sided, with statistical significance set at <.05.

Results

Baseline Characteristics of the Study Population

Overall, 9227 participants (4315 males and 4912 females), with 25,175 person-years of follow-up, were included in our analyses, and the average baseline age was 50.21 (SD 11.04) years, with the majority of the participants being normal or underweight (5459/9227, 59.2%), living in rural environments (6124/9227, 66.4%), from Central or Eastern regions (4854/9227, 52.6%), and surveyed in 2004 and 2006 (6159/9227, 66.7%). As shown in Table 1, in terms of compliance with the guidelines, 112 (1.2%) participants did not meet any of the recommendations at baseline, and 2171 (23.5%), 6716 (72.8%), and 228 (2.5%) participants met 1, 2, and all components of the 24HGs, respectively. Compared to participants who met none, those who adhered to the overall recommendations of the guidelines were more likely to be older, live in urban environments, have lower energy intake, drink less coffee, never smoke, never drink alcohol, have a higher family net income, and have higher education.
Table 1. Characteristics of the participants by the category of adhering to the 24-hour movement guidelines at baseline.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall (N=9227)</th>
<th>Category of adhering to 24-hour movement guidelines</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Category</td>
<td>Measuring compliance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>none (n=112)</td>
<td>1 standard (n=2171)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 standards (n=6716)</td>
<td>3 standards (n=228)</td>
<td></td>
</tr>
<tr>
<td>Age at baseline (years), mean (SD)</td>
<td>50.21 (11.04)</td>
<td>49.01 (10.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>53.57 (12.51)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>49.11 (10.35)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>51.38 (9.79)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td>.054</td>
</tr>
<tr>
<td>Male</td>
<td>4315 (46.8)</td>
<td>62 (55.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1016 (46.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3146 (46.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>91 (39.9)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4912 (53.2)</td>
<td>50 (44.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1155 (53.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3570 (53.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>137 (60.1)</td>
<td></td>
</tr>
<tr>
<td>BMI category, n (%)</td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Normal weight</td>
<td>5072 (55)</td>
<td>62 (55.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1196 (55.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3695 (55)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>119 (52.2)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>3093 (33.5)</td>
<td>30 (26.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>697 (32.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2278 (33.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>88 (38.6)</td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>675 (7.3)</td>
<td>15 (13.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>164 (7.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>479 (7.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17 (7.5)</td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>387 (4.2)</td>
<td>5 (4.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>114 (5.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>264 (3.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 (1.8)</td>
<td></td>
</tr>
<tr>
<td>Household register, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Urban</td>
<td>3103 (33.6)</td>
<td>58 (51.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>772 (35.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2129 (31.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>144 (63.2)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>6124 (66.4)</td>
<td>54 (48.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1399 (64.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4587 (68.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>84 (36.8)</td>
<td></td>
</tr>
<tr>
<td>Regiona, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Northern</td>
<td>1950 (21.1)</td>
<td>29 (25.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>407 (18.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1445 (21.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>69 (30.3)</td>
<td></td>
</tr>
<tr>
<td>Eastern</td>
<td>2273 (24.6)</td>
<td>24 (21.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>529 (24.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1647 (24.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>73 (32)</td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>2581 (28)</td>
<td>39 (34.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>665 (30.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1837 (27.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>40 (17.5)</td>
<td></td>
</tr>
<tr>
<td>Southern</td>
<td>979 (10.6)</td>
<td>2 (1.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>192 (8.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>767 (11.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18 (7.9)</td>
<td></td>
</tr>
<tr>
<td>Western</td>
<td>1444 (15.6)</td>
<td>18 (16.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>378 (17.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1020 (15.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>28 (12.3)</td>
<td></td>
</tr>
<tr>
<td>Baseline year, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2004 and 2006</td>
<td>6159 (66.7)</td>
<td>68 (60.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1419 (65.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4561 (67.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>111 (48.7)</td>
<td></td>
</tr>
<tr>
<td>2009 and 2011</td>
<td>3068 (33.3)</td>
<td>44 (39.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>752 (34.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2155 (32.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>117 (51.3)</td>
<td></td>
</tr>
<tr>
<td>Total energy intake (kcal/day), mean (SD)</td>
<td>2136.22 (703.25)</td>
<td>2091.14 (657.13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2079.97 (663.95)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2160.60 (717.64)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1973.80 (603.33)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Regular intake of sugar-sweetened beverageb, n (%)</td>
<td></td>
<td></td>
<td>.31</td>
</tr>
<tr>
<td>No</td>
<td>6953 (75.4)</td>
<td>90 (80.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1636 (75.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5068 (75.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>159 (69.7)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2251 (24.4)</td>
<td>22 (19.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>531 (24.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1629 (24.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>69 (30.3)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>23 (0.2)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 (0.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>19 (0.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Regular intake of tea b, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>9006 (97.6)</td>
<td>104 (92.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2100 (96.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6586 (98.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>216 (94.7)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>202 (2.2)</td>
<td>8 (7.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>65 (3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>118 (1.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11 (4.8)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>19 (0.2)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 (0.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12 (0.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Regular intake of coffee b, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>3650 (96.8)</td>
<td>42 (93.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>819 (95.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2688 (97.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>101 (96.2)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>112 (3)</td>
<td>3 (6.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>41 (4.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>64 (2.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 (3.8)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>6 (0.2)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 (0.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 (0.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Smoking status, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>6159 (66.7)</td>
<td>61 (54.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1404 (64.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4520 (67.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>174 (76.3)</td>
<td></td>
</tr>
<tr>
<td>Yes/ever</td>
<td>3056 (33.1)</td>
<td>51 (45.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>765 (35.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2186 (32.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>54 (23.7)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>12 (0.1)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 (0.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 (0.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Association between the number of 24-hour movement guidelines met at baseline and the onset of chronic diseases in the overweight/obese group.

<table>
<thead>
<tr>
<th>Adherence</th>
<th>Participants, n</th>
<th>Events, n</th>
<th>Model 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Model 2&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Model 3&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk ratio (95% CI)</td>
<td>Risk ratio (95% CI)</td>
<td>Risk ratio (95% CI)</td>
</tr>
<tr>
<td></td>
<td>Risk ratio</td>
<td>Reference</td>
<td>P value</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>None</td>
<td>45</td>
<td>17</td>
<td>0.61 (0.44-0.94)</td>
<td>0.36 (0.19-0.73)</td>
<td>0.32 (0.14-0.73)</td>
</tr>
<tr>
<td>Meet 1 guideline</td>
<td>861</td>
<td>219</td>
<td>.01</td>
<td>.003</td>
<td>.006</td>
</tr>
<tr>
<td>Meet 2 guidelines</td>
<td>2757</td>
<td>659</td>
<td>0.09</td>
<td>.002</td>
<td>.003</td>
</tr>
<tr>
<td>Meet all guidelines</td>
<td>105</td>
<td>22</td>
<td>0.48 (0.29-0.81)</td>
<td>0.32 (0.14-0.73)</td>
<td>0.32 (0.14-0.73)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Adherence with participants not adhering to any recommendations of 24HGs, those adhering to 1 (RR 0.37, 95% CI 0.19-0.74; P=.004), 2 (RR 0.37, 95% CI 0.19-0.72; P=.003), and all (RR 0.32, 95% CI 0.14-0.73; P=.006) components had a significantly lower risk of NCDs (Table 2).

<sup>b</sup>In the normal weight or underweight group, compared with participants not adhering to any of the activity guidelines, those adhering to 1 (RR 0.49, 95% CI 0.27-0.96; P=.03) and overall components of the 24HGs (RR 0.40, 95% CI 0.17-0.94; P=.03) were associated with a lower risk of NCDs in the fully adjusted model. However, the association was not significant in the group adhering to 2 recommendations (Table 3).
Table 3. Association between the number of 24-hour movement guidelines met at baseline and the onset of chronic diseases in the normal/underweight group.

<table>
<thead>
<tr>
<th>Adherence</th>
<th>Participants, n</th>
<th>Events, n</th>
<th>Model 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Model 2&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Model 3&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risk ratio (95% CI)</td>
<td>P value</td>
<td>Risk ratio (95% CI)</td>
</tr>
<tr>
<td>None</td>
<td>67</td>
<td>15</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Meet 1 guideline</td>
<td>1310</td>
<td>230</td>
<td>0.55 (0.31-1.05)</td>
<td>.06</td>
<td>0.54 (0.30-1.04)</td>
</tr>
<tr>
<td>Meet 2 guidelines</td>
<td>3959</td>
<td>614</td>
<td>0.65 (0.37-1.23)</td>
<td>.16</td>
<td>0.63 (0.35-1.20)</td>
</tr>
<tr>
<td>Meet all guidelines</td>
<td>123</td>
<td>13</td>
<td>0.39 (0.17-0.89)</td>
<td>.03</td>
<td>0.43 (0.18-1.01)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Age*age and sex were adjusted; per score increase = 1.03 (95% CI 0.89-1.20); P=.68.<br><sup>b</sup>The type of household register, province, survey year, smoking status, alcohol status, frequency of drinking sugar-sweetened beverage/tea/coffee, family net income, and light physical activity were additionally adjusted; per score increase = 1.03 (95% CI 0.89-1.20); P=.66.<br><sup>c</sup>The whole movement time was additionally adjusted; per score increase = 1.01 (95% CI 0.87-1.18); P=.90.

Additionally, we performed a compositional analysis with proportions of ratios of the actual time spent on each kind of activity and recommended time (the constituent ratios of 3 movements for each participant) to compare the results more comprehensively. Figure 1 shows the proportions of sleep, MVPA, and screen time, which are displayed using dotted blue and red lines, surrounded by 3 vertices in the heatmap. We observed a positive association between the proportion of screen time and risk of NCDs and a negative association between the proportions of MVPA the risk of NCDs.

Figure 1. Heatmap ternary diagrams of expected risk ratios based on model 3 against different percentage time allocation of the movement behavior time-use composition in the (A) normal/underweight group and (B) overweight/obese group. MVPA: moderate to vigorous physical activity.

**Stratified Analyses and Sensitivity Analyses**

For the specific combination of 24HGs, after fully adjusting for all covariates, we also observed an inverse association between adhering to 24HGs and risk for NCDs across specific combinations of the guidelines in the overweight and obese group. Compared with participants not adhering to any of the activity guidelines, those adhering to meeting screen time guideline (RR 0.34, 95% CI 0.18-0.69; P=.002), meeting screen time + MVPA guideline (RR 0.19, 95% CI 0.06-0.54; P=.002), meeting screen time + sleep guideline (RR 0.37, 95% CI 0.19-0.72; P=.003), and meeting all (RR 0.32, 95% CI 0.14-0.72; P=.006) had a significantly lower risk of NCDs (Table S1 of Multimedia Appendix 1). Moreover, in the subgroup analysis, we observed a similar inverse association between adhering to 24HGs and risk for overweight and obesity across age, sex, highest education level, and family net income subgroups in the fully adjusted model. We did not observe a significant interaction effect between these groups and adherence to 24HGs (Figure 2, Table S2 and Table S3 of Multimedia Appendix 1).
Figure 2. Subgroup analysis of the association between the number of 24-hour movement guidelines met at baseline and the onset of chronic diseases in the (A) normal/underweight group and (B) overweight/obese group.

Discussion

Principal Findings

In this prospective cohort study, we found that compliance with 24HGs had a protective effect against major chronic diseases such as high blood pressure, stroke, diabetes, cancer, and acute myocardial infarction. These associations were independent of the potential confounding effects of age, sex, and socioeconomic status. Compared with individuals who met none of the recommendations, overweight or obese individuals who met 1, 2, or 3 recommendations had a 63%, 63%, and 68% lower risk of being diagnosed with NCDs, respectively. Restricted leisure screen time was observed to have a strong protective effect on chronic disease prevention. Among normal weight or underweight individuals, adhering to 1 or all components of 24HGs were linked with 51% or 60% lower risk of NCDs, respectively.

Although previous studies have performed several investigations regarding association between lifestyle behavior and life expectancy without chronic diseases [21,22], only few studies have assessed the compliance to 24HGs. Findings from a prospective multicohort study of 116,043 European adults aged 40-75 years showed that individuals who followed all the healthy lifestyle behaviors gained about 9 additional disease-free years of life compared to those who did not adopt any healthy behavior [23]. Nevertheless, the Nurses’ Health Study and Health Professionals Follow-Up Study did not observe significant improvements in life expectancy upon adherence to more risk-reducing lifestyle behaviors [2]. This inconsistency in past evidence may be due to the inconclusive combination of healthy lifestyle behavior assessments [24,25]. In our study, 24-hour daily movements were considered an essential part of a healthy lifestyle, and the rationale for recommending 24HGs is already supported by evidence from past randomized controlled trials [26]. Furthermore, independent associations among adequate physical activity [4], limited leisure screen time, sleep duration [6,27], and NCD prevention have also been observed in previous observational and experimental investigations. However, previous studies have not been able to comprehensively assess the association between daily movements and NCDs, and only very few studies have evaluated the association between daily movements and the onset of major NCDs [28]. Our prospective
The association between obesity and the onset of NCDs has potential biological significance and noticeable health improvements. The association of obesity with COVID-19 hospitalization [33] is due to the vulnerability of the overweight/obese group, which are at greater risk due to chronic diseases [10], and face greater risks of NCDs [28], experience more substantial declines in quality of life due to chronic diseases [3].

Despite following the same recommendations of 24HGs, the normal/underweight and overweight/obese groups experienced different health benefits, as indicated by the point estimates. This divergence may be attributable to the particular vulnerabilities of the overweight/obese group, which are supported by well-established biological mechanisms. In comparison to the general population, obese individuals are at an increased risk. They are more susceptible to environmental risk factors for health, have a higher prevalence of mental health disorders [28], experience more substantial declines in quality of life due to chronic diseases [10], and face greater risks of COVID-19 hospitalization [33]. Due to the vulnerability of the obese population, even minor changes could lead to more significant and noticeable health improvements. The association between obesity and the onset of NCDs has potential biological mechanisms. A recent review [34] reported that groups with obesity and adiposity had long-term higher leptin or lower adiponectin concentrations, which might cause dysregulation of the immune system response that leads to abnormally higher concentrations of proinflammatory cytokines. Obesity is usually a situation of chronic low-grade systemic inflammation that leads to numerical and functional alterations of lymphocytes [35]. In the basal state, B-cell and T-cell responses are impaired in individuals with excess body weight, and these alterations may upsurge vulnerability to viral infection [35]. In addition, because several inflammatory components inhabit the tumor microenvironment and foster the phenotype of cancer, adipose inflammation is also considered to be a mediator of cancer development.

In addition, the higher level of concentrations of adipocytokines are associated with fluctuations in the levels of angiotensinogen, plasminogen activator inhibitor-1, interleukin-6, and resistin, which might progress the onset of insulin resistance and atherosclerotic lesions [35,36]. Compared with general population, individuals with adipose inflammation and metabolic dysfunction exhibit higher cancer and cardiovascular disease risk [35,36]. For individuals who are overweight or obese, setting achievable goals and performing daily movements to intentionally lose weight not only reduce the levels of chronic disease-related factors but also reduce the risks of comorbidity incidence and disease progression. Adhering to physical activity recommendations may induce the beneficial effects of decrease in the levels of inflammatory biomarkers and improved endothelial function in obese children and adolescents [36].

Based on the analysis results of 59,005 adults in an English and Scottish Health Survey, vigorous-intensity leisure-time physical activity showed a significant effect on decreasing the cardiovascular disease mortality risk in the obese group [37]. Compared with active females with excess body weight, inactive females were exposed to 20%-41% higher risk of the onset of type 2 diabetes, while inactive males were exposed to 36%-38% higher risk than active overweight and obese males [38]. In addition, we found that meeting MVPA and leisure screen use recommendations would decline 81% risk of chronic diseases. Besides the salutary benefits conferred by physical activity, our analysis revealed the concurrent reduction in the risk of NCD onset when leisure screen time is curtailed, particularly when combined with adequate sleep duration. The nonsignificant values observed in the results for the association between MVPA and NCD risk may be attributable to the smaller sample size rather than a lack of association between MVPA and the prevention of NCDs. A smaller sample size can diminish the statistical power of the study, potentially leading to a type II error, where a true effect is present but not detected. Thus, the observed significance value should not be interpreted as definitive evidence of no relationship between MVPA and NCD prevention. Further research with a larger sample size is warranted to accurately determine the extent of this association.

Future research is needed to investigate the mechanisms that link adhering to 24HGs with NCD risk as well as to explore effective daily movement interventions needed for substantial NCD risk reduction.

**Strengths and Limitations**

One strength of our study was that our analysis was based on a large nationwide sample size of adults with long-term follow-up.
data, which provided high statistical power and the ability to infer temporality. Another important strength was that the data collection of CHNS was conducted under robust quality control management and performed by qualified staff.

There are, however, several limitations to our study. First, the data on lifestyle behaviors were collected via self-reported questionnaires, which might be subject to misclassification or self-report biases. Nonetheless, the results of our study were largely consistent with past research [2,23-28], supporting the validity of our findings. Second, there are likely various potential unmeasured factors related to daily activities or NCD status that could not be adjusted for in the analyses. Third, in the CHNS, sleep duration was collected through self-reported questionnaires. Therefore, we were unable to assess participants’ sleep quality, as sleep quality is typically measured by electroencephalography monitoring, which would not have been feasible in a population-based survey, given its complexity and cost. Thus, in our study, for meeting the guidelines, we were primarily concerned with the duration rather than quality. Fourth, the outcome data on NCDs were collected through self-reported questionnaires. Although participants reported health conditions based on their recollection of hospital diagnoses and medical records, the staff did not have access to the exact diagnosis dates. As a result, the ascertainment of NCDs relied on self-reported data, which may have introduced recall biases. Finally, as we enrolled only Chinese adults in our cohort study, our estimates may not be generalizable to other population subgroups with different social structures and characteristics.

Conclusion
Based on the CHNS data from the 2004 to 2015 waves, our study shows that actively following 24HGs was associated with lower risks of NCDs, especially among overweight or obese individuals. Furthermore, overweight and obese individuals who met at least 1 component of 24HGs were at a significantly lower risk of NCDs. However, this protective effect was not observed in normal or underweight groups. Individuals with excess body weight who tend to be prone to greater health risks may gain larger health benefits by adhering to these recommendations compared to the general population.

Acknowledgments
This investigation received financial support from Zhejiang Provincial Natural Science Foundation of China Funding (LY18C110001) and Fundamental Research Funds for the Central Universities and institutional support from the Department of Sport and Exercise Science, College of Education, Zhejiang University. We express our gratitude to all the staff and participants in the China Health and Nutrition Survey (CHNS) whose efforts made this study possible and thank CHNS for making their large data set available to us. We greatly appreciate guidance from Professor Huijie Qiao in R language, Associate Professor Luis E Escobar in manuscript writing, and Professor Huakang Tu in statistics. This work was supported by the Zhejiang Provincial Natural Science Foundation of China Funding (grant LY18C110001) and Fundamental Research Funds for the Central Universities. The funders had no role in study design, data analysis, data interpretation, writing of this report, or in the decision to submit this paper for publication.

Data Availability
The data sets generated and analyzed during this study were sourced from the China Health and Nutrition Survey. Researchers can access the data on the official China Health and Nutrition Survey website [39] after submitting an application. The data sets used in this study are also available from the corresponding author upon reasonable request.

Authors’ Contributions
SH and YG are credited as co-first authors for their significant contributions. XW (wenxu@zju.edu.cn) and RZ (rhzhang@cdc.zj.cn) are recognized as co-corresponding authors. They supervised the manuscript's completion process, including task allocation, manuscript writing, and budget management. Both have dedicated an equal amount of time and effort to the writing of this article and to overseeing the project that funded the article processing charges. SH and YG jointly worked on the study's concept and design. SH also took on the analysis and interpretation of data, as well as its organization into tables. The drafting of the manuscript was a joint effort by SH and YG, with guidance from XW and RZ, and with additional support from SHA and JX. YG, SH, SHA, and JX concentrated on assessing the results, especially focusing on the methods and analytics. All authors participated in critically revising the manuscript for important intellectual content, had full access to all the study's data, reviewed and revised the manuscript drafts, and gave their final approval for the version to be published.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File, 25 KB - publichealth_v10i1e47517_app1.docx ]
References


Abbreviations

24HGs: Canadian 24-Hour Movement Guidelines
CHNS: China Health and Nutrition Survey
MVPA: moderate to vigorous physical activity
NCD: noncommunicable disease
RR: risk ratio
Evaluating the Knowledge of and Behavior Toward COVID-19 and the Possibility of Isolating at a City Level: Survey Study

Elise Verot1,2,3, MSc, PhD; Robin Chaux4, MD; Julie Gagnaire2,5,6,7, MD, PharmD, PhD; Paul Bonjean4, MD; Amandine Gagneux-Brunon2,6,7,8, MD, PhD; Philippe Berthelot5,6,7,8,9, MD, PhD; Carole Pelissier10,11, MPH, PhD; Billal Boulamail4, MD; Franck Chauvin1,2,3, MD, PhD; Bruno Pozzetto7,9, MD, PhD; Elisabeth Botelho-Nevers2,6,7, PhD

1CIC EC 1408 INSERM Saint-Etienne, Saint-Etienne cedex 2, France
2Laboratoire Parcours Sante’ Système-UR4129, Université Jean Monnet, Université de Lyon, St Priest-en-Jarez, France
3Chaire Hygiène, Institut PRESAGE, Université Jean Monnet, Université de Lyon, Saint-Etienne, France
4Unité de Recherche Clinique, Centre Hospitalier Universitaire de Saint-Etienne, Saint-Etienne, France
5Unité de Gestion des Risques infectieux, Centre Hospitalier Universitaire (CHU) de Saint-Etienne, Saint-Etienne, France
6Service d’Infectiologie, Centre Hospitalier Universitaire de Saint-Etienne, Saint-Etienne, France
7Team GIMAP, CIRI—Centre International de Recherche en Infectiologie, Université Jean Monnet, Université de Lyon, Saint-Etienne, France
8Chaire PreVacCI, Institut PRESAGE, Université Jean Monnet, Université de Lyon, Saint-Etienne, France
9Laboratoire des Agents Infectieux et d’Hygienne, Centre Hospitalier Universitaire de Saint-Etienne, Saint-Etienne, France
10Occupational Health Service University Hospital Center of Saint-Etienne, Saint-Etienne, France
11UMRESTTE, Université Lyon 1, Université Gustave Eiffel—IFSTTAR, UMR t 9405, Lyon, France

Corresponding Author:
Elise Verot, MSc, PhD
CIC EC 1408 INSERM Saint-Etienne
CHU de Saint-Etienne
Bâtiment le Canopée- 2ème étage
Saint-Etienne cedex 2, 42055
France
Phone: 33 04 77 12 77 88
Email: elise.verot@univ-st-etienne.fr

Abstract

Background: Mass testing campaigns were proposed in France during the first wave of the COVID-19 pandemic to detect and isolate asymptomatic individuals infected by SARS-CoV-2. During mass testing in Saint-Étienne (February 2021), we performed a survey of the general population.

Objective: We evaluated, on the scale of a city’s population, the literacy level about SARS-CoV-2 transmission, barrier gesture respect, and isolation acceptability or possibility in case of SARS-CoV-2 infection.

Methods: We used the validated CovQuest-CC questionnaire. Data were analyzed and correlated with volunteer characteristics and their SARS-CoV-2 screening results using multivariate analysis.

Results: In total, 4707 participants completed the CovQuest-CC questionnaire. Multivariate analysis revealed that female sex was a determinant of a higher score of knowledge about SARS-CoV-2 transmission (adjusted β coefficient=0.14, 95% CI 0.04-0.23; corrected P=.02). Older ages of 50-59 years (adjusted β coefficient=0.25, 95% CI 0.19-0.31; corrected P<.001) and ≥60 years (adjusted β coefficient=0.25, 95% CI 0.15-0.34; corrected P<.001) were determinants of a higher score on barrier gesture respect compared to ages 20-49 years considered as reference. Female sex was also a determinant of a higher score on barrier gesture respect (adjusted β coefficient=0.10, 95% CI 0.02-4.63; corrected P<.001). The knowledge score was correlated with the score on barrier gesture respect measures (adjusted β coefficient=0.03, 95% CI 0.001-0.004; corrected P=.01). Older ages of 50-59 years (adjusted β coefficient=0.21, 95% CI 0.13-0.29; corrected P<.001) and ≥60 years (adjusted β coefficient=0.25, 95% CI 0.1-0.38; corrected P<.001) were determinants of a higher score on isolation acceptability or possibility compared to the age of 20-49 years considered as reference. Finally, the knowledge score regarding SARS-CoV-2 transmission was significantly associated with a lower risk of RT-PCR (reverse transcriptase–polymerase chain reaction) positivity (adjusted odds ratio 0.80, 95% CI 0.69-0.94; corrected P<.03), implying that a 1-point increase in the knowledge score lowers the risk of positivity by 20% on average.
Conclusions: This study identified factors associated with health literacy regarding SARS-CoV-2 infection in asymptomatic individuals in a large French city’s population. We can confirm that in the context of the COVID-19 pandemic, the determinants of better health literacy are not the same as in other contexts. It seems critical to obtain a more detailed understanding of the determinants of individual citizens’ behavior, as part of a strategy to combat the large-scale spread of the virus. The harsh experience of this pandemic should teach us how to nurture research to structure customized interventions to encourage the adoption of ad hoc behaviors to engage citizens in adapting behaviors more favorable to their health.

KEYWORDS
SARS-CoV-2; COVID-19; health literacy; knowledge, attitude, and perception/practices (KAP); public health; population; mass testing; screening; pandemic; sociological trends; COVID-19 screening

Introduction
Since the beginning of the COVID-19 crisis, and independent of the wave of the pandemic, one of the main goals of scientists and governments has been to reduce the burden of SARS-CoV-2 infection [1].

Mass Testing Campaign to Contain the Spread of SARS-CoV-2
Among the different strategies used worldwide, mass testing was proposed early [2], with the aim to detect highly infectious individuals who are asymptomatic or presymptomatic, allowing their isolation together with the rapid identification and testing of their close contacts to reduce virus spread [3]. The performance of this strategy is influenced by the prevalence of active infection in the group being tested. Mass testing was then carried out in high-incidence settings with the hope of mitigating the transmission dynamics and avoiding lockdown measures [2]. The impact of such mass testing was considered controversial [2]. Indeed, although the “test, trace, and isolate” strategy has proven effective in suppressing early spread of SARS-CoV-2 [4,5], this strategy has also shown serious limitations as it has been overwhelmed by the increasing number of cases [5] and associated costs [6]. However, if these interventions were used notably before the availability of safe and effective COVID-19 vaccines, with the emergence of variants of concern exhibiting a high level of transmissibility, testing and isolation-based strategies are likely to remain viable tools for the control of epidemic waves.

The Citizen as a Player in the Fight Against the Spread of the Virus
Beyond mass nonpharmaceutical interventions conducted by cities or governments [7], each citizen has also a role, and may be an actor, in controlling the epidemics [8-11]. Indeed, a good understanding and knowledge of the transmission route of SARS-CoV-2, adhesion to barrier gesture respect measures, and having realistic possibilities of individual isolation contribute significantly to fight against the pandemic [10-12]. Moreover, after the launch of mass testing or even after a negative test result is obtained, efforts should continue to further protect oneself and others [13]. As West et al [14] pointed out in May 2020, effective interventions are urgently needed to increase adherence of the general population to the proper implementation of health measures for protecting people individually and collectively.

The Need to Understand Individual Behavior to Tackle the Crisis More Effectively
As Rodon et al [15] point out, assessing health literacy about COVID-19 is crucial to understanding any difficulties individuals may have in adopting protective measures and implementing social distancing measures, bearing in mind that age, education, and the way in which they gather information about COVID-19 are all factors that determine individual behavior. As part of the French health policy in 2021, the city of Saint-Étienne, France (174,082 residents) [16] took part in a French national experiment of city-wide mass testing operations for COVID-19. The campaign took place over 2 full weeks separated by a 5-week interval; the first one occurred from January 13 to 19, markedly after the end-of-year holidays, and the second one occurred from February 22 to 28, when the winter school vacations were over. The city of Saint-Étienne was chosen to be part of this experiment of mass testing since it had high incidence rates, peaking at 1000 cases per 100,000 inhabitants in late 2020 [17,18]. At the time, having been amid this unprecedented crisis, we seized the opportunity of this mass testing operation to survey people about their attitudes and beliefs regarding SARS-CoV-2, which was not undertaken on a city-wide scale. Hence, to our knowledge, this type of evaluation in such a mass testing context among asymptomatic citizens was not performed. While collecting these data may seem outdated at present, given that COVID-19 is no longer a public health emergency of international concern, this large-scale work carried out in a particular situation has enabled us to assess knowledge of, behaviors toward, and opportunities for COVID-19 isolation to better understand our fellow citizens in a pandemic context. At that time, the results of this approach were intended to help develop public health strategies to contain local epidemics using a population approach.

Methods
The study took place during the second week of the mass testing campaign of Saint-Étienne, France, that is, from February 22 to 28, 2021.

Background Concerning the Mass Testing Campaign in Saint-Étienne
Adults and children older than 10 years were eligible for inclusion. People wishing to participate in the study had to be older than 10 years and able to read and understand the French language. For minors, parental permission was required. An
information note was provided to each participant. After obtaining verbal consent of the participant or of his or her guardian, the participant was included in the study. For SARS-CoV-2 screening, saliva samples were obtained from the participants in accordance with the French guidelines of February 2021 on the use of salivary RT-PCR (reverse transcriptase–polymerase chain reaction) tests in the context of large-scale iterative screening in a closed population. Samples were collected and analyzed using quantitative RT-PCR in accordance with the methods described previously [18].

**Recruitment**

Individuals who voluntarily participated in the COVID-19 city-scale mass screening campaign were also offered to complete the CovQuest-CC questionnaire. Screenings were offered free of charge at 12 ephemeral sites, and in parallel, mobile teams were deployed to target populations (adolescents, students, people living in low-income neighborhoods, businesses, etc.). Study participants were provided a paper questionnaire with a pen and asked to complete it while they waited to be tested for the presence or absence of SARS-CoV-2. Everyone then handed in the questionnaire at the time of the screening test. The self-administered CovQuest-CC questionnaire explores 3 main areas: knowledge of SARS-CoV-2, adherence to barrier gesture respect, and the ability to isolate themselves in the event of a SARS-CoV-2 infection. For these 3 concept areas, all response modalities are offered in the form of a Likert scale. The precise content of the CovQuest-CC questionnaire is presented elsewhere [19]. The questionnaire was psychometrically tested and validated in a general population aged 10 years and older. The following age groups were represented: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, and >90 years. As explained in a previous report on the CovQuest-CC questionnaire validation process [19], we first carried out a pretest with a representative sample of the target population, to assess the comprehensibility of the questions and their wording. [19] The psychometric validation procedure showed this questionnaire was valid, consistent and reliable. Therefore, our study participants were those who both completed the CovQuest-CC questionnaire and were screened for SARS-CoV-2 infection in the context of the mass screening campaign.

**Information Collected From Participants**

Participants provided information on demographic characteristics, occupation, place of living, and symptoms through a paper questionnaire. The CovQuest-CC questionnaire provided data through a global knowledge score on SARS-CoV-2 transmission ranging from 0 to 6, a score on barrier gesture respect ranging from 0 to 4, and a score on isolation acceptability or possibility ranging from 0 to 4. The highest score for each item corresponded to the best level of response. The French version of the European Deprivation Index (EDI) was also used in this study to approximate the social deprivation level of participants [20]. This score was calculated in 2007 using homogenous clusters for statistical information (called “IRIS” [regrouped statistical information blocks], originally constructed by the organization in charge of the French census) [21]. EDI quintiles were used to group IRIS into 5 deprivation levels [21]. In this study, participants were associated with a deprivation level (increasing from 1 to 5, with 5 being the most deprived) according to the IRIS area they lived in.

**Ethical Considerations**

The CROSS (Checklist for Reporting of Survey Studies) guided the reporting of this survey study [22]. No financial compensation was provided to participants. This study was conducted in accordance with the guidelines set out in the Declaration of Helsinki. The research in which this validation took place was approved by the Ethics Committee of IRB ILE-DE-FRANCE I (I ORG0009918; Protocole No. EudraCT: 2021-A00390-41) and all participants provided their written consent. This study is registered on ClinicalTrials.gov (NCT04859023).

**Statistical Analysis**

**Power**

The margin of error for our sample of 4707 individuals is 1.45% and corresponds to what is expected for survey studies [23], with a recommended margin of error of 1%-10%.

Sample characteristics were described using frequencies and proportions for categorical variables, and mean, SD, and median and IQR for numerical variables.

The relationship between CovQuest-CC questionnaire scores and variables of interest were analyzes using multivariate linear regression models. The models were constructed a priori, without knowledge of the data. One model was developed to analyze the associations between the knowledge score and age, sex, profession, and EDI quintiles. One model was developed to analyze the associations between the barrier gesture respect score and age, sex, profession, EDI quintiles, and the knowledge score. One model was developed to analyze the associations between the isolation possibility score and age, sex, profession, EDI quintiles, and the number of children at home (one or more versus none). The results are presented as multivariate linear $\beta$ coefficients with their 95% CI. Colinearity between variables was systematically assessed using correlation matrices and calculation of the variance inflation factor. Models’ validity and robustness were systematically assessed via graphical residual analysis.

Relations between RT-PCR positivity and age, sex, occupation, global SARS-CoV-2 transmission knowledge score, barrier gesture respect score, number of children at home, and EDI were assessed using a multivariate logistic regression model. The results are presented as odds ratios with their 95% CI. To account for multiple testing, Bonferroni correction was used for adjusting the $P$ value to the number of variables tested in each model. All statistical tests were 2-sided, with $P<.05$ considered significant. Statistical analysis was performed with the R software (version 4.0.3; The R Foundation).

**Data Exclusion**

Interviewees with at least 1 missing value in one of the variables of the different models were excluded from the said model.
Results

User Statistics
Among the 7020 participants in the mass testing campaign, 4707 (67%) responded to the CovQuest-CC questionnaire and were therefore included in this study. The participants’ mean age was 50.3 (SD 18.54) years and their median age was 52 (IQR 36-66) years. Women represented 56.1% (n=2634) of the sample. Among participants, 72.2% (n=3399) lived in the city of Saint-Étienne, and 97.0% (n=4566) of them lived in the Loire or the adjacent Haute-Loire departments; 65.3% (n=2485) of participants lived in areas belonging to the 2 most disadvantaged quintiles of EDI (EDI 4 and EDI 5). Table 1 presents an overview of the sample characteristics.
Table 1. Characteristics of study participants (N=4707).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>10-19</td>
<td>272 (5.78)</td>
</tr>
<tr>
<td>20-49</td>
<td>1884 (40.06)</td>
</tr>
<tr>
<td>50-59</td>
<td>810 (17.22)</td>
</tr>
<tr>
<td>≥60</td>
<td>1737 (36.93)</td>
</tr>
<tr>
<td>Missing data</td>
<td>4 (0.09)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2634 (56.14)</td>
</tr>
<tr>
<td>Missing data</td>
<td>15 (0.32)</td>
</tr>
<tr>
<td><strong>Number of children in the household, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>3042 (67.42)</td>
</tr>
<tr>
<td>1 or more</td>
<td>1470 (32.58)</td>
</tr>
<tr>
<td>Missing data</td>
<td>195 (4.14)</td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Employees</td>
<td>1820 (38.91)</td>
</tr>
<tr>
<td>Students</td>
<td>370 (7.92)</td>
</tr>
<tr>
<td>High school or college students</td>
<td>152 (3.25)</td>
</tr>
<tr>
<td>Retired</td>
<td>1474 (31.54)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>234 (5.01)</td>
</tr>
<tr>
<td>Health workers</td>
<td>228 (4.88)</td>
</tr>
<tr>
<td>Self-employed workers</td>
<td>142 (3.04)</td>
</tr>
<tr>
<td>Others</td>
<td>254 (5.43)</td>
</tr>
<tr>
<td>Missing data</td>
<td>33 (0.7)</td>
</tr>
<tr>
<td><strong>EDI(^a) quintile, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>First quintile</td>
<td>807 (21.21)</td>
</tr>
<tr>
<td>Second quintile</td>
<td>396 (10.41)</td>
</tr>
<tr>
<td>Third quintile</td>
<td>117 (3.07)</td>
</tr>
<tr>
<td>Fourth quintile</td>
<td>374 (9.83)</td>
</tr>
<tr>
<td>Fifth quintile</td>
<td>2111 (55.48)</td>
</tr>
<tr>
<td>Missing data</td>
<td>902 (19.16)</td>
</tr>
<tr>
<td><strong>SARS-CoV-2 RT-PCR(^b) result, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>3906 (98.14)</td>
</tr>
<tr>
<td>Positive</td>
<td>74 (1.86)</td>
</tr>
<tr>
<td>Missing data</td>
<td>727 (15.45)</td>
</tr>
<tr>
<td>Knowledge score(^c), median (IQR)</td>
<td>5 (4.5-6)</td>
</tr>
<tr>
<td>Barrier gesture respect score(^d), median (IQR)</td>
<td>3 (2.50-3.50)</td>
</tr>
<tr>
<td>Isolation respect score(^e), median (IQR)</td>
<td>2.75 (2.25-3.33)</td>
</tr>
</tbody>
</table>

\(^a\)EDI: The French version of the European Deprivation Index.

\(^b\)RT-PCR: reverse transcriptase–polymerase chain reaction.

\(^c\)Data missing for 204 (4.33%) participants.

\(^d\)Data missing for 266 (5.65%) participants.

\(^e\)Data missing for 389 (8.26%) participants.
Evaluation Outcomes

Knowledge Score on the Transmission of SARS-CoV-2 According to the CovQuest-CC Questionnaire

The median score was 5 (IQR 4.5-6). Factors significantly associated with the knowledge score in multivariate analysis are depicted in Figure 1 and Multimedia Appendix 1. Female sex was shown as a determinant of a higher score for knowledge about transmission of SARS-CoV-2 (adjusted $\beta$ coefficient=0.14, 95% CI 0.04-0.23; corrected $P$=.02). Compared to health workers taken as reference, middle and high school students exhibited a lower score of knowledge about transmission of SARS-CoV-2 (adjusted $\beta$ coefficient=–0.57, 95% CI –1.01 to –0.13; corrected $P$=.04).

Figure 1. Adjusted multivariate linear regression exploring the associations between the knowledge global score and the variables of interest, with Bonferroni corrected $P$.

Barrier Gesture Respect Measure Scores According to the CovQuest-CC Questionnaire

The median score was 3 (IQR 2.5-3.5). Factors significantly associated with this score in multivariate analysis are depicted in Figure 2 and Multimedia Appendix 2. Older ages of 50-59 years (adjusted $\beta$ coefficient=0.25, 95% CI 0.19-0.31; corrected $P$<.001) and $\geq$60 years (adjusted $\beta$ coefficient=0.25, 95% CI 0.15-0.34; corrected $P$<.001) were shown as determinants of a higher score on barrier gesture respect compared to ages 20-49 years considered as reference. Female sex also appeared as a determinant of a higher score on barrier gesture respect (adjusted $\beta$ coefficient=0.10, 95% CI 0.02-0.63; corrected $P$<.001). Finally, the level of the knowledge score was also correlated to that of the score for barrier gesture respect measures (adjusted $\beta$ coefficient=0.03, 95% CI 0.001-0.004; corrected $P$=.001). By contrast, higher education was significantly associated with a lower barrier gesture respect score (adjusted $\beta$ coefficient=–0.30, 95% CI –0.43 to –0.17; corrected $P$<.001).
Figure 2. Adjusted multivariate linear regression exploring the associations between the barrier gesture respect score and the variables of interest, with Bonferroni corrected P.

Score for Isolation Acceptability or Possibility According to the CovQuest-CC Questionnaire

The median score was 2.75 (IQR 2.25-3.33). Factors significantly associated with this score in multivariate analysis are depicted in Figure 3 and Multimedia Appendix 3. Older ages of 50-59 years (adjusted $\beta$ coefficient=0.21, 95% CI 0.13-0.29; corrected $P<.001$) and ≥60 years (adjusted $\beta$ coefficient=0.25, 95% CI 0.1-0.38; corrected $P<.001$) were shown to be determinants of a higher score on isolation acceptability or possibility compared to ages 20-49 years considered as reference. In contrast, having at least 1 child living at home versus having none exhibited a lower score on isolation possibility (adjusted $\beta$ coefficient=–0.12, 95% CI –0.19 to –0.05; corrected $P=.006$).

Figure 3. Adjusted multivariate linear regression exploring the associations between the isolation possibility score and the variables of interest, with Bonferroni corrected P.

RT-PCR Positivity Determinants

Results of univariate and multivariate analyses of RT-PCR positivity determinants are presented in Figure 4 and Multimedia Appendix 4. Interestingly, the knowledge score for SARS-CoV-2 transmission was significantly associated with a lower risk of RT-PCR positivity (adjusted odds ratio 0.80, 95% CI 0.69-0.94; corrected $P<.03$), implying that a 1-point increase in knowledge score lowers the risk of positivity by 20% on average.
Discussion

Background

In the middle of a city-wide mass screening campaign, beyond proposing massive SARS-CoV-2 screening for asymptomatic citizens, as accomplished in several other cities [24,25], we were able to assess the knowledge of, behavior toward, and potential for isolation of participants through a validated questionnaire on SARS-CoV-2–related health literacy [19]. It therefore provided a global picture of the determinants of the population’s adherence to a national control strategy in the context of a pandemic at the level of a French semiurban agglomeration of 174,082 inhabitants.

Indeed, the sanitary crisis that has lasted for 3 years now, has been associated with the necessary implementation of social and individual constraints that have not always been accepted or adopted by the general population despite their crucial role to curb the epidemic, reflecting a lack of health literacy in the general population [12]. These deficiencies were further reinforced by discordant messages issued from medical and nonmedical influencers or by deviation of control strategies that were misunderstood because of their complexity [12]. As Hannah Spring pointed out in an editorial on the COVID-19 pandemic, “good health literacy has never been more crucial for survival” [12].

Principal Results

Our results show that female sex was an independent determinant of better knowledge of SARS-CoV-2. We were also able to find evidence that middle and high school status was related to a low score of knowledge about SARS-CoV-2. Same gender and age distribution were also associated with a better knowledge score in a recent study conducted by our Belgian neighbors [26]. Better health literacy among older people seems, however, to be specific to the COVID-19 pandemic, as previous studies have shown that older age is a determinant of lower health literacy [27-30].

According to our findings, the social gradient did not appear as a determinant of better health literacy regarding COVID-19. The same observation was reported by Okan et al [31] in the German population. They suggested that the lack of a social gradient may mean that a great deal of information about the COVID-19 epidemic and SARS-CoV-2 has been made available and understandable, and that the informational environment helps to build health literacy on the topic [31]. It contrasts with other contexts than the COVID-19 pandemic, where better health literacy is usually observed in higher social classes [32]. Our results emphasize that the EDI cannot be settled as a determinant of health literacy, SARS-CoV-2 positivity, or compliance with existing health measures. During the first wave of the epidemic, it appeared that French and Italian citizens—2 populations generally known for their tolerance of deviation from sanitary norms—appeared to be much more disciplined and more respectful of the recommendations issued by public health authorities than Chinese, German, or British citizens, which can be considered as paradoxical [33,34]. Indeed, several studies have suggested that preventive health behavior is strongly influenced by social norms [33,35,36]. We can then hypothesize that, in France, during the beginning of the pandemic, citizens became compliant to sanitary rules as a new normative reference [37].

Factors independently associated with a higher respect of barrier gestures were older age and female sex. Of note, we also found a positive correlation with global knowledge scores. In a recent study on the determinants of preventive behaviors in response to the COVID-19 pandemic in France, we showed that men and young adults were less likely to follow guidelines to contain the spread of COVID-19 [33]. Our results are consistent with this in terms of the age and gender distribution in this study. Other international studies highlighted that better health literacy...
during the pandemic was associated with adherence to preventive behaviors, which is in line with our results [38].

We also had the opportunity to evaluate willingness and the possibility of isolation in case of SARS-CoV-2 infection. Isolation of COVID-19–positive individuals is crucial to contain an epidemic and is a factor that, to our knowledge, has not been well studied before. Our results show that determinants of a higher self-isolation possibility are older age and having no children living under the same roof. Previous literature suggests that socioeconomic status and a fear of loss of income due to home quarantine are the main barriers for adhering to self-isolation [39-41]. It is particularly true among the lowest socioeconomic status quintiles where people are usually unable to work from home [41]. During the pilot study of asymptomatic testing in Liverpool, the United Kingdom, one of the main barriers to adherence to self-isolation was the fear of testing positive and not having sufficient support to implement this individual measure or experiencing a loss of income [42]. In France, home quarantine was not associated with loss of income as the government compensated the loss of salary [43]. In our study, we showed that the major obstacle to the effective implementation of self-isolation was housing conditions, especially with households having more than 1 child. Once again, the EDI was not independently associated with isolation. Self-discipline is, however, not enough: the right conditions must be in place to minimize the risk of disrupting family or social dynamics and thus to produce the expected effects. Practical dispositions such as living conditions, ability to access basic supplies and needs, and access to health care are important factors in individuals’ decisions to comply or not with self-isolation and also possibly participate in a screening campaign [40].

Regarding RT-PCR positivity, it was interesting to observe that a significant determinant of a SARS-CoV-2 infection was the knowledge score regarding SARS-CoV-2 transmission. A 2021 report from the Norwegian Institute of Public Health showed that people with low education and low family income have higher rates of confirmed positive cases [44]. Nevertheless, we cannot make the same claim for our population, given that we did not record the school education level of our respondents. Other studies have confirmed the existence of sociodemographic factors associated with COVID-19 such as deprivation [45]. In our study, belonging to deprived EDI strata or to a low socio-professional category was not independently associated with SARS-CoV-2 infection. However, detection of SARS-CoV-2 infection in the study participants was performed only on 1 day, which was not the case in other studies.

As Schmidt et al [38] recently reported, one of the key elements to be retained from the various recent studies carried out is that the level of health literacy specific to COVID-19 seems to be a strong determinant of the individual’s commitment to adopting preventive attitudes and behaviors toward the virus [38].

Limitations

Our study has some limitations. First, Saint-Étienne ranks 14th among France’s 75 most populous cities, and 17th among 50 of France’s main urban areas. So, despite its position in the top half of France’s most populous cities, it is by no means representative of all French cities [16,46]. Second, we used EDI and IRIS to determine socioeconomic status and geographic areas. We were, therefore, unable to classify individuals by rural, periurban, or urban catchment area since EDI was created in 2007 and the urban landscape has been reshaped. Thus, it was not possible for us to verify whether there is a difference in the level of health literacy according to urban or rural areas. However, it is important to note that the literature is mixed on this specific issue [47-51]. Moreover, EDI is an area indicator and its transposition at an individual level may not be representative of the true deprivation status in numerous cases. Recording the socioeconomic status of individuals would have been more precise but would have increased the length of the questionnaire and may have stigmatized people; hence, we did not make this choice.

We excluded participants with missing values, which can limit the representativeness of the study sample relative to the target population. The most frequent reason for exclusion was the impossibility to determine the IRIS area where participants lived. A proportion of exclusions was also due to incomplete filling of the questionnaire, resulting in an impossibility to calculate scores for 4% to 8% of participants, depending on the score. We cannot rule out a bias of social desirability in the declarative responses to questions about respondents’ behavior with respect to their adherence to existing barrier practices. Moreover, we cannot exclude a representation bias as volunteer participants self-presented to take part in a mass screening campaign and were probably already aware of strategies deployed to contain the epidemic, and their behaviors were perhaps not representative of those of the general population. Especially since the demographic characteristics of our population are slightly different from those of the French general population, we had a notably higher proportion of older individuals [52]. However, we analyzed their behaviors according to their sociodemographic characteristics. Nevertheless, the context, the population surveyed led us to acknowledge that the results of this study have limited generalization. Furthermore, the results obtained were not used to implement public health interventions since the epidemic situation improved. However, our results could help customize specific avenues in case of a future emergency.

Comparison With Prior Work

Added Value of This Study

To our knowledge, this is the first study to assess the knowledge of, behavior toward, and isolation for COVID-19 in a large-scale population during a mass city-wide testing using a validated questionnaire (CovQuest-CC questionnaire) [19]. This evaluation was correlated with the results of the SARS-CoV-2 RT-PCR testing performed for each individual during the mass city-wide testing. A correlation was found between knowledge scores and positive findings on RT-PCR testing among asymptomatic people. In the setting of the COVID-19 pandemic, determinants of better health literacy were not the same as those in other contexts. Other works have evaluated such types of data on a large scale but never in the context of city-wide mass testing and with a lower number of respondents [26,31,33].
Implications of all Available Evidence

As pointed out by Mühlbacher et al [53], restrictive measures in the event of a pandemic can only be successful if they are accepted by the population and if political decision makers can count on the approval of a large with the majority of citizens.

In this way, knowledge of behavioral determinants can help to implement appropriate nonpharmaceutical interventions, supported by public commitment, in the event of a large-scale health threat.

Thus, assessing health-related knowledge about SARS-CoV-2 is crucial, particularly in a large population, in order to transition to public health intervention research, which will provide essential multilevel responses at organizational, societal, and individual levels, while enabling the implementation of structured, individualized interventions aimed at different typologies of individuals.

Conclusions

In conclusion, this study identifies factors associated with health literacy regarding SARS-CoV-2 infection in asymptomatic individuals in a large French population. We can confirm that, in the context of the COVID-19 pandemic, the determinants of better health literacy are not the same as those in other contexts. In particular, we were able to establish a significant relationship between a low health literacy score on SARS-CoV-2 and positivity in an asymptomatic population. We were also able to highlight that female gender is an independent determinant of a better level of health literacy regarding SARS-CoV-2, and that middle and high school status signaled a low level. It seems extremely important to obtain a more detailed understanding of the determinants of individual citizens’ behavior as part of a strategy to combat the large-scale spread of a virus. The harsh experience of this pandemic should teach us how to nurture research to structure customizable interventions to encourage the adoption of ad hoc behaviors to engage citizens in adapting behaviors more favorable to their health.

Acknowledgments

This study was supported by the French Ministry of Health and Prevention.

Data Availability

The data sets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed univariate and adjusted multivariate linear regression exploring the associations between the Knowledge Global Score and the variables of interest, with non-corrected P and Bonferroni corrected P for the multivariate analysis.

[DOCX File, 20 KB - publichealth_v10i1e47170_app1.docx ]

Multimedia Appendix 2
Detailed univariate and adjusted multivariate linear regression exploring the associations between the Barrier Gesture Respect Score and the variables of interest, with non-corrected P and Bonferroni corrected P for the multivariate analysis.

[DOCX File, 21 KB - publichealth_v10i1e47170_app2.docx ]

Multimedia Appendix 3
Detailed univariate and adjusted multivariate linear regression exploring the associations between the Isolation Possibility Score and the variables of interest, with non-corrected P and Bonferroni corrected P for the multivariate analysis.

[DOCX File, 21 KB - publichealth_v10i1e47170_app3.docx ]

Multimedia Appendix 4
Detailed univariate and adjusted multivariate logistic regression exploring the associations between PCR positivity and the variables of interest, with non-corrected P and Bonferroni corrected P for the multivariate analysis.

[DOCX File, 22 KB - publichealth_v10i1e47170_app4.docx ]

References


16. Populations l


Abbreviations

CROSS: Checklist for Reporting of Survey Studies
EDI: European Deprivation Index
IRIS: regrouped statistical information blocks
RT-PCR: reverse transcriptase–polymerase chain reaction

©Elise Verot, Robin Chaux, Julie Gagnaire, Paul Bonjean, Amandine Gagneux-Brunon, Philippe Berthelot, Carole Pelissier, Billal Boulamaï, Franck Chauvin, Bruno Pozzetto, Elisabeth Botelho-Nevers. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 11.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Review

Community Engagement in Vaccination Promotion: Systematic Review and Meta-Analysis

Yao Jie Xie¹,²*, PhD; Xiaoli Liao¹*, PhD; Meijuan Lin¹, MM; Lin Yang¹, PhD; Kin Cheung¹, PhD; Qingpeng Zhang³,⁴, PhD; Yan Li¹, PhD; Chun Hao⁵, PhD; Harry HX Wang⁵,⁶, PhD; Yang Gao⁷, PhD; Dexing Zhang⁸, PhD; Alex Molassiotis⁹, PhD; Gilman Kit Hang Siu¹⁰, PhD; Angela Yee Man Leung¹,¹¹, PhD

¹School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)
²Research Centre for Chinese Medicine Innovation, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)
³Muskeeters Foundation Institute of Data Science, The University of Hong Kong, Hong Kong, China (Hong Kong)
⁴Department of Pharmacology and Pharmacy, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
⁵School of Public Health, Sun Yat-sen University, Guangzhou, China
⁶Usher Institute, Deanery of Molecular, Genetic & Population Health Sciences, The University of Edinburgh, Edinburgh, United Kingdom
⁷Department of Sport, Physical Education and Health, Hong Kong Baptist University, Hong Kong, China (Hong Kong)
⁸JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
⁹Health and Social Care Research Centre, University of Derby, Derby, United Kingdom
¹⁰Department of Health Technology and Informatics, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)
¹¹Research Institute on Smart Aging (RISA), The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

* these authors contributed equally

Corresponding Author:
Yao Jie Xie, PhD
School of Nursing
The Hong Kong Polytechnic University
11 Yuk Choi Road
Hung Hom, Kowloon
Hong Kong
China (Hong Kong)
Phone: 852 34003798
Fax: 852 23649663
Email: grace.yj.xie@polyu.edu.hk

Abstract

Background: Community engagement plays a vital role in global immunization strategies, offering the potential to overcome vaccination hesitancy and enhance vaccination confidence. Although there is significant backing for community engagement in health promotion, the evidence supporting its effectiveness in vaccination promotion is fragmented and of uncertain quality.

Objective: This review aims to systematically examine the effectiveness of different contents and extent of community engagement for promoting vaccination rates.

Methods: This study was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. A comprehensive and exhaustive literature search was performed in 4 English databases (PubMed, Embase, Web of Science, and Cochrane Library) and 2 Chinese databases (CNKI and Wan Fang) to identify all possible articles. Original research articles applying an experimental study design that investigated the effectiveness of community engagement in vaccination promotion were eligible for inclusion. Two reviewers independently performed the literature search, study selection, quality assessment, and data extraction. Discrepancies were resolved through discussion, with the arbitration of a third reviewer where necessary.

Results: A total of 20 articles out of 11,404 records from 2006 to 2021 were retrieved. The studies used various designs: 12 applied single-group pre-post study designs, 5 were cluster randomized controlled trials (RCTs), and 3 were non-RCTs. These studies targeted multiple vaccines, with 8 focusing on children’s immunization, 8 on human papillomavirus vaccine, 3 on hepatitis B virus vaccine, and 1 on COVID-19 vaccine. The meta-analysis revealed significant increases in vaccination rates both in pre-post
comparison (rate difference [RD] 0.34, 95% CI 0.21-0.47, \( I^2 = 99.9\% \), \( P < .001 \)) and between-group comparison (RD 0.18, 95% CI 0.07-0.29, \( I^2 = 98.4\% \), \( P < .001 \)). The meta-analysis revealed that participant recruitment had the largest effect size (RD 0.51, 95% CI 0.36-0.67, \( I^2 = 99.9\% \), \( P < .001 \)), followed by intervention development (RD 0.36, 95% CI 0.23-0.50, \( I^2 = 100.0\% \), \( P < .001 \)), intervention implementation (RD 0.35, 95% CI 0.22-0.47, \( I^2 = 99.8\% \), \( P < .001 \)), and data collection (RD 0.34, 95% CI 0.19-0.50, \( I^2 = 99.8\% \), \( P < .001 \)). The meta-analysis indicated that high community engagement extent yielded the largest effect size (RD 0.49, 95% CI 0.17-0.82, \( I^2 = 100.0\% \), \( P < .001 \)), followed by moderate community engagement extent (RD 0.45, 95% CI 0.33-0.58, \( I^2 = 99.6\% \), \( P < .001 \)) and low community engagement extent (RD 0.15, 95% CI 0.05-0.25, \( I^2 = 99.2\% \), \( P < .001 \)). The meta-analysis revealed that “health service support” demonstrated the largest effect sizes (RD 0.45, 95% CI 0.25-0.65, \( I^2 = 99.9\% \), \( P < .001 \)), followed by “health education and discussion” (RD 0.39, 95% CI 0.20-0.58, \( I^2 = 99.7\% \), \( P < .001 \)), “follow-up and reminder” (RD 0.33, 95% CI 0.23-0.42, \( I^2 = 99.3\% \), \( P < .001 \)), and “social marketing campaigns and community mobilization” (RD 0.24, 95% CI 0.06-0.41, \( I^2 = 99.9\% \), \( P < .001 \)).

Conclusions: The results of this meta-analysis supported the effectiveness of community engagement in vaccination promotion with variations in terms of engagement contents and extent. Community engagement required a “fit-for-purpose” approach rather than a “one-size-fits-all” approach to maximize the effectiveness of vaccine promotion.

Trial Registration: PROSPERO CRD42022339081; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=339081

(JMIR Public Health Surveill 2024;10:e49695) doi:10.2196/49695

KEYWORDS

community engagement; community-based participatory research; vaccination rate; health promotion; vaccine.

Introduction

Vaccination stands as one of the top 10 great public health achievements of the last century. It has made significant strides in eliminating and controlling various vaccine-preventable diseases, as evidenced by the reduction in morbidity, mortality, and disability caused by these diseases [1,2]. A notable illustration is the use of vaccines as a crucial measure to mitigate the COVID-19 pandemic in the past 3 years [3,4]. A previous study analyzed the economic advantages of vaccination against 10 diseases across 73 countries from 2001 to 2020. It reported that vaccinations have prevented over 20 million deaths and saved approximately US $350 billion in disease costs [5]. A modeling study examined the health implications of vaccination against 10 pathogens across 98 countries from 2000 to 2030. It revealed that vaccinations have prevented 69 million deaths [6].

Both the Global Vaccine Action Plan 2011-2020 and Immunization Agenda 2030 have established strategic objectives to immunize every eligible individual with appropriate vaccines and to ensure equitable coverage of immunization benefits for all. However, the immunization coverage of many vaccines has yet to reach the expected level. For instance, between 2006 and 2014, only 47 million women across 80 countries and territories received the full course of human papillomavirus (HPV) vaccines, representing a mere 1.4% coverage of the total female population [7]. In addition, a study assessing the coverage of childhood vaccines across 136 administrative regions in 43 countries revealed that only one-third of children in 4 countries had fully received routine childhood vaccines [8]. In terms of adult vaccination, only 11 out of 204 countries achieved the World Health Organization (WHO) target of 90% coverage for 11 routine vaccines by 2019 [9]. Various reasons and barriers contribute to the lack of vaccination, with a significant obstacle being vaccine hesitancy. Vaccine hesitancy has been steadily rising worldwide over the past decade [10,11], emerging as one of the top 10 threats to global health listed by the WHO in 2019.

Community engagement is a process that involves engaging and motivating diverse partners to collaborate in harnessing community potential and enhancing community health [12,13]. It first gained prominence in the public health sphere with the Declaration of Alma-Ata and has since become increasingly prominent, particularly with the introduction of the new Sustainable Development Goals [14]. The WHO defines community engagement as “a process of developing relationship which enables stakeholders work together to address health issues” [15]. The United Nations Children’s Fund (UNICEF) defines community engagement as “an action for working with community stakeholders to improve community health” [13]. The definition of community engagement often intersects, competes with, and contradicts definitions of other terms such as community participation and community involvement, among others. Despite the extensive literature on community engagement, there is a lack of comprehensive guidelines to clarify the content and scope of community engagement, including what constitutes community engagement and the extent of its involvement. The levels of community engagement are structured along a continuum that spans from informing and consulting to involving, collaborating, and empowering [16,17]. The elements of community engagement manifest across a spectrum of initiatives, encompassing participant recruitment, intervention development, intervention implementation, and data collection [18,19]. Community engagement is characterized as a dynamic process rather than a singular intervention, operating within diverse contexts to address various issues through multiple mechanisms involving different actors.

A meta-analysis, incorporating 131 individual studies, supported the positive impact of community engagement on health and psychosocial outcomes for disadvantaged groups across various
conditions [20]. It plays a prominent role in global immunization strategies, as it has the capacity to alleviate vaccination hesitancy and enhance vaccination confidence. A systematic review, which included 14 studies, examined the effectiveness of community interventions on HPV vaccine coverage. Of these, 12 studies reported that community interventions led to an increase in the uptake of the HPV vaccine [21]. Another review, spanning across 19 countries, assessed studies indicating that community engagement enhanced the timeliness and coverage of routine childhood immunization vaccines [22]. Despite robust evidence supporting the role of community engagement in promoting health within diverse populations, the evidence for community engagement specifically in vaccination promotion remains fragmented. Thus, we conducted a systematic review and meta-analysis to investigate the effectiveness of various aspects and levels of community engagement in enhancing vaccination rates.

**Methods**

**Overview**
This study was conducted following the guidelines outlined in the Cochrane Handbook for Systematic Reviews of Interventions [23], and the results were reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [24]. The review protocol was registered in the PROSPERO database (CRD42022339081). Two reviewers (ML and YJX) conducted the literature search, study selection, quality assessment, and data extraction independently. Any discrepancies were resolved through discussion, and a third reviewer (LY) was consulted for arbitration when necessary.

**Ethics Approval**
This review paper was a secondary analysis of existing data from original studies published before, rather than a direct collection of new data, and thus, does not require ethical approval.

**Search Strategies**
A comprehensive and exhaustive literature search was conducted across 4 English databases (PubMed, Embase, Web of Science, and Cochrane Library), as well as 2 Chinese databases (CNKI and Wan Fang). The search strategy involved combining terms related to “community engagement” and “vaccination” using specific vocabulary terms (MeSH and Emtree) and their corresponding free-text terms [25,26]. These terms were identified based on key publications in relevant fields, and the search strategy was adjusted to suit each database. Boolean operators, specifically “OR” between terms and “AND” between concepts, were used to combine search terms effectively.

No restrictions were placed on language, study design, country of origin, or publication date. Studies were searched in the selected databases from their inception to April 30, 2023. The initial literature searches were performed in June 2022, with an updated search conducted in April 2023. In addition, the reference lists of relevant articles and previous reviews were manually reviewed to identify any additional relevant studies. The ProQuest Dissertations & Theses Database was consulted to identify unpublished dissertations and theses. Furthermore, Google and Google Scholar were searched to identify gray literature for potential inclusion. Clinical trial registries, including ClinicalTrials.gov and the WHO International Clinical Trials Registry, were also searched to identify trials with outcomes that had not yet been published.

Details of the full search strategy for each database are listed in Table S1 in Multimedia Appendix 1.

**Selection Criteria**
The inclusion and exclusion criteria were established based on the participants, interventions, comparisons, outcomes, and study design (PICOS) strategy [27]. Initially, these criteria were applied to titles and abstracts, and subsequently to full-text articles, to determine their final inclusion status (Table 1).

All records retrieved from the literature search were imported into the bibliographic database EndNote (Clarivate), which was used to manage records and eliminate duplicates. Two reviewers (ML and XL) independently screened the records based on the eligibility criteria. Any discrepancies between the 2 reviewers were resolved through discussion, and a third reviewer (YJX) was consulted if consensus could not be reached. The search terms and selection criteria were designed to provide inclusive flexibility and discretion, considering the various permutations of community engagement.
Table 1. Inclusion and exclusion criteria for literature.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population (P)</td>
<td>• All age groups</td>
<td>• No restrictions</td>
</tr>
<tr>
<td>Intervention (I)</td>
<td>• Community engagement was required to meet 2 compulsory criteria [28,29]: (1) identify community partners in research and (2) engage community partners in intervention.</td>
<td>• Inability to identify community partners or failure to engage community partners.</td>
</tr>
<tr>
<td></td>
<td>• Partner engagement was required to meet 4 optional criteria [28,29]: (1) participant recruitment, (2) intervention development, (3) intervention implementation, and (4) data collection.</td>
<td></td>
</tr>
<tr>
<td>Comparison (C)</td>
<td>• Blank control, active control, and any other intervention</td>
<td>• No restrictions</td>
</tr>
<tr>
<td>Outcome (O)</td>
<td>• Vaccine rates that involved full immunization, partial immunization, and up-to-date immunization [30-32].</td>
<td>• No data on vaccine rates</td>
</tr>
<tr>
<td>Study design (S)</td>
<td>• Experimental designs that included randomized controlled trials, quasi-randomized controlled trials, non-randomized controlled trials, or controlled pre-post studies.</td>
<td>• Descriptive or conceptual studies</td>
</tr>
</tbody>
</table>

**Data Extraction**

A data extraction form was developed and piloted on 6 randomly selected sample studies to establish consensus on the data abstraction procedures. Subsequently, 2 independent investigators (ML and XL) extracted information including the first author, publication year, study design, country, participant number, intervention details, control condition, vaccine rates, and effect size of the intervention, where reported. In cases where a study provided data for both vaccine series initiation and completion, only the latter was included in the summary table. If a study evaluated multiple vaccine types and reported a combined vaccination rate, that result was selected; otherwise data for the primary vaccine under focus were presented. In instances where a study reported incomplete data, the authors were contacted via email to obtain the required information.

**Assessment of the Risk of Bias**

The revised Cochrane Tool for Risk of Bias 2.0 (RoB2) was used to assess the risk of bias in randomized controlled trials (RCTs) [33]. For nonrandomized trials and controlled pre-post studies, the Risk of Bias in Non-Randomized Studies-of Interventions (ROBINS-I) tool was used to evaluate the risk of bias [34].

Each study was assessed and categorized as having low, moderate, or high risk of bias for each domain. Studies with low risk in 3 or more domains and moderate risk in any remaining domain(s) were classified as having an overall low risk of bias. Studies with moderate risk in 3 or more domains and low or unclear risk in any remaining domain(s) were classified as having an overall moderate risk of bias. Studies with high risk in 3 or more domains and moderate risk in any remaining domain(s) were classified as having an overall high risk of bias. Studies with moderate risk in 3 or more domains and high risk in any remaining domain(s) were also classified as having an overall high risk of bias.

**Data Synthesis**

Descriptive statistics were used to summarize the key variables of the included studies. Meta-analysis was conducted using Stata version 15.1 (StataCorp LLC) to investigate the effectiveness of community engagement in promoting vaccination.

Vaccination rates were computed as the proportion of vaccinated individuals to the total targeted population. Effect sizes were represented as the rate difference (RD) of vaccination rates, along with 95% CIs [35,36]. Random effects models were used to calculate pooled effect sizes, considering the expected heterogeneity among studies. Standard errors were adjusted for clustering effects when trials used a cluster randomized controlled design.

Forest plots were used to display individual and pooled vaccination rates. Heterogeneity was assessed using the Cochrane $Q$ test ($P_{CQ} < .10$) and the $I^2$ statistics. Subgroup analyses were conducted based on age groups, vaccine types, and immunization. A meta-regression analysis was performed to explore the effects of study design or quality on the pooled effect size [37,38]. Sensitivity analysis using a single-study knockout approach was performed to determine the contribution of each study to the pooled effect size. Publication bias was evaluated through visual inspection of the funnel plot, and the asymmetry of the funnel plot was further assessed using the Egger test [39]. The Egger tests required a minimum of 10 publications to examine the association between SE and effect size in the funnel plot [37]. We classified the evidence quality into different levels according to the recommendations from van Tulder et al [40].
Results

Study Identification and Selection
The flowchart depicting the study selection process is presented in Figure 1. The literature search was conducted across 6 electronic databases from July 5, 2022, to July 12, 2022, yielding a total of 11,404 records. After removing duplicates, 9512 articles remained. Following the preliminary review of titles and abstracts, 83 articles were retained for full-text assessment. Subsequently, after reviewing the full texts, the final selection of 19 eligible articles was made. An additional article was identified through a manual search of reference lists. Therefore, a total of 20 eligible articles published in English were identified that met all inclusion criteria.

Figure 1. The flowchart of study selection. Community engagement is a process that involves engaging and motivating diverse partners to collaborate in harnessing community potential and enhancing community health.

Characteristics of the Included Studies
The characteristics of the included studies are summarized in Table S2 in Multimedia Appendix 1. This review did not restrict the timeframe of the literature search to provide a broad temporal perspective. The included studies were published between 2006 and 2021, with the majority (n=8) in the last 5 years. These studies were conducted in various countries, with the highest number (n=13) conducted in the United States [41-53], followed by Nigeria (n=3) [54-56] and Peru (n=2) [57,58], and 1 each in Pakistan [29] and India [59]. The studies used various designs, with over one-half (n=12) adopting single-group pre-post study designs [41,44,46,48,53,54-56,59], while the rest used cluster RCTs (n=5) [29,45,52,55,59] and non-RCTs (n=3) [49-51]. The studies recruited participants across all age groups, spanning from children (n=8) [29,43,44,46,54-56,59], to adolescents (n=7) [41,49-52,57,58], and to adults (n=5) [42,45,47,48,53]. The sample sizes of pre-post studies ranged from 30 to 12,103, with a median of 323, while the sample sizes of RCTs ranged from 337 to 2598, with a median of 349. These included studies targeted multiple vaccines, with 8 studies focusing on children’s immunization [29,43,44,46,54-56,59], 8 studies on HPV vaccine [41,47,49-52,57,58], 3 studies on hepatitis B virus (HBV) vaccine [42,45,53], and 1 study on COVID-19 vaccine [48]. Vaccination coverage was calculated using either individual-reported or officially recorded data.

Conceptualization of Community Engagement
Community engagement does not neatly fit into predefined typologies, as it encompasses a variety of contexts, extents, and outcomes [60,61]. To address this complexity, a conceptual framework of community engagement was developed. This framework aims to delineate the different contents and extent of community engagement, drawing from the WHO definition of community engagement [62] and the utilitarian perspective...
of community engagement [63]. The contents of community engagement were delineated into 4 main categories: participant recruitment, intervention development, intervention implementation, and data collection. The extents of community engagement were categorized as low, moderate, and high [64]. Specifically, a low extent of community engagement indicated that studies fulfilled 1 or 2 contents of community engagement; a moderate extent of community engagement indicated that studies fulfilled 3 contents of community engagement; and a high extent of community engagement indicated that studies fulfilled all 4 contents of community engagement [64].

Most studies incorporated 2 engagement contents, with the majority engaged in intervention implementation (19/20, 95%) [29,41-46,48-59] and intervention development (13/20, 65%) [41-43,45-50,52-54,56], followed by participant recruitment (12/20, 60%) [41,43-49,51,56-58] and outcome evaluation (11/20, 55%) [29,42-44,46,48,51-53,55-57] (Table 2). Furthermore, most studies fell into the moderate engagement extent category (n=10) [41,42,44,45,49,51,53,54,56,57], followed by low engagement extent (n=7) [29,47,50,52,55,58,59] and high engagement extent (n=3) [43,46,48] (Table 2).

### Table 2. The contents and extent of community engagement in included studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Participant recruitment (n=12)</th>
<th>Intervention development (n=13)</th>
<th>Intervention implementation (n=19)</th>
<th>Data collection (n=11)</th>
<th>The number of community engagement content</th>
<th>The extent of community engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bailey et al [53]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Ma et al [45]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Weir et al [42]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Levinson et al [57]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Abuelo et al [58]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>2</td>
<td>Low</td>
</tr>
<tr>
<td>Parra-Medina et al [51]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Lee et al [47]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>2</td>
<td>Low</td>
</tr>
<tr>
<td>Paskett et al [52]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>2</td>
<td>Low</td>
</tr>
<tr>
<td>Sanderson et al [50]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>2</td>
<td>Low</td>
</tr>
<tr>
<td>Lennon et al [41]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Ma et al [49]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Findley et al [43]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>4</td>
<td>High</td>
</tr>
<tr>
<td>Willis et al [46]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>4</td>
<td>High</td>
</tr>
<tr>
<td>More et al [59]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td>Habib et al [29]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>2</td>
<td>Low</td>
</tr>
<tr>
<td>Bawa et al [54]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Oyo-Ita et al [55]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>2</td>
<td>Low</td>
</tr>
<tr>
<td>Akwataghibe et al [56]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Suryadevara et al [44]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3</td>
<td>Moderate</td>
</tr>
<tr>
<td>Marquez et al [48]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>4</td>
<td>High</td>
</tr>
</tbody>
</table>

Community engagement in these studies took various forms of intervention strategies, including social marketing campaigns, community mobilization, health education and discussions, health service support, and follow-up and reminders. These interventions were often combined into intervention packages, which included combinations such as health education and discussion with follow-up and reminders, health education and discussion with health service support, health education and discussion with health service support and follow-up reminders, social marketing campaigns and community mobilization with health service support and follow-up reminders, and social marketing campaigns and community mobilization with health education and discussion as well as follow-up reminders. Community engagement varied in geographical coverage, ranging from localized sites in 1 village or city to broader areas encompassing 1 district or more. However, many interventions failed to consider implementation constraints and practicalities on the ground, which in turn limited the fidelity of community engagement and the efficient utilization of community resources.

https://publichealth.jmir.org/2024/1/e49695
**Risk-of-Bias Assessment**

These studies exhibited variable quality across different study designs, with none meeting all the elements of a good quality design. Individual domain ratings and overall bias risk ratings for each study are presented in Tables S3 and S4 in Multimedia Appendix 1.

Two cluster RCT studies [29,59] were identified as having a low risk of bias, 1 [55] as a moderate risk of bias, and 2 [45,52] as a high risk of bias. This variability in the risk of bias may result from incorrect randomization procedures, deviations in intervention implementation, and incomplete outcome reporting. In addition, 6 quasi-experimental studies [44,48,49,53,54,57] were rated as having a low risk of bias, 4 [42,46,47,56] were rated as having a moderate risk of bias, and 5 [41,43,50,51,58] were rated as having a high risk of bias. The sources of bias in these studies may include confounding effects, missing outcome data, and selective reporting of results.

**Overall Meta-Analysis of Community Engagement on Vaccination Rates**

The pooled meta-analysis incorporated usable data from 21 intervention groups across 20 studies. The random effects meta-analysis of pre-post intervention effects revealed a moderate positive effect size of community engagement on promoting vaccination rates (RD 0.34, 95% CI 0.21-0.47, $I^2=99.9\%$, $P_{CQ} < .001$; Figure 2, see also [29,41-59]). Similarly, the random effects meta-analysis of between-group intervention effects showed a small positive effect size of community engagement on promoting vaccination rates (RD 0.18, 95% CI 0.07-0.29, $I^2=98.4\%$, $P_{CQ} < .001$; Figure 2).

**Figure 2.** Meta-analysis of the effects of overall community engagement on vaccination rates. See also [29,41-59]. RD: rate difference.
Meta-Analysis of Community Engagement Contents and Extent on Vaccination Rates

Regarding the contents of community engagement, the random effects meta-analysis revealed that participant recruitment yielded the largest effect size (RD 0.51, 95% CI 0.36-0.67, \( I^2 = 99.9\%, P_{CQ} < .001 \)), followed by intervention development (RD 0.36, 95% CI 0.23-0.50, \( I^2 = 100.0\%, P_{CQ} < .001 \)), intervention implementation (RD 0.35, 95% CI 0.22-0.47, \( I^2 = 99.8\%, P_{CQ} < .001 \)), and data collection (RD 0.34, 95% CI 0.19-0.50, \( I^2 = 99.8\%, P_{CQ} < .001 \); Figure 3, see also [29,41-59]).

![Figure 3](https://publichealth.jmir.org/2024/1/e49695/fig3.png)

Figure 3. Meta-analysis of the effects of different contents of community engagement on vaccination rates. See also [29,41-59]. RD: rate difference.

With regard to the extent of community engagement, the random effects meta-analysis found that high community engagement extent yielded the largest effect size (RD 0.49, 95% CI 0.17-0.82, \( I^2 = 100.0\%, P_{CQ} < .001 \)), followed by moderate...
community engagement extent (RD 0.45, 95% CI 0.33-0.58, $I^2=99.6\%$, $P_{CQ}<.001$) and low community engagement extent (RD 0.15, 95% CI 0.05-0.25, $I^2=99.2\%$, $P_{CQ}<.001$; Figure 4, see also [29,41-59]).

**Figure 4.** Meta-analysis of the effects of different extents of community engagement on vaccination rates. See also [29,41-59]. RD: rate difference.

---

Meta-Analysis of Intervention Strategies on Vaccination Rates

With regard to single types of intervention strategies, the meta-analysis of 4 intervention strategies found that “health service support” yielded the largest effect sizes (RD 0.45, 95% CI 0.25-0.65, $I^2=99.9\%$, $P_{CQ}<.001$), followed by “health education and discussion” (RD 0.39, 95% CI 0.20-0.58, $I^2=99.7\%$, $P_{CQ}<.001$), “follow-up and reminder” (RD 0.33, 95% CI 0.23-0.42, $I^2=99.3\%$, $P_{CQ}<.001$), and “social marketing campaigns and community mobilization” (RD 0.24, 95% CI 0.06-0.41, $I^2=99.9\%$, $P_{CQ}<.001$; Figure 5, see also [29,41-59]).
With regard to combined types of intervention strategies, the meta-analysis of 5 intervention strategy packages found that intervention package 2 yielded the largest increase in vaccination rates (RD 0.64, 95% CI 0.31-0.97, $I^2=99.3\%$, $P_{CQ}<.001$), followed by intervention package 3 (RD 0.58, 95% CI 0.05-1.11, $I^2=99.1\%$, $P_{CQ}<.001$), intervention package 4 (RD 0.31, 95% CI 0.20-0.41, $I^2=99.2\%$, $P_{CQ}<.001$), and intervention package 1 (RD 0.25, 95% CI 0.09-0.41, $I^2=98.6\%$, $P_{CQ}<.001$). However, intervention package 5 had no statistically significant impact on vaccination rates (RD 0.07, 95% CI 0.00-0.14, $I^2=72.7\%$, $P_{CQ}=.03$; Figure 6, see also [41,43-45,47,49-57,59]). Data from 5 studies were not synthesized because of high heterogeneity in their intervention strategies [29,42,46,48,58].
Subgroup Analyses of Age Groups, Vaccine Types, and Immunization Definitions on Vaccination Rates

Subgroup analyses revealed that adults (RD 0.50, 95% CI 0.16-0.85, $I^2=100.0\%$, $P_{CQ}<.001$) exhibited a larger effect size compared with adolescents (RD 0.44, 95% CI 0.18-0.70, $I^2=99.3\%$, $P_{CQ}<.001$) or children (RD 0.18, 95% CI 0.04-0.33, $I^2=99.7\%$, $P<.001$; Figure S1 in Multimedia Appendix 1). HPV vaccination (RD 0.44, 95% CI 0.18-0.70, $I^2=99.3\%$, $P_{CQ}<.001$) exhibited a larger effect size compared with HBV vaccination (RD 0.42, 95% CI 0.12-0.72, $I^2=99.8\%$, $P_{CQ}<.001$) or children immunization (RD 0.18, 95% CI 0.04-0.33, $I^2=99.7\%$, $P_{CQ}<.001$; Figure S2 in Multimedia Appendix 1). Full immunization (RD 0.41, 95% CI 0.30-0.53, $I^2=99.5\%$, $P_{CQ}<.001$) exhibited a larger effect size compared with partial immunization (RD 0.20, 95% CI 0.08-0.33, $I^2=93.3\%$, $P_{CQ}<.001$). However, no significant increase was found in the vaccine rate of up-to-date immunization (RD 0.25, 95% CI –0.10 to 0.60, $I^2=100.0\%$, $P_{CQ}<.001$; Figure S3 in Multimedia Appendix 1).

Sensitivity, Meta-Regression, Publication Bias, and Evidence Level

Sensitivity analysis showed that no significant changes were observed in the effect size of the pre-post intervention effect analysis (Figure S4 in Multimedia Appendix 1). However, the pooled effect size decreased dramatically when eliminating the study of Ma et al [45] in the between-group intervention effect analysis (RD 0.08, 95% CI 0.02-0.20; Figure S5 in Multimedia Appendix 1). Meta-regression analyses did not show any association between effect size and study design or study quality for the pre-post intervention effect analysis ($P=.16$ or $P=.65$; Figure S6 in Multimedia Appendix 1). As evidenced by the funnel plot and Egger test, no discernible signs of publication bias were detected either in the pre-post or in the between-group intervention effect analyses ($P=.25$; Figures S7 and S8 in Multimedia Appendix 1). According to recommendations from van Tulder et al [40], evidence quality in our meta-analysis was graded as moderate in both the pre-post and between-group
intervention effect analyses (Table S5 in Multimedia Appendix 1).

**Discussion**

**Principal Findings**

Community engagement drives interventions operated in a bottom-up manner rather than the traditional top-down approach. This approach supports stakeholders coming together to achieve global vaccination coverage goals from childhood to adulthood.

This study reported that community engagement strategies resulted in a 34% increase in vaccination rates through a pre-post intervention effect analysis and an 18% increase in vaccination rates through a between-group intervention effect analysis. The random effects meta-analyses indicated that participant recruitment exhibited the largest effect size, followed by intervention development, intervention implementation, and data collection. Consistent with previous evidence [64], intervention implementation constituted the primary engagement approach of most included studies and yielded a moderate effect size, while participant recruitment represented the engagement approach of over half of the included studies and yielded the largest effect size. Community partners who possess the knowledge and skills to effectively approach the target population and actively engage in participant recruitment hold the most potential to achieve relatively high recruitment and retention rates for participants. This meta-analysis found that the effect size increased with the extent of community engagement, with the highest community engagement extent exhibiting the largest effect size. Similar to previous evidence [65], a higher extent of community engagement resulted in greater vaccination promotion. Previous systematic reviews, which classify community engagement into different levels from low to high, also reported positive correlations between community engagement extents and intervention effects [65].

Regarding intervention strategies, the meta-analyses showed that health service support yielded the largest effect size, followed by health education and discussion, follow-up and reminder, and social marketing campaigns and community mobilization. Similar to previous studies, health service support increased routine childhood vaccine coverage [66]; health education and discussion increased HPV vaccine coverage among adolescents [21,67] and influenza vaccine coverage among older adults [68]; follow-up and reminder increased HBV vaccine coverage among adults [69]; and social marketing campaigns and community mobilization increased routine childhood vaccine coverage [70]. Health service support, whether used alone or in combination with other strategies, demonstrated effectiveness for vaccination promotion. Consistent with our analysis, previous studies have supported the effectiveness of health service support in the form of free vaccination outreach or mobile clinic vaccination [66,71], and flexible vaccination schedules [72]. Intervention packages that combined health service support with the other 3 strategies resulted in a significant boost in vaccine rates. The intervention packages with adaptability and flexibility, which incorporated diverse intervention strategies, could effectively meet the needs of the community population and maximize intervention benefits.

Meta-analyses across a broad range of topics, populations, and interventions often encounter a disjunction between considerable heterogeneity arising from broad questions and the limited statistical methods available for variance analysis. The limited number of included studies precluded the performance of subgroup analyses and meta-regressions to fully address the sources of heterogeneity. The development of a conceptual framework provided homogeneity at the theoretical level despite the unavoidable nature of situational heterogeneity.

The geographic coverage of the included studies spanned across 5 countries, with most studies located in the United States, which could reflect a type of publication bias along with the skewed nature of global health research. These included studies were published between 2006 and 2021, with the majority in the last 5 years, which could reflect increased academic enthusiasm and enhanced policy support in recent years. However, most studies failed to disclose the social characteristics of community participants, which highlights the reality of known social hierarchies within communities.

Many studies proposed operational definitions of community engagement, and some studies suggested empirical models to explain its connotation. However, few articles made references to definitions or frameworks, reflecting a lack of theoretical basis and critical perspective. The lack of common definitions, along with the absence of conceptual frameworks, has led to diversified procedures and contents of engagement across diverse contexts and practices. Despite the wide acceptance of community engagement in theory and practice, considerable challenges remain in identifying the best engagement approach and evaluating engagement effectiveness [73-75]. Community engagement shares similar spirits but varies in practices, as the extent of engagement spans a spectrum from minimal superficial involvement to fully collaborative participation. Operating community engagement is cost-intensive rather than cost-neutral, requiring labor, capital, and time to establish, develop, and sustain fruitful partnerships, thus posing challenges to its successful and sustainable implementation. These included studies failed to report any analysis of costs, which precluded conclusions about the economic case for community engagement. While studies support the value of community engagement, the evaluation of community engagement has largely focused on health outcomes and ignored economic information. Future studies should incorporate economic analysis to explore the potential cost-utility and cost-effectiveness of community engagement in real-world contexts. This will help close the research-practice gap and facilitate evidence-based policy making.

The inclusion of experimental designs allowed the identification of a clear link between community engagement and vaccination promotion. However, none of these included studies were located at the top level of the evidence hierarchy, which limited the direct contribution of community engagement to vaccination promotion. Future studies with more rigorous designs should be performed to draw more definitive linkages about which participant group benefits the most from which engagement
type in what community context. Randomized trials followed allocation sequence methods to ensure between-group comparability, but most interventions differed from comparisons in more ways than just community engagement. The comparator for community engagement always involved a completely different multicomponent intervention rather than the conventional health promotion activity without community engagement. The lack of a pure comparator in most community engagement interventions could cloud the interpretation of this meta-analysis. Community engagement often operates in nonlinear pathways synergized between various components and multiple outcomes, thus complicating effect evaluation compared with simple dose-response relationships. Community engagement functions as a dynamic process rather than as a discrete intervention, implying that evaluation should fully account for intrinsic complexities rather than simply focusing on outcome indicators. The primary studies should conduct thorough process evaluations to incorporate a spectrum of outcome measures and complement qualitative evaluations to elucidate the active ingredients of community engagement and the potential unintended effects of community engagement.

The effects of community engagement on vaccination promotion did not occur as a linear progression, but rather consisted of complex processes influenced by facilitators or challenges. These included studies identified individual- and community-level factors that facilitated or challenged community engagement in the context of vaccination promotion. At the individual level, the sense of confidence and ownership, along with the development of leadership skills and knowledge, facilitated community partners to engage with participatory processes. Conversely, the lack of interest and capacity, as well as the ambiguity of role and responsibility, challenged community partners to engage with participatory processes. At the community level, trust facilitated effective community engagement, while mistrust inhibited genuine community engagement. Further work should adopt a broader range of study designs that encompass both quantitative and qualitative engagement. Further work should adopt a broader range of study designs that encompass both quantitative and qualitative methodologies to measure these intangible facilitators or challenges in the area of community engagement.

These included studies faced the challenge of measuring the level of community engagement, as engagement levels span a spectrum from more passive involvement to more active participation. This study proposed operationalized extents of community engagement beyond levels of community engagement from a pragmatic perspective. Further studies should be performed to develop tools or standards to measure and evaluate the levels of community engagement effectively.

As most studies narrowly defined community engagement as an intervention program imposed on the community, they framed the effectiveness of community engagement in terms of short-term individual-level outcomes [60] while neglecting multidimensional community-level outcomes. A narrow definition of community engagement, along with a restricted view of effectiveness, excludes a conceptually coherent and methodologically sound evaluation of community engagement [15]. Evaluating community engagement raises a unique set of challenges around conceptual, methodological, and practical aspects [76]. The interaction between the engagement strategy and the community system creates a degree of complexity beyond the detail of intervention implementation [77]. This complexity grows in concert with the delivery of the engagement strategy, which may, in some instances, reshape the intervention and the community context [77]. Future work should focus on intervention theories, logic models, and outcome frameworks to clarify the relationship between community engagement and health outcomes.

Community engagement can function independently or in conjunction with other initiatives. However, when combined with other initiatives, it becomes challenging to isolate the specific contribution of community engagement to health outcomes [78]. On the other hand, some studies treated community engagement as a discrete intervention rather than a dynamic process. This oversight has resulted in a lack of alternative process evaluations to explore how community engagement contributes to vaccination promotion [79-81]. Despite the widespread use and recognized significance of community engagement [82], there are still gaps in measuring and evaluating its implementation. While there is a vast body of literature on community engagement spanning various disciplines, comprehensive guidelines and frameworks for community engagement are lacking. The adoption of consistent guidelines and frameworks can formalize the implementation and evaluation of community engagement efforts.

Limitations
This study faces some challenges and limitations that warrant consideration and point toward future directions. The first challenge was the range of different definitions and terminology referring to engagement versus involvement and participation. The second challenge was the disjunction between the conceptual heterogeneity inherent in such broad questions and the limited statistical methods available to analyze variance. The third limitation was the possibility of study omission due to search deficiencies or publication bias, despite the extensive and rigorous literature search conducted.

Conclusions
The findings of this meta-analysis support the effectiveness of community engagement in promoting vaccination, with variations observed in terms of the contents and extent of engagement. Experimental studies often involve differences between the intervention and comparison groups beyond just community engagement. Studies designed to specifically isolate community engagement as the only differing factor between the intervention and comparison groups are suggested, which allows for a clearer understanding of its added value in vaccination promotion. Comprehensive process evaluations and qualitative evaluations should be used, to provide insights into the active ingredients of community engagement and uncover any unintended effects it may have. A further scientific agenda on community engagement should focus on theory development, framework construction, and effectiveness evaluation. Future studies will benefit from the adoption of standard guidelines and frameworks to enable cross-study or cross-country comparisons of community engagement, promoting effective, sustainable, and appropriate community initiatives.
Acknowledgments
This study is funded by the Health and Medical Research Fund (HMRF)–Commissioned Research on the Novel Coronavirus Disease (COVID-19; reference number COVID1903006-A). The funding body has (and will continue to have) no role in the study design, the collection, analysis, and interpretation of any data, or in the decision to submit the paper for publication.

Data Availability
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Disclosure of AI Use
The generative artificial intelligence was not used in any portion of the manuscript writing.

Authors’ Contributions
YJX conceptualized the study. XL and ML drafted the initial manuscript. YJX checked and revised the manuscript. All other authors contributed to article revisions for important intellectual content and approved the final draft.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Literature search results, characteristics of included studies, risk-of-bias assessment, and subgroup and sensitivity analyses.

References


Abbreviations

- HBV: hepatitis B virus
- HPV: human papillomavirus
- PICOS: participants, interventions, comparisons, outcomes, and study design
- PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- RCT: randomized controlled trial
- RD: rate difference
- ROBINS-I: Risk of Bias in Non-randomized Studies-of Interventions
- WHO: World Health Organization

©Yao Jie Xie, Xiaoli Liao, Meijuan Lin, Lin Yang, Kin Cheung, Qingpeng Zhang, Yan Li, Chun Hao, Harry HX Wang, Yang Gao, Dexing Zhang, Alex Molassiotis, Gilman Kit Hang Siu, Angela Yee Man Leung. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 10.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Relationship Between Physical Exercise and Cognitive Function Among Older Adults in China: Cross-Sectional Population-Based Study

Fubaihui Wang¹, PhD; Changqing Gao², BS; Yantao Wang³, PhD; Zhuo Li⁴, MS; Feiran Zheng⁵, MS; Yanan Luo⁶, PhD

¹Social Science of Sport Research Center, China Institute of Sport Science, Beijing, China
²Mental Health Center, Kunming Medical University, Kunming, China
³Institute for Crime Prevention, Ministry of Justice, Beijing, China
⁴School of Social Sciences, Tsinghua University, Beijing, China
⁵School of Ethnology and Sociology, Minzu University of China, Beijing, China
⁶Department of Global Health, School of Public Health, Peking University, Beijing, China

Corresponding Author:
Yanan Luo, PhD
Department of Global Health
School of Public Health
Peking University
No 38 Xuyuan Road, Haidian District
Beijing, 100191
China
Phone: 86 18519621115
Email: luoyanan@bjmu.edu.cn

Abstract

Background: The existing literature reveals several significant knowledge gaps that hinder health care providers in formulating exercise prescriptions for cognitive health.

Objective: This study endeavors to elucidate the relationship between the level of physical activity and cognitive function in older adults in China. Moreover, it seeks to explore the associations between distinct exercise behaviors—such as exercise types, the purpose motivating engagement in exercise, the accessibility of exercise fields, and the inclination toward exercise—and cognitive function.

Methods: Using data from the China Longitudinal Aging Social Survey (CLASS conducted in 2016, cognitive function was meticulously assessed through the modified Chinese version of the Mini-Mental State Examination, encompassing measures of orientation, memory, and calculation. Using self-report structured questionnaires, a myriad of information about physical activity during leisure time, exercise engagement, exercise intensity, primary exercise types, reasons for exercise participation, availability of sports facilities, and exercise willingness was diligently gathered. Robust ordinary least squares regression models were then used to compute coefficients along with 95% CIs.

Results: A discernible inverted U-shaped trend in cognitive scores emerged as the level of physical activity surpassed the threshold of 500 metabolic equivalents of task (MET) minutes per week. Notably, individuals with a physical activity level between 500 and 999 MET minutes per week exhibited a coefficient of 0.31 (95% CI 0.09 to 0.54), those with a physical activity level between 1000 and 1499 MET minutes per week displayed a coefficient of 0.75 (95% CI 0.52 to 0.97), and those with a physical activity level above 1500 MET minutes per week demonstrated a coefficient of 0.45 (95% CI 0.23 to 0.68). Older individuals engaging in exercise at specific MET levels showcased superior cognitive function compared to their inactive counterparts. Furthermore, individuals driven by exercise motivations aimed at enhancing physical fitness and health, as well as those using sports facilities or public spaces for exercise, exhibited notably higher cognitive function scores.

Conclusions: The findings underscore the potential of exercise as a targeted intervention for the prevention and treatment of dementia or cognitive decline associated with aging in older individuals. Leveraging these insights to formulate informed exercise recommendations holds promise in addressing a significant public health challenge linked to aging populations.
Introduction

Cognitive function plays a crucial role within the framework of healthy aging proposed by the World Health Organization [1]. The decline in cognitive function can have adverse effects on social interaction, sense of purpose, independent living abilities, as well as the capacity to recover from illnesses or injuries [2,3]. Approximately 55.2 million people worldwide have cognitive impairments, and this number is projected to rise to 83.2 million by 2030 and further to 152.8 million by 2050 [4,5]. Given the rapid acceleration of global population aging, it has become imperative to identify and address potential cognitive decline. While cognitive decline is an inevitable part of the aging process, evidence suggests that cognitive plasticity can still occur, even in later stages of life [6]. The 2020 Lancet Commission on dementia prevention, intervention, and care indicates that approximately 40% of dementia cases worldwide are potentially modifiable [7], and a significant contribution can be attributed to engaging in physical activity and exercise can potentially reduce the risk of cognitive function [6,7].

Physical activity refers to any movement of the body that requires energy expenditure, such as gardening, shopping, or doing household chores [8], while exercise is a planned and structured form of physical activity that aims to improve or maintain physical fitness components [8]. Both physical activity and exercise have been extensively studied and shown to enhance cognitive performance through various molecular mechanisms, including the stimulation of brain-derived neurotrophic factors, learning, and memory [9,10]. As a result, clinical guidelines and public health recommendations often emphasize the importance of incorporating physical activity and exercise as fundamental strategies for preventing and managing dementia [7,11]. The World Health Organization suggests that older adults should aim for 150-300 minutes of moderate-intensity aerobic activity per week or 75-150 minutes of vigorous-intensity aerobic activity, along with regular muscle-strengthening exercises involving major muscle groups at least 3 times per week [12].

However, the existing literature reveals several significant knowledge gaps that hinder health care providers in formulating exercise prescriptions for cognitive health. First, although some evidence suggests the minimum effective, optimal doses, and maximum safe threshold of physical activity to enhance cognitive function in older adults [13], most of these studies have focused on high-income countries rather than limited-income countries. This discrepancy poses a challenge when attempting to develop exercise prescriptions for cognitive health in limited-income regions. Second, there is a limited number of studies that have distinguished the diverse effects on cognitive function between general physical activity and planned, structured exercise behaviors. Additionally, only a few studies have examined the specific influence of exercise itself on cognitive function while considering the controlling factors of physical activity metabolic equivalents. Third, there is insufficient evidence to determine the most effective types of exercise for improving cognitive function in older adults. Moreover, research is scarce regarding the relationship between different exercise behaviors, such as exercise purpose, chosen venue, exercise willingness, and cognitive function in older adults. However, this information concerning exercise behaviors plays a crucial role in guiding the development of exercise prescriptions aimed at enhancing cognitive health.

Therefore, to fill these gaps, this study used a nationally representative population-based study, the China Longitudinal Aging Social Survey (CLASS), to examine the relationship between exercise and cognitive function and its 3 domains among old adults. Here, we have included the assumptions of this study: first, we assume that there is a statistically significant relationship between the level of physical activity and cognitive function among older adults in China, and then, we assume that the aforementioned relationship will exhibit significant variations across different specific exercise behaviors (including various exercise types, the purpose for engaging in exercise, the availability of exercise fields, and the willingness of exercise).

Methods

Study Design and Participants

For this study, data were obtained from the CLASS conducted from 2014 to 2020. CLASS is a nationally representative survey that encompasses 28 provinces in China. It was designed by the Institute of Gerontology and implemented by the China Survey and Data Center at Renmin University of China. The survey used a stratified multistage probability sampling method to obtain participants from communities which involved 134 counties and 462 communities across the county in China. Participants aged 60 years and older were included in this study, and their demographic information, family and community details, health status, social functioning, physical activity, and retirement information were examined. Specially trained interviewers collected all the information through face-to-face interviews. More comprehensive information about this study’s design of CLASS can be found in a previous publication [14].

In 2016, the collection of metabolic equivalents of task (MET) values and detailed exercise information was limited to that specific year, while cognitive function and other covariates were collected across all 4 waves of data collection. As a result, our analytic sample includes individuals who responded to MET, detailed exercise information, cognitive function, and other covariates in 2016. In the 2016 survey, a total of 11,492 participants aged 60 years and older were involved. Among them, 2027 participants had missing data on cognitive function scores, 18 were missing information on the intensity of physical activity metabolic equivalents, 18 were missing information on the intensity of physical activity metabolic equivalents, and 2027 participants had missing data on cognitive function.
activities, 63 were missing information on exercise behaviors, and 186 were missing information on covariates. Individuals with missing values were excluded from our sample, resulting in a final analysis involving 9198 participants.

Measures

Cognitive Function

The cognitive function of the respondents, including orientation, memory, and calculation, was assessed using the modified Chinese version of the Mini-Mental State Examination (MMSE). Orientation was evaluated using 5 questions that required respondents to provide information such as the current date, the village’s name, the date of National Day, the name of the current president, and the lunar calendar year. Each correct answer was assigned a score of 1, while incorrect answers received a score of 0, resulting in a range of 0 to 5 for orientation. Memory function was measured through immediate and delayed word recall tasks, where respondents were asked to recall 3 simple Chinese words immediately and several minutes later, respectively. The scores for immediate and delayed recall were combined, resulting in a range of 0 to 6. Calculation ability was assessed by having respondents count backwards from 100 by 7 seconds for 5 consecutive times, yielding a score ranging from 0 to 5. The sum of these 3 dimensions provided a representation of respondents’ cognitive functioning, with higher scores indicating better cognitive performance. According to a previous study [15], respondents who correctly answered at least three questions related to orientation were considered to have the cognitive ability to complete the full MMSE test. Therefore, the final cognitive functioning score ranged from 3 to 16.

Physical Activity

The self-report questionnaires were used to collect information on the extent of physical activity during leisure time. The questionnaire consisted of 9 questions that aimed to measure the time taken for different levels of activity, establishing based on the standard MET, which is a physiological measure expressing the energy cost of physical activities. It represents the ratio of the rate of energy expended during a specific physical activity to the rate of energy expenditure at rest. Further, 1 MET is defined as the energy expenditure at rest, typically around 3.5 mL of oxygen per kilogram of body weight per minute (3.5 mL O2/kg/minute). In practical terms, MET values are used to quantify the intensity of various activities. This study translated physical activities to METs involved assigning MET values to various activities based on the standard MET scale according to a previous study [16]. In our study, physical activities, such as light-intensity activities (including reading, playing an instrument, working on a computer, or walking at a slow or leisurely pace), moderate-intensity activities (including brisk walking, slow cycling, or playing tennis doubles), and vigorous-intensity activities (including jogging or running, bicycling, climbing briskly up a hill, or participating in an aerobics class), were involved. MET minutes per week were derived by quantifying the energy expenditure of different activities in terms of MET values and summing these values based on the duration of each activity. According to that, the level of leisure-time physical activity in this study was then categorized into several groups based on the calculated energy expenditure. The categories included: 0 (indicating a completely sedentary lifestyle), less than 500 MET minutes per week, 500 to 999 MET minutes per week, 1000 to 1499 MET minutes per week, and greater than 1500 MET minutes per week.

Exercise Behaviors

We investigated the relationship between exercise behaviors and cognition by examining aspects including whether participants engaged in exercise, the intensity of the exercise, the primary types of exercise, the reasons for engaging in the exercise, the availability of a sports field, and the willingness to participate in the exercise. These aspects were measured by (1) exercise: the presence of exercise behavior was measured by asking the question, “have you participated in physical exercise in the past year?”; (2) the intensity of exercise: the intensity of exercise was assessed using the questions, “how many days per week do you engage in high-intensity physical exercise that lasts at least 10 min and significantly increases your heart rate and sweating?” (frequency of high-intensity exercise per week) and “how many days per week do you engage in moderate-intensity physical exercise that lasts at least 10 min, increases your heart rate, and causes slight sweating?” (frequency of moderate-intensity exercise per week); these variables were measured as continuous variables; (3) the primary types of exercise: the primary types of exercise were determined by asking the question, “what types of physical exercise do you regularly participate in?” and categorized as follows: without exercise, endurance and speed exercises (eg, walking, running, swimming, cycling, badminton, and table tennis), apparent aesthetic exercises (eg, aerobics, dance, martial arts, and yoga), and other; (4) reasons for engaging in exercise: the reasons for participating in exercise were assessed by asking, “why do you engage in physical exercise?” and categorized as follows: without exercise, physical fitness and health enhancement (eg, disease prevention, maintaining physical fitness, and improving health), no specific purpose (ie, engaging in exercise without a specific goal and just for the sake of being active), and other (eg, self-realization and encouragement from children to be active); (5) the availability of sports fields: the availability of sports fields was measured by asking, “have you used the following sports facilities?” and categorized as follows: without exercise, athletic fields (eg, public sports fields and community sports facilities), parks, squares, or community fitness equipment (eg, parks, squares, residential area open spaces, and community fitness equipment), and home’s yard or curbside (eg, roadside and personal residence); (6) the willingness to participate in exercise: the willingness to engage in exercise was measured by asking, “if suitable conditions were provided for your exercise, would you be prepared to participate in exercise?” and categorized as “without willingness,” “uncertainty,” and “willingness.”

Covariates

Age (continuous covariates), gender (men or women), marital status (living without a spouse or living with a spouse), residency (urban or rural), education level (illiteracy, primary school, junior high school, senior high school, and above), working for pay (no or yes), household wealth (Renminbi, yuan), activities of daily living (ADL) impairment (no or yes), and
having cardiometabolic disease (no or yes) were involved as covariates. Of these, cardiometabolic disease was consistent with the combination of diabetes, heart disease, and stroke according to previous studies [17-21], which was determined through the self-reported history obtained through an in-person visit with study personnel via a questionnaire.

Statistical Analysis

Descriptive statistics are presented as means (SDs) for numerical variables and as numbers (%) for categorical variables. Ordinary least squares regression analysis was used to examine whether there is a relationship between the intensity of physical activity and cognitive function which was to calculate β coefficient and 95% CIs while adjusting for various factors including age, sex, marital status, residency, education level, occupation, household wealth, ADL impairment, and presence of cardiometabolic disease. Additionally, we use 8 ordinary least squares regression models to verify the associations between specific exercise behaviors and cognitive function. To examine the robustness of our findings, we conducted a subgroup analysis to explore the association between exercise and cognitive function among different subgroups. Specifically, we analyzed the male and female subgroups, individuals residing in rural and urban areas, individuals with low or high education levels, and individuals with or without cardiometabolic disease. All statistical analyses were performed using Stata (version 15; Stata Corp).

Ethical Considerations

The data collection in CLASS was issued by the Biomedical Ethics Review Committee of Peking University (IRB00001052-11015). All participants were compensated and ethical approval for collecting data on human participants was received and updated annually at Peking University Institutional Review Board.

Equity, Diversity, and Inclusion Statement

Our study embraces diversity by examining the relationship between exercise and cognitive function specifically among Chinese older adults. By focusing on this population, we contribute to the broader understanding of the benefits of exercise across diverse demographics. We strive for equity by ensuring equal representation and access to our study, regardless of socioeconomic background or other factors. We acknowledge that equitable opportunities for participation are essential in producing meaningful and inclusive findings. Inclusion is at the heart of our research, as we value the experiences, perspectives, and contributions of all individuals involved. We recognize that diversity and inclusion enhance the validity and applicability of our findings, promoting a more comprehensive understanding of the topic.

Results

Characteristics of Participants

Table 1 displays the characteristics of participants aged 60 years and older in China. The mean age of the participants was 69.55 (SD 7.21) years. Among the participants, 51.84% (n=4768) of them were men, 50.3% (n=4627) of them resided in urban areas, and 74.36% (n=6840) of them lived with their spouse. Regarding educational background, 23.05% (n=2120) of individuals were classified as illiterate and 13.71% (n=1261) of them were engaged in paid employment. The average household wealth was CN ¥241,267 (SD CN ¥425,622 [US $33,385.6, SD US $58,895.9]). Additionally, 5.97% (n=1414) of participants exhibited ADL impairment and 38.3% (n=3523) had cardiometabolic diseases.

The average global cognition score was 13.47 (SD 3.23), with specific mean scores of 4.67 (SD 0.78) for orientation, 4.90 (SD 1.67) for memory, and 3.90 (SD 1.72) for calculation. Among the participants, 16.37% (n=1507) of them reported engaging in exercise within the past year. On average, participants dedicated 0.20 (SD 0.94) days per week to high-intensity physical exercise lasting at least 10 minutes, which substantially increased heart rate and sweating. Additionally, participants allocated an average of 0.27 (SD 1.07) days per week to moderate-intensity physical exercise. Regarding the primary types of exercise, 14% (n=1288) of participants engaged in endurance and speed exercises, 1.89% (n=174) of them were involved in apparent aesthetic exercises, and 0.49% (n=45) of them fell into other categories. Furthermore, 12.95% (n=1191) of participants exercised for physical fitness and health enhancement, while 1.34% (n=123) of them had no specific purpose for engaging in exercise. In terms of available sports field facilities, 5.74% (528) of individuals had access to athletic fields when participating in exercise, 8.56% (n=787) of them could use parks, squares, or community fitness equipment, and 2.09% (n=192) of them used their home’s yard or curbside. As for willingness to engage in exercise, 40.41% (n=3717) of them expressed uncertainty, while 41.19% (n=3789) of them confirmed their readiness to participate in exercise if suitable conditions were provided.
Table 1. Characteristics of participants aged 60 years and above in China, 2016 (N=9198).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive function scores, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Global cognition</td>
<td>13.47 (3.23)</td>
</tr>
<tr>
<td>Orientation</td>
<td>4.67 (0.78)</td>
</tr>
<tr>
<td>Memory</td>
<td>4.90 (1.67)</td>
</tr>
<tr>
<td>Calculation</td>
<td>3.90 (1.72)</td>
</tr>
<tr>
<td><strong>Physical activity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Totally sedentary</td>
<td>954 (10.37)</td>
</tr>
<tr>
<td>&lt;500 MET&lt;sup&gt;a&lt;/sup&gt;, minutes per week</td>
<td>1603 (17.43)</td>
</tr>
<tr>
<td>500-999 MET, minutes per week</td>
<td>2275 (24.73)</td>
</tr>
<tr>
<td>1000-1499 MET, minutes per week</td>
<td>2024 (22.01)</td>
</tr>
<tr>
<td>≥1500 MET, minutes per week</td>
<td>2342 (25.47)</td>
</tr>
<tr>
<td><strong>Exercise, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>7691 (83.63)</td>
</tr>
<tr>
<td>Yes</td>
<td>1507 (16.37)</td>
</tr>
<tr>
<td><strong>Frequency for high-intensity exercise, per week, mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Frequency for high-intensity exercise, per week, mean (SD)</td>
<td>0.20 (0.94)</td>
</tr>
<tr>
<td><strong>Frequency for middle-intensity exercise, per week, mean (SD)</strong></td>
<td>0.27 (1.07)</td>
</tr>
<tr>
<td><strong>Main type of exercise, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Without exercise</td>
<td>7691 (83.62)</td>
</tr>
<tr>
<td>Endurance and speed</td>
<td>1288 (14)</td>
</tr>
<tr>
<td>Apparent aesthetic</td>
<td>174 (1.89)</td>
</tr>
<tr>
<td>Other</td>
<td>45 (0.49)</td>
</tr>
<tr>
<td><strong>Reasons for exercise, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Without exercise</td>
<td>7691 (83.62)</td>
</tr>
<tr>
<td>Physical fitness and health enhancement</td>
<td>1191 (12.95)</td>
</tr>
<tr>
<td>No specific purpose</td>
<td>123 (1.34)</td>
</tr>
<tr>
<td>Other</td>
<td>193 (2.1)</td>
</tr>
<tr>
<td><strong>Availability of sports field, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Without exercise</td>
<td>7691 (83.62)</td>
</tr>
<tr>
<td>Athletic field</td>
<td>528 (5.74)</td>
</tr>
<tr>
<td>Park, square, or community fitness equipment</td>
<td>787 (8.56)</td>
</tr>
<tr>
<td>Home yard or curbside</td>
<td>192 (2.09)</td>
</tr>
<tr>
<td><strong>Willingness to exercise, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Without willingness</td>
<td>1692 (18.4)</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>3717 (40.41)</td>
</tr>
<tr>
<td>Willingness</td>
<td>3789 (41.19)</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td>69.55 (7.21)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>4768 (51.84)</td>
</tr>
<tr>
<td>Women</td>
<td>4430 (48.16)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Living without spouse</td>
<td>2358 (25.64)</td>
</tr>
<tr>
<td>Living with spouse</td>
<td>6840 (74.36)</td>
</tr>
</tbody>
</table>
### Distribution of Cognitive Function Scores According to Physical Activity Levels and Exercise Behaviors

Figure 1 shows the distribution of cognitive function scores according to physical activity levels. While participants with physical activity levels below 500 MET minutes per week exhibited lower cognitive function scores compared to those who were completely sedentary, cognitive function scores increased as the physical activity level increased from 500 to 1499 MET minutes per week (13.44-14.11). However, cognitive function scores started to decline after reaching a physical activity level of 1500 MET minutes per week or higher (13.57). Similar patterns were found in scores of orientation, memory, and calculation.

Regarding the scores of cognition based on the purpose of exercise, individuals with the purpose of exercise focused on physical fitness and health enhancement attained the highest cognitive function score (14.30). Conversely, individuals who exercised without a specific purpose had the lowest cognitive function score among those who engaged in exercise. Moreover, individuals who had access to athletic fields or parks, squares, or community fitness equipment exhibited higher cognition scores (>14.0) compared to those who exercised in their home’s yard or curbside. Lastly, individuals who expressed willingness to exercise if suitable conditions were provided had the highest cognitive function score (13.68), while those with uncertainty about exercise demonstrated a slightly lower score (13.66), and those without a willingness to exercise had the lowest cognitive function score (12.55). Comparable trends were observed in the cognitive domains of orientation, memory, and calculation. More specific details can be found in Figure 2 and Table S1 in Multimedia Appendix 1.
Figure 1. Distribution of cognition score according to physical activity levels.

Figure 2. Distribution of cognition score according to exercise.
Association Between Exercise Behaviors and Cognitive Function Among Older Adults in China

When compared to the cognition scores of individuals who were completely sedentary, there was no significant difference in the cognition scores of individuals with a physical activity level below 500 MET minutes per week (β=−0.04, 95% CI −0.28 to 0.20). However, as the physical activity level increased beyond the 500 MET minutes per week threshold, cognition scores showed an inverted U-shaped trend. The coefficient for individuals with a physical activity level between 500 and 999 MET minutes per week was 0.31 (95% CI 0.09 to 0.54), for individuals with a physical activity level between 1000 and 1499 MET minutes per week, the coefficient was 0.75 (95% CI 0.52 to 0.97), and for individuals with a physical activity level above 1500 MET minutes per week, the coefficient was 0.45 (95% CI 0.23 to 0.68).

Individuals who engaged in exercise within the past year exhibited higher levels of cognitive function compared to those without exercise engagement (β=0.49, 95% CI 0.32 to 0.66). Higher frequency of high-intensity or moderate-intensity exercise per week was associated with higher cognitive function scores (high-intensity: β=0.11, 95% CI 0.04 to 0.17; moderate-intensity: β=0.13, 95% CI 0.07 to 0.19). Specifically, individuals primarily involved in endurance and speed exercises had increased cognitive function compared to those without exercise (β=0.56, 95% CI 0.38 to 0.74). Participants who exercised to improve physical fitness or health demonstrated a coefficient of 0.49 (95% CI 0.30 to 0.68), whereas those with no specific purpose for exercise did not show a significant difference. Additionally, compared to individuals without exercise, the availability of athletic fields, parks, squares, or community sports facilities was associated with higher cognitive function scores. Individuals with exercise willingness or uncertainty if suitable conditions were provided exhibited higher cognitive function scores compared to those without exercise willingness (uncertainty: β=0.68, 95% CI 0.51 to 0.89; willingness: β=0.60, 95% CI 0.43 to 0.78). More details can be found in Table 2.
Table 2. Risk of cognitive function decline associated with exercise behaviors in China, 2016. The models were adjusted for age, sex, marital status, residency, education level, occupation, household wealth, ADL\(^a\) impairment, and having cardiometabolic disease or not.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical activity, (\beta) coefficient (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totally sedentary</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>&lt;500 MET(^b), minutes per week</td>
<td>−0.04 (−0.28 to 0.20)</td>
<td>0.92 (−0.05 to 1.89)</td>
<td>0.97 (0.00 to 1.94)</td>
<td>0.96 (−0.01 to 1.93)</td>
<td>0.92 (−0.05 to 1.89)</td>
<td>0.94 (−0.04 to 1.91)</td>
<td>1.13 (0.16 to 2.10)</td>
<td></td>
</tr>
<tr>
<td>500-999 MET, minutes per week</td>
<td>0.31 (0.09 to 0.54)</td>
<td>1.00 (−0.89 to 2.88)</td>
<td>1.14 (−0.75 to 3.02)</td>
<td>1.12 (−0.76 to 3.01)</td>
<td>1.02 (−0.86 to 2.91)</td>
<td>1.00 (−0.89 to 2.88)</td>
<td>1.02 (−0.87 to 2.90)</td>
<td>1.31 (−0.58 to 3.19)</td>
</tr>
<tr>
<td>1000-1499 MET, minutes per week</td>
<td>0.75 (0.52 to 0.97)</td>
<td>1.79 (0.36 to 3.21)</td>
<td>1.81 (0.38 to 3.24)</td>
<td>1.80 (0.37 to 3.23)</td>
<td>1.81 (0.38 to 3.24)</td>
<td>1.79 (0.36 to 3.21)</td>
<td>1.82 (0.39 to 3.25)</td>
<td>2.04 (0.61 to 3.46)</td>
</tr>
<tr>
<td>≥1500 MET, minutes per week</td>
<td>0.45 (0.23 to 0.68)</td>
<td>0.66 (0.04 to 1.28)</td>
<td>0.72 (0.10 to 1.34)</td>
<td>0.70 (0.08 to 1.32)</td>
<td>0.66 (0.04 to 1.28)</td>
<td>0.66 (0.03 to 1.28)</td>
<td>0.65 (0.03 to 1.27)</td>
<td>0.85 (0.23 to 1.47)</td>
</tr>
<tr>
<td>Exercise, (\beta) coefficient (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>N/A (^c)</td>
<td>Reference</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Yes</td>
<td>N/A</td>
<td>0.49 (0.32 to 0.66)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Frequency of high-intensity exercise, per week</td>
<td>N/A</td>
<td>N/A</td>
<td>0.11 (0.04 to 0.17)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Frequency of middle-intensity exercise, per week</td>
<td>N/A</td>
<td>N/A</td>
<td>0.13 (0.07 to 0.19)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Main type of exercise, (\beta) coefficient (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without exercise</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Endurance and speed</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.56 (0.38 to 0.74)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Apparent aesthetic</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.13 (−0.33 to 0.58)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Other</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.31 (−0.58 to 1.19)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Reasons for exercise, (\beta) coefficient (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without exercise</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Physical fitness or health enhancement</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.49 (0.30 to 0.68)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>No specific purpose</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.40 (−0.14 to 0.93)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Other</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.55 (0.12 to 0.99)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Availability of sports field, (\beta) coefficient (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without exercise</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Athletic field</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.30 (0.03 to 0.57)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Park, square, or community</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.70 (0.48 to 0.93)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Home yard or curbside</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.23 (−0.20 to 0.66)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Willingness to exercise, (\beta) coefficient (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without willingness</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Reference</td>
<td>N/A</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.68 (0.51 to 0.86)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Our research has revealed that older individuals who participate in exercise exhibit higher cognitive function levels than those who are inactive, considering a specific level of MET for physical activity. This suggests that purposeful, organized, and planned exercise activities, in contrast to general physical activity, positively influence cognitive function. Physical exercise has been scientifically proven to decrease the risk of various negative health outcomes and positively impact individuals with mild cognitive impairment or dementia. These effects may be attributed to the elevation of neurotrophic proteins, reduction in the accumulation of harmful proteins, mediation of neuroinflammation, or inhibition of neuronal functional deficits [22-24]. Additionally, our findings indicate that endurance and speed exercises exert a more pronounced influence on cognitive function. This aligns with previous research that highlights the significant role of aerobic exercise and low-intensity resistance training in inducing clinically significant cognitive changes [13,25]. Our robust analysis was indicated by similar findings of the main results.

The results of our study reveal a robust link between exercise aimed at improving physical fitness and health and higher cognitive function scores. Furthermore, we noted that individuals who possess exercise motivation tend to exhibit higher cognitive function scores compared to those lacking motivation. These findings underscore the importance of emphasizing the impact of exercise on health through health education, boosting exercise motivation, and amplifying the cognitive benefits of exercise. Moreover, our research indicates a positive correlation between engaging in exercise in sports facilities or public spaces and achieving higher cognitive function scores. This suggests the potential advantages of constructing sports facilities or park areas in the future to maximize the benefits of exercise for the older adult population. In summary, our study represents a significant step toward providing precise exercise recommendations for enhancing the overall cognitive abilities of older individuals in China. It can serve as a valuable resource for developing tailored exercise plans and implementing patient-centered care approaches to enhance cognitive function in limited-income countries moving forward [26,27].

Discussions

Our study is the first study to examine the specific associations of exercise itself (including the purpose of exercise, the choice of exercise place, and the willingness to exercise) with cognitive function while taking into account controls for the metabolic equivalent of physical activity. This study has several limitations. First, those with impaired cognitive function may have a limited capacity to exercise, which may lead to a reverse causal relationship between exercise and cognitive function. However, due to the cross-sectional design of this study, it could not exclude this influence. Subsequent research could use longitudinal designs to explore the bidirectional relationship and the causal impact of exercise behavior on cognitive function. Second, this study only incorporated respondents who completed cognitive function measurements in the CLASS survey, thereby excluding nonrespondents. This may result in an underestimation of the strength of the relationship between exercise and cognitive function. Future research should replicate this issue using a wider array of data sources. Third, due to data limitations, some important variables that could potentially impact cognitive function (such as depression, dementia, or Alzheimer disease) were not included in this study. Therefore, when interpreting our results, it is important to be mindful of this, and we hope that future research can use other data sources to further validate and delve deeper into this study’s findings. Fourth, the physical activity was measured through self-reporting based on the standard MET. Future studies should use more accurate measurement methods, such as accelerometer-measured relative physical activity intensity, to further validate our results. Fifth, this study used the MMSE to assess cognitive function, which may result in a ceiling effect for older adults. Future research should consider using other cognitive function measurement scales, such as the Montreal Cognitive Assessment, to validate our findings.

Conclusions

Our results found that there was a statistically significant relationship between the intensity of physical activity and cognitive function among older adults in China. Our findings suggested that the optimal exercise dose associated with

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willingness</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.60 (0.43 to 0.78)</td>
</tr>
</tbody>
</table>

aADL: activities of daily living.
bMET: metabolic equivalent of task.
cN/A: not applicable.
cognitive changes falls within the range of 500-1499 MET minutes per week. However, the benefits of exercise become less clear beyond 1500 MET minutes per week. Older individuals who engage in exercise at a specific level of MET for physical activity demonstrate higher levels of cognitive function compared to inactive individuals. Moreover, this relationship exhibited significant variations across different specific exercise behaviors: individuals with exercise motivation, aiming to improve physical fitness and health, and those who engage in exercise in sports facilities or public spaces, exhibit higher cognitive function scores than their counterparts. Our findings may offer practical guidance for health care professionals, enabling them to enhance cognitive health outcomes in older individuals through targeted exercise interventions and recommendations. Specifically, health care professionals should recognize that the impact of exercise motivation on cognitive function can be addressed through motivational strategies to encourage older individuals to engage in physical activity. Additionally, physicians or physiotherapists may need to advise older adults to monitor their exercise intensity within the specified range. This approach can optimize cognitive benefits while avoiding the potential diminishing returns associated with exceeding 1500 MET minutes per week. By using these findings to provide informed exercise recommendations, we can better address one of the significant public health challenges we face. Future studies can delve deeper into understanding the nuances of the optimal exercise dose associated with cognitive changes and use longitudinal designs to establish temporal relationships between exercise patterns and cognitive changes over time.

Acknowledgments
This study was supported by the Fundamental Research Funds for the China Institute of Sport Science (Basic2242) to FW. The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors accept responsibility for the decision to submit for publication.

Authors FW and YL are co-corresponding authors of this paper. Author FW can be reached at the Social Science of Sport Research Center, China Institute of Sport Science, No 11 Tiyuguan Rd, Dongcheng District, Beijing,100061, China, Email: wangfubaihui@ciss.cn.

Data Availability
The data sets generated or analyzed during this study are not publicly available owing to security protocols and privacy regulations but are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary Files.
[DOCX File, 26 KB - publichealth_v10i1e49790_app1.docx ]

References


Abbreviations

- ADL: activities of daily living
- CLASS: China Longitudinal Aging Social Survey
- MET: metabolic equivalent of task
- MMSE: Mini-Mental State Examination

©Fubaihui Wang, Changqing Gao, Yantao Wang, Zhuo Li, Feiran Zheng, Yanan Luo. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 30.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Participatory Approaches in the Context of Research Into Workplace Health Promotion to Improve Physical Activity Levels and Reduce Sedentary Behavior Among Office-Based Workers: Scoping Review

Aidan John Buffey, MSc; Christina Kate Langley, PhD; Brian P Carson, PhD; Alan E Donnelly, PhD; Jon Salsberg

Department of Physical Education and Sport Sciences, Faculty of Education and Health Sciences, University of Limerick, Limerick, Ireland

Physical Activity for Health Research Cluster, Health Research Institute, University of Limerick, Limerick, Ireland

University Academy 92 (Old Trafford Campus), Manchester, United Kingdom

Public and Patient Involvement Research Unit, School of Medicine, University of Limerick, Limerick, Ireland

Public and Patient Involvement Research Unit, Health Research Institute, University of Limerick, Limerick, Ireland

Corresponding Author:
Aidan John Buffey, MSc
Department of Physical Education and Sport Sciences
Faculty of Education and Health Sciences
University of Limerick
Castletroy
Limerick, V94 T9PX
Ireland
Phone: 353 061 202808
Email: Aidan.Buffey@ul.ie

Abstract

Background: Participatory research (PR) involves engaging in cocreation with end users and relevant stakeholders throughout the research process, aiming to distribute power equitably between the end users and research team. Engagement and adherence in previous workplace health promotion (WHP) studies have been shown to be lacking. By implementing a PR approach, the insights of end users and stakeholders are sought in the co-design of feasible and acceptable intervention strategies, thereby increasing the relevance of the research.

Objective: This scoping review aims to explore, identify, and map PR techniques and their impact when used in office-based WHP interventions designed to improve physical activity (PA) or reduce sedentary behavior (SB).

Methods: The reporting of this scoping review followed the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews). A systematic literature search of 5 electronic databases—Web of Science, PubMed, Scopus, Google Scholar, and OpenGrey—was conducted, searching from January 1, 1995, to February 8, 2023. In total, 2 independent reviewers first screened the retrieved articles by title and abstract, and then assessed the full texts based on the inclusion and exclusion criteria. The search strategy and eligibility criteria were developed and guided by an a priori population (office-based working adults), intervention (a PA WHP intervention that took a PR approach), comparison (no comparison required), and outcome (PA or SB) framework. Data were charted and discussed via a narrative synthesis, and a thematic analysis was conducted. The included studies were evaluated regarding the degree of end user engagement throughout the research process and power shared by the researchers, using Arnstein’s ladder of citizen participation.

Results: The search retrieved 376 records, of which 8 (2.1%) met the inclusion criteria. Four key strategies were identified: (1) end user focus groups, (2) management involvement, (3) researcher facilitators, and (4) workplace champions. The degree of engagement and power shared was relatively low, with 25% (2/8) of the studies determined to be tokenistic, 25% (2/8) determined to be nonparticipation studies, 25% (2/8) determined to be tokenistic, and 50% (4/8) determined to provide citizen power.

Conclusions: This review provides a foundation of evidence on the current practices when taking a PR approach, highlighting that previous office-based PA WHP studies have been largely tokenistic or nonparticipative, and identified that the end user is only engaged with in the conception and implementation of the WHP studies. However, a positive improvement in PA and...
Reduction in SB were observed in the included studies, which were largely attributed to implementing a PR approach and including the end user in the design of the WHP intervention. Future studies should aim to collaborate with workplaces, building capacity and empowering the workforce by providing citizen control and letting the end users “own” the research for a sustainable WHP intervention.

**International Registered Report Identifier (IRRID):** RR2-10.1136/bmjopen-2021-054402

**KEYWORDS**

participant research approach; workplace health promotion; physical activity; sedentary behavior; end user involvement; office based; desk based; intervention; cocreation; public and patient involvement

**Introduction**

**Background**

A growing body of literature in occupational health research, and specifically research into individuals with office- and desk-based occupations, has begun to focus on prolonged occupational sitting [1]. The workplace environment and increased use of computers have been associated with a significant reduction in physical activity (PA) and an increased prevalence of prolonged sitting, especially in desk-based office workplaces [2,3]. Sitting is classified as a sedentary behavior (SB), which is a term used to define any waking behavior with an energy expenditure of ≤1.5 metabolic equivalents of task when sitting, reclined, or in a lying posture [4]. The workplace environment and organizational culture can often facilitate and promote prolonged SB [2,5]. Emerging evidence from 2 previous studies that measured sedentary time via accelerometers indicated that office workers were sedentary for a mean of 75.8% [6] to 81.8% (438.8, SD 51.5 min) of their working hours [7].

Previous research has attempted to reduce or interrupt prolonged occupational sitting in the workplace environment to varying degrees of success, by increasing PA. However, these previous workplace interventions or workplace health promotion (WHP) studies can often be characterized as “one size fits all” interventions, which have taken a traditional top-down, research-driven approach where the end users are considered passive subjects [8-10]. In this review, we define an end user as an individual working in a desk-based office environment or communities such as the organization’s workforce as this population is the target (participants) in WHP studies and the population of interest in this scoping review. Furthermore, this scoping review acknowledged the involvement of relevant stakeholders, following the description provided by Leask et al [11] regarding stakeholders as individuals or groups that are interested or involved in the implementation of an intervention but not the direct end users. Example stakeholders in this scoping review context, but not an exhaustive list, may be family of desk-based office employees, employees not in administrative or desk-based roles within the company, office managers, and company owners. While all these stakeholders are not the specific end users, they have lived experience and knowledge of the workplace and the end users and the ability to inform the design of a relevant, feasible, and acceptable WHP intervention from a stakeholder perspective and facilitate or support changes in the workplace culture, practices, and policy.

To shift research away from top-down “one size fits all” interventions, researchers have begun to take a participatory research (PR) approach to conducting studies. PR incorporates the knowledge and expertise of the end users and relevant stakeholders, thereby increasing the relevance of the research [6,12], and was described by Jagosh et al [10] as the coconstruction of research among people affected by the issues under study and researchers, stakeholders, and decision makers who have the capacity to apply the research findings. Therefore, PR allows for the tailoring of interventions by incorporating the relevant end users and stakeholders within the research process, which has been shown to promote a sense of ownership and aid the acceptability of the research when implemented, if conducted well [6,13]. When a participatory approach is not taken, research has shown that the WHP study may lead to an intervention approach, concept, or format that is inappropriate [14].

Previous research has suggested that the workplace is an ideal and valuable setting for the delivery of preventative health interventions when targeting adults, both healthy and especially those at increased risk of developing chronic diseases [15-17]. Earlier WHP interventions have targeted different aspects of the workplace environment, commuting habits, work schedules or implemented behavior change strategies to increase PA or reduce SB. However, studies targeting behavior change that did not take a PR approach have been shown to be weaker in intervention design [18] as it has been demonstrated to be beneficial for the investigating research team to acquire an understanding of the influences on the targeted behavior in the context in which they occur [19]. For example, in the workplace setting, a manager is an important stakeholder and can provide insights into the acceptability and feasibility of potential intervention strategies [20].

A research priority in WHP interventions is to create a sustainable WHP program after the completion of the study and researcher involvement [21]. Furthermore, maintaining end user adherence throughout the WHP intervention can be difficult, with documented high rates of attrition shown in previous WHP studies [21]. For instance, participants who are highly sedentary before an intervention are likely to return to their previous levels of SB due to increasing work pressures [2]. Including end users and relevant stakeholders with the aim of collaboration, education, and community action can promote active ownership of the research process and sustainability [8,22,23].
PR is a distinctive approach to research and not a particular research method that aims to distribute equitably the power between the research team and the research participants [24]. Therefore, PR is not a research method in and of itself and can take multiple forms and use varying methodologies, methods, processes, and tools [25]. When taking a PR approach, end user and stakeholder involvement can vary in intensity at distinct phases of an intervention (eg, conception, planning, conduct, evaluation, reporting, and dissemination). Conventional research methods such as focus groups and surveys can be adapted and applied in a participatory way, and therefore, any method, tool, or activity can be participatory if chosen or used collaboratively among end users, relevant stakeholders, and their academic partners [25]. The level of end user involvement in selecting, adapting, and implementing a method, process, activity, or tool could be considered more important than the method or technique itself in terms of impact on the WHP intervention. This is supported by Andersson [26], who identified that the quality and impact of the research can increase when end users are involved early in the research process, and reinforced by Vaughn and Jacquez [25], who stated that the level of participation is closely tied to the impact that research will have in real-world settings. The level of participation and power shared between researchers and end users can vary, and when power is not shared between the end user and academic partner in the decision-making process, the research cannot truly be “participatory” [8]. Arnstein [27] states that there is a crucial difference between going through an empty ritual of participation and end users being provided with the real power needed to affect the outcome of the process.

Several systematic reviews and meta-analyses have investigated workplace interventions that were designed to improve PA [21,28-30] or reduce SB within the workplace [31]. These reviews have reported positive overall benefits and that workplace interventions are generally effective in improving PA or reducing SB. However, previous systematic reviews investigating PA WHP studies have stated that the evidence is inconclusive [21,32] and called for more research into the elements of WHP studies that are likely to increase adoption and efficacy within the occupational setting [21]. Previous literature has shown that a PR approach can increase efficacy and lead to successful implementation and greater adherence in health promotion studies [13,18,22]. Thus, this scoping review provides evidence on the use of PR in WHP studies that may lead to greater adoption of PR and success of WHP interventions.

Rationale
To the authors’ knowledge, the use of PR in WHP studies has not been synthesized, and by examining how PR is currently incorporated within WHP research, we can identify the current available evidence, key approaches and methods, and the scope of reported impacts of PR, thereby providing an overview and identifying key characteristics of the current research that has used PR in WHP interventions aimed at increasing PA and reducing SB.

Objectives
This study had the following objectives: (1) to identify and map previous literature in which office-based adults have been involved in PR studies and how their involvement shaped the design of the WHP intervention, (2) to identify and discuss the methods implemented in the PR WHP studies, and (3) to discuss the evaluation and outcomes measured in the PR WHP studies included in the scoping review.

Research Question
How have previous PA WHP studies investigating office-based workers incorporated PR and the end user in their studies and to what reported benefit or detriment?

Methods
Protocol and Registration
An a priori protocol was published with BMJ Open and is available for this scoping review [33].

The reporting of this scoping review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews). The PRISMA-ScR consists of a 22-item checklist [34]. The completed checklist for this scoping review can be found in Table S1 in Multimedia Appendix 1 [16,18,35-40].

This scoping review followed the guidelines and framework published previously by Levac et al [41], who expanded and developed the methodology for scoping reviews by Arksey and O’Malley [42], as planned in the published protocol [33]. This scoping review also followed the more recent methodological guidance published by Peters et al [43,44] for conducting and reporting scoping reviews.

Eligibility Criteria
In the planning of this scoping review, the research team a priori developed a population, intervention, comparison, and outcome (PICO) framework to assist in the development of the search strategy and the inclusion and exclusion criteria. The PICO framework, as previously published in the scoping review protocol [33], was as follows: the population was office-based working adults; the intervention was PA WHP interventions that used a PR approach; for comparison, we did not wish to compare interventions or treatments (this is typical in some PICO analysis frameworks, where a comparison is not always present); and the outcome was PA or SB levels.

Articles were screened for eligibility related to our inclusion and exclusion criteria (Textbox 1), excluding non–English-language articles.
Inclusion criteria
- Population: working adults in office environments
- Language: English
- Years considered: January 1, 1995, to February 8, 2023
- Study focus: articles investigating workplace health promotion (WHP) in office-based workplaces that implemented participatory health research techniques that including a physical activity (PA) aspect to the intervention study, for example, increasing PA or decreasing sedentary behavior (SB) using steps or walking, breaks in sitting, exercise, and yoga.
- Publication status: published peer-reviewed journal articles and relevant gray literature, which was defined within this scoping review as theses or dissertations, conference papers, research and government reports, ongoing research, editorials, and textbooks.

Exclusion criteria
- Population: home-based "office" workers
- Language: studies written in a language other than English
- Years considered: studies published before January 1, 1995, or after February 8, 2023
- Study focus: studies conducted within the workplace with the aim of improving health; studies based in the community or home and not in the office environment; and health promotion (HP) interventions measuring and targeting psychological or work performance improvements and not measuring or reporting on PA or SB.
- Publication status: any other literature that was not listed in the inclusion criteria, such as websites

Rationale
- Population: the focus of this scoping review was to investigate participatory research (PR) in WHP studies in office-based participants and workplaces. Children, teenagers, and retired adults would not fit our eligibility criteria of “working adults.” Non-office workers and home-based workers may have different “health” needs related to the working environment.
- Language: the reviewers only speak English, and feasibility considerations (eg, limited resources) prevented the use of translation services.
- Years considered: a wide period was established to capture all relevant WHP research. The years considered were cut off at 1995 as this is when guidance was published by Green et al [45] for the development of PR in HP and, therefore, implemented into research practices following this year.
- Study focus: the focus of the overall research question of this scoping review was specific to PR in WHP research in office-based workplaces. Other work-based environments may carry different health-associated risks, priorities, or safety concerns, which would not be comparable to those of an office-based environment. Including PA or SB as outcome measures would allow for an evaluation of the included studies and a discussion on the effectiveness of taking a PR approach in those WHP studies. We excluded studies that did not measure or reported PA or SB and included those that did to address our research question as these outcome variables were the primary outcome variables of interest. Of the included studies, those that reported further outcome variables such as psychological well-being, diet, or work performance were eligible and included; however, we did not report or discuss these additional outcome variables in this scoping review as they were outside the scope.
- Publication status: the aim of this scoping review was to capture a wide range of literature, so including gray literature ensured a more complete search and minimized publication bias.

Data Sources, Searches, and Study Selection
A total of 5 electronic databases was systematically searched by 2 independent reviewers (AJB and CKL). These databases were Web of Science, PubMed, Scopus, Google Scholar, and OpenGrey. The full electronic search strategy for each database was previously published with the protocol [33].

The first search was executed on January 17, 2022, retrieving articles between this date and January 1, 1995. This criterion was used for all the included databases. These years were considered as part of our inclusion and exclusion criteria (Textbox 1). The second search to ensure all recent and relevant literature was retrieved was executed on February 8, 2023, retrieving articles between this date and January 17, 2022.

The retrieved articles were exported from the 5 electronic databases to EndNote (Clarivate Analytics), where duplicates were removed via the EndNote function and manually when a duplicate was missed by the software.

Following duplicate removal, retrieved articles were reviewed by 2 researchers independently (AJB and CKL); the retrieved articles were screened first by title and abstract based on the inclusion and exclusion criteria and then by full text. Any discrepancy between the 2 reviewers regarding eligibility was discussed until consensus was reached. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram was populated to show the number of articles retrieved, screened, and excluded (with reasons) before reaching our number of included studies (Figure 1).
A comprehensive evaluation of the retrieved articles was conducted, which involved an informal multistep iterative approach and screening system by the lead researcher (AJB) and second reviewer (CKL) assessing whether a retrieved study had taken a PR approach if the article met all other inclusion criteria. The first step or indication that both reviewers would try to identify when screening a retrieved article was the keyword “participatory” or a related synonym; when this keyword (“participatory”) or related synonyms were not present, the reviewers would search for words such as “co-design,” “co-development,” and “co-produced.” If these words were still not present within a retrieved manuscript, a search would be conducted for key terms such as “end user,” “stakeholder,” “manager,” and “volunteer.” Finally, a further screening of the
retrieved articles would then be conducted looking for specific nuances or characteristics of PR, such as focus groups, surveys, meetings or community boards, panels, or groups that were conducted or mentioned to be formed in the retrieved article with the objective of designing, tailoring, or facilitating an intervention. If none of the aforementioned key terms or characteristics of PR were present within a retrieved manuscript, the study was excluded.

No in-depth assessment of whether a study was participatory was completed during the screening phase. For example, studies did not need to meet certain requirements to be considered participatory. However, the level of participation was examined during data extraction and when writing the narrative synthesis of all the included studies in this scoping review.

Data Extraction and Charting Process
Data from the included studies were charted independently by one reviewer (AJB). The second reviewer (CKL) charted 2 randomly selected articles independently to duplicate and confirm the data charting process. It was previously proposed that the second reviewer would independently duplicate 10% of the included articles [29]; due to <10 articles being included, 2 articles were deemed more than the agreed upon 10% and sufficient for confirming the data charting process.

Data were extracted and charted using a Microsoft Excel (Microsoft Corp) sheet table. The Microsoft Excel table had been piloted by the research team and peer reviewed during publication of the scoping review protocol.

Data Items
When charting the data from the included studies, we sought to retrieve, extract, and chart information on variables such as study design, study purpose and aims, PR approach taken, whether oversight was included and whether this oversight was a participatory group or researcher oversight, level of involvement from end users and stakeholders throughout the study, intervention focus, data collection methods, study outcomes (primarily related to PA and SB), data analysis, and whether the studies self-evaluated the PR techniques they implemented. Further information on each of the data items described has been provided within the published protocol of this scoping review [33]. These data items were the headings used in the data extraction and charting process. Each data item had an associated question that was used to retrieve and chart the extracted information from the included articles.

Data Synthesis and Analysis
As is a standard approach when conducting scoping reviews [41,42,46], a critical appraisal of the included studies or data was not conducted. This was decided a priori, as documented in the study protocol [33].

Synthesis of Results
Following data extraction and charting, we summarized our findings via a narrative synthesis, providing a descriptive summary of the included studies and charted data. A qualitative thematic analysis was conducted using Microsoft Excel according to the approach published by Bree and Gallagher [47] and following the guidance of Braun and Clarke [48,49].

The studies were evaluated based on the degree of participation and power shared with end users when making important decisions with the public, end users, and relevant stakeholders from nonparticipation to tokenism to citizen power using the ladder of citizen participation by Arnstein [27], which is an 8-rung ladder. There are 2 rungs (manipulation and therapy) categorized under nonparticipation, 3 rungs (informing, consultation, and placation) that fall within tokenism, and 3 rungs (partnership, delegated power, and citizen control) that form citizen power [27]. Therefore, the ladder by Arnstein [27] allowed the authors to distinguish and compare the level of participation and access to power in the included studies [30]. Each of the 8 different rungs of the ladder relates directly to the extent to which end users have attained decision-making power, with the highest rung signifying complete citizen control [51].

Data from the included studies were charted independently by one reviewer (AJB). The second reviewer (CKL) charted 2 randomly selected articles independently to duplicate and confirm the data charting process. It was previously proposed that the second reviewer would independently duplicate 10% of the included articles [29]; due to <10 articles being included, 2 articles were deemed more than the agreed upon 10% and sufficient for confirming the data charting process.

Data were extracted and charted using a Microsoft Excel (Microsoft Corp) sheet table. The Microsoft Excel table had been piloted by the research team and peer reviewed during publication of the scoping review protocol.

Data Items
When charting the data from the included studies, we sought to retrieve, extract, and chart information on variables such as study design, study purpose and aims, PR approach taken, whether oversight was included and whether this oversight was a participatory group or researcher oversight, level of involvement from end users and stakeholders throughout the study, intervention focus, data collection methods, study outcomes (primarily related to PA and SB), data analysis, and whether the studies self-evaluated the PR techniques they implemented. Further information on each of the data items described has been provided within the published protocol of this scoping review [33]. These data items were the headings used in the data extraction and charting process. Each data item had an associated question that was used to retrieve and chart the extracted information from the included articles.

Data Synthesis and Analysis
As is a standard approach when conducting scoping reviews [41,42,46], a critical appraisal of the included studies or data was not conducted. This was decided a priori, as documented in the study protocol [33].

Synthesis of Results
Following data extraction and charting, we summarized our findings via a narrative synthesis, providing a descriptive summary of the included studies and charted data. A qualitative thematic analysis was conducted using Microsoft Excel according to the approach published by Bree and Gallagher [47] and following the guidance of Braun and Clarke [48,49].
Results

Study Selection

The systematic search strategy of 5 electronic databases identified 297 articles upon first search and an additional 79 articles with the updated second search, totaling 376 articles. Following the screening process of all eligible articles, which was conducted independently by 2 reviewers (AJB and CKL), of the 376 studies, 8 (2.1%) met the inclusion criteria (Figure 1). Of the 8 included studies, 2 (25%) provided additional reports, one in the form of a study protocol [52] and the second in the form of a second manuscript reporting findings obtained from the same study [53]; however, these reports were combined and only discussed as the manuscript retrieved and included via the search strategy, and therefore only 8 studies will be referred to.

Study Characteristics

A total of 38% (3/8) of the included studies were pilot studies [35-37] that did not randomize participants and used a pretest-posttest [35,36] or a posteriori quasi-experimental study design [37], whereas 50% (4/8) were cluster-randomized trials [16,18,38,39]. The final study had a longitudinal pretest-posttest study design [40]. In total, 38% (3/8) of the studies took a mixed methods research approach collecting and analyzing both quantitative and qualitative data in the same study [35-37,40]. The duration of the studies ranged from 3 weeks [36] to 18 months [40]. The sample size ranged from 5 [36] to 585 [16] across the included studies, and they were conducted in various countries, such as the United Kingdom [35,36], Australia [18,37], China [38,39], Sweden [40], and Singapore [16]. One of the included studies sampled women only [16], whereas the rest were mixed-sex studies [18,35-40]. Participants’ occupations varied across the included studies, and they were based in large IT organizations [38], telecommunications organizations (administrative and clerical workers) [37], government organizations (clerical, data entry, and call center office workers) [18], and workplaces in sectors or industries that were primarily office based and sedentary in nature (government administration and finance departments) [16]. Some participants were office workers in desk-based occupations (education or research, administration, human resources, accountancy, sales, and IT) [36]; worked within a department in an academic institution (university) [35]; and were part of a singular organization that serviced a municipality with 56,000 inhabitants and carried out assignments such as running and developing schools, providing social services, and conducting urban planning [40]. The final study did not provide specific occupations but stated that the included worksites were located in the Yangtze River Delta in China and comprised primarily desk-based occupations [39]. Further study characteristics can be found in Table 1, and the intervention components and reported outcome variables can be found in Table S2 in Multimedia Appendix 1.
Table 1. Characteristics of the included studies, including the year of publication, country of origin, number of participants, number of end users involved in the participatory research approach, and the inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Design</th>
<th>Country</th>
<th>Participant composition</th>
<th>Inclusion and exclusion criteria</th>
</tr>
</thead>
</table>
| Blake et al [38] | 2019 | 2-arm cluster-randomized waitlist control trial | China | Total: N=282  
• Intervention: n=196 (96 male participants, 49%; 97 female participants, 49.5%; and 3 not specified, 1.5%)  
• Control: n=86 (46 male participants, 53.5%; 37 female participants, 43%; and 3 not specified, 3.5%)  
• Participatory research: organizational committee=4 (2 team leaders and 2 HR officers); team leaders=31 (invited by the organizational committee to act as intervention facilitators) | Inclusion criteria: eligible clusters were 2 sites of a large IT organization, and eligible participants were employees of the organization. |
| Gilson et al [37] | 2016 | Pilot—a posteriori quasi-experimental design | Australia | Intervention protocol 1 (strategies and no prompts): n=33 (27 male and 6 female participants)  
• Exclusion criteria: N/A  
• Intervention protocol 2 (strategies and prompts): n=24 (19 male and 5 female participants)  
• Participatory research: participatory 1-hour workshop=10-15; aimed to identify occupational strategies for “sitting less and moving more” | Inclusion criteria: all members of each work team were eligible to participate.  
Exclusion criteria: N/A$^b$ |
| Griffiths et al [36] | 2022 | Pilot—mixed methods intervention | United Kingdom | Pilot: n=5 (2 male and 3 female participants)  
“Needs analysis” questionnaire: n=157 (19 male and 138 female participants)  
• Participatory research: coproduction development focus group=11 (4 male and 7 female participants); range of employees—management=3; sales=2; IT=3; HR=3  
• 4 fields of employment from 6 different organizations: public health=2; IT=5; energy supplier=3; education=2 | Inclusion criteria: (1) adults (aged ≥18 years), (2) occupying seated job roles (defined as sitting for ≥26 hours during work hours), and (3) being physically inactive (defined as not meeting UK PA$^c$ guidelines)  
Exclusion criteria: (1) individuals who currently used active workstations (defined as reporting regular use of sit-stand, treadmill, or pedal desks), (2) inability to complete any desk-based focused PA, (3) failure to occupy sedentary jobs, and (4) meeting and exceeding UK PA guidelines |
| Kong et al [39] | 2022 | Group randomized controlled trial | China | Total enrolled: N=955 (4 worksites)  
• Intervention: n=464 (2 worksites)  
• Control: n=491 (2 worksites)  
• Baseline—intervention: n=216 (2 worksites); control: n=172 (2 worksites)  
• Final evaluation—intervention: n=159 (2 worksites); control: n=119 (2 worksites)  
• Intention-to-treat analysis—intervention: n=216 (2 worksites); control: n=172 (2 worksites)  
• Participatory research: an EAB$^d$=4 to 7 employees from all occupational sectors in the worksite in each worksite (4 worksites × 4 to 7 = approximately 16 to 28; unknown number of occupational sectors and whether the EABs where present in the control worksite, so may be 2 worksites × 4 to 7 = approximately 8 to 14) | Participant inclusion criteria: (1) ages ≥18 years, (2) full-time employees, (3) not having received clinical weight-loss treatment, (4) not pregnant at the time of recruitment, and (5) having signed informed consent form  
Participant exclusion criteria: N/A  
Worksites inclusion criteria: (1) large proportion (>50%) of desk-based employees, (2) operating for >3 years, and (3) never having hosted a health management program |
• Completers (statistical analysis): n=17 (4 male and 13 female participants)  
• Participatory research: intervention development 1-hour focus group=7/11; 4/11 were unable to attend and submitted suggestions via email | Inclusion criteria: all employees of SchARR$^e$ were eligible to participate in the study.  
Exclusion criteria: N/A |
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Design</th>
<th>Country</th>
<th>Participant composition</th>
<th>Inclusion and exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parry et al [18]</td>
<td>2013</td>
<td>Parallel-arm cluster-randomized trial</td>
<td>Australia</td>
<td>• Total: N=133 (18% male and 82% female participants)</td>
<td>• Inclusion criteria: workers participating in office-bound duties for ≥26 hours per day and working ≥4 days per week</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Participants with complete data and included in the analysis: n=62 (19% male and 81% female participants)</td>
<td>• Exclusion criteria: participants were only excluded if they were unable to wear an accelerometer due to disability or if they were confined to a wheelchair.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Participatory research: participants from all 3 workplaces or interventions were asked to attend 2 structured meetings at their workplace to discuss and develop their specific workplace intervention.</td>
<td></td>
</tr>
<tr>
<td>Tan et al [16]</td>
<td>2016</td>
<td>2-arm cluster-randomized trial</td>
<td>Singapore</td>
<td>• Total: N=585 (585 female)</td>
<td>• Workplace inclusion criteria: (1) workplaces in sectors or industries that were primarily office based and sedentary in nature (ie, government administration and finance departments), (2) workplaces that were able to recruit at least 30 female employees engaged in desk-based jobs (sitting ≥50% of working hours), and (3) agreement to permit up to 10 hours of paid work time during the course of the study (12 months) for the recruited employees to participate in pretest-posttest data collection and intervention activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Intervention: n=287 (287 female); intervention—PA analyzed: n=234</td>
<td>• Participant inclusion criteria: (1) being female, (2) being aged 25 to 49 years, and (3) having a sedentary job (at least 50% of work hours seated)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Control: n=298 (298 female); control—PA analyzed: n=196</td>
<td>• Participant exclusion criteria: (1) being pregnant or lactating, (2) diagnosis of osteoporosis, (3) diagnosis of kidney problems, and (4) participation in another health program that addressed diet or PA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Participatory research: all participants in the tailored intervention received 3 participatory workshops focused on participatory skill-building activities, peer support, goal-setting exercises, and problem-solving discussions to attain individual goals and overcome individual barriers.</td>
<td></td>
</tr>
<tr>
<td>Wahlström et al [40]</td>
<td>2019</td>
<td>Longitudinal mixed methods</td>
<td>Sweden</td>
<td>• Total (baseline characteristics): N=152 (50 male and 102 female participants)</td>
<td>• Inclusion criteria: (1) aged 18 to 63 years, (2) working ≥75%, (3) &gt;60% of work hours inside the office, and (5) not planning to retire or relocate to another worksite during the study period</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Interviews: n=70 (17 male and 53 female participants)</td>
<td>• Exclusion criteria: N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Focus groups: n=15: focus office [activity-based work office where there are no fixed workstations, but instead various spaces in the office, which are designed to support the performance of different work tasks]: n=43; cell office [the most common office types are cell offices and open landscapes with fixed workstations]: n=43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• From within the organization, 7 female employees volunteered to be health promoters or inspirers. Managers were involved along with a health strategist from within the organization and the organization’s communication department; however, the PA-promoting program was initiated by the researchers with collaboration between researchers and workplace representatives, but the exact number was not stated.</td>
<td></td>
</tr>
</tbody>
</table>

^aHR: human resources.  
^bN/A: not applicable.  
^cPA: physical activity.  
^dEAB: employee advisory board.
Synthesis of Results

Overview

In this section, we provide a narrative synthesis that focuses on how participants, end users, and relevant stakeholders of the included studies were involved in the WHP studies and how their involvement shaped the design of the WHP intervention. This narrative synthesis was guided by a thematic analysis that identified themes related to how end users, stakeholders, and researchers were involved in the research process when taking a PR approach and what role they played (Table 2). Further information on the PR methods implemented in the included studies, including how they engaged end users and relevant stakeholders, the number and duration of any meetings and activities, how many end users and stakeholders were involved in the PR method or methods, and the content and agenda of any meetings or activities when engaging with the end users and relevant stakeholders, can be found in Table S2 in Multimedia Appendix 1.
Table 2. Components of the thematic analysis of end user, stakeholder, and researcher involvement in the participatory research process and overall conduct of the study. The table highlights primary themes and subthemes and provides a description of each theme with supporting illustrative quotes with references.

<table>
<thead>
<tr>
<th>Themes, subthemes, and descriptions</th>
<th>Illustrative quotes and references</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theme 1: participant workshops, focus groups, meetings, or brainstorming sessions</strong>&lt;br&gt;This theme describes the format of how the participants were typically asked to express their opinions, views, and personal experiences in relation to designing or evaluating a WHP intervention.</td>
<td></td>
</tr>
<tr>
<td><strong>Subtheme 1.1: researcher-provided evidence</strong>&lt;br&gt;This subtheme describes how, as part of meetings with participants, end users, and stakeholders, the research team would provide evidence that supported what they wished to codevelop with the participants, which was a WHP intervention designed to increase PA and reduce SB.</td>
<td></td>
</tr>
<tr>
<td>• “...During the workshop, researchers reviewed evidence on the benefits of reducing sitting and increasing physical activity...” [37]</td>
<td></td>
</tr>
<tr>
<td>• “In the first focus group, participants were educated via an online presentation on the importance of breaking up sitting time regularly for cardiovascular health...” [36]</td>
<td></td>
</tr>
<tr>
<td>• “...this involved an initial description of the associations between prolonged sitting and health.” [35]</td>
<td></td>
</tr>
<tr>
<td><strong>Subtheme 1.2: workers identified strategies for intervention development or promotion</strong>&lt;br&gt;This subtheme describes how participants and workers were asked to contribute and discuss their ideas and strategies for the development, implementation, and promotion of a WHP intervention at their workplace.</td>
<td></td>
</tr>
<tr>
<td>• “...asked to comment and discuss the concept of breaking up sitting time and the initial perceived challenges.” [36]</td>
<td></td>
</tr>
<tr>
<td>• “...workers identified and discussed occupational strategies for 'sitting less and moving more.'” [37]</td>
<td></td>
</tr>
<tr>
<td>• “...were asked to comment on and discuss how they would prefer to break up sitting time.” [36]</td>
<td></td>
</tr>
<tr>
<td>• “...‘brainstorming’ session where strategies were identified by participants on how to reduce workplace sitting time.” [35]</td>
<td></td>
</tr>
<tr>
<td>• “Eleven staff from (ScHARR) volunteered to be a part of an intervention development focus group...participants who were unable to attend the meeting submitted suggestions via email.” [35]</td>
<td></td>
</tr>
<tr>
<td>• “Participants from all 3 interventions were asked to attend two structured meetings at their workplace to discuss and develop interventions...to develop workplace specific interventions as part of the participatory approach.” [18]</td>
<td></td>
</tr>
<tr>
<td>• “During the first meeting participants ‘brainstormed’ options to promote their specific intervention (active office, physical activity, or office ergonomics)...Between meetings participants were encouraged to think about specific strategies...At the second meeting, 2-3 weeks following the first meeting, participants shared their ideas and rated potential strategies in terms of feasibility and effectiveness.” [18]</td>
<td></td>
</tr>
<tr>
<td>• “An employee advisory board, which consisted of four to seven employees from all occupational sectors in the work site, was established in each work site and worked closely with the research team to design and implement intervention activities.” [39]</td>
<td></td>
</tr>
<tr>
<td><strong>Subtheme 1.3: workplace centered</strong>&lt;br&gt;This subtheme describes how meetings and workshops took place at the workplace rather than asking participants to come to the research team.</td>
<td></td>
</tr>
<tr>
<td>• “...workers attended a one-hour workshop (n = 10-15) held at the work site.” [37]</td>
<td></td>
</tr>
<tr>
<td>• “Participants from all 3 interventions were asked to attend two structured meetings at their workplace to discuss and develop interventions.” [18]</td>
<td></td>
</tr>
<tr>
<td><strong>Theme 2: manager or management</strong>&lt;br&gt;This theme describes how managers or management of the workplace acted as gatekeepers and were often used or asked to contribute in different formats or circumstances compared to workers.</td>
<td></td>
</tr>
<tr>
<td><strong>Subtheme 2.1: management creating study materials</strong>&lt;br&gt;This subtheme describes how management took an active participation in creating study materials, thereby bringing more relevance and familiarity to the developed posters and videos.</td>
<td></td>
</tr>
<tr>
<td>• “Management were involved in the design and development of promotional posters and exercise videos.” [38]</td>
<td></td>
</tr>
</tbody>
</table>
This subtheme describes how the managers acted as gatekeepers to the workplace and were used to pass information or study materials from the research team to the participants and workplace.

**Subtheme 2.3: management leading in data collection and setting up intervention components**

This subtheme describes how individuals with managerial roles assisted with data collection and study design and provided support throughout the study.

**Subtheme 2.4: invitation from the management or workplace**

This subtheme describes how invitations to participate in the development of the WHP study were sent via the workplace itself rather than through the research team.

**Theme 3: researcher facilitator**

This theme describes how the research team or lead research facilitator would communicate with team leaders or managers to help with the facilitation of the WHP intervention.

**Theme 4: workplace champions, health inspirers, team leaders, and health strategists**

This theme describes how the studies recruited participants from the workplace for elevated positions within the study to lead by example or provide more in-depth insights when codeveloping components of the WHP study.

**Subtheme 4.1: involved in intervention or implementation delivery or provision of support in increasing PA**

This subtheme describes how participants, managers, workplace champions, health inspirers, team leaders, or health strategists, and stakeholders were involved in the delivery of the intervention and supported WHP and increasing participants’ PA.

---

To evaluate the level of involvement and power shared with the end users and relevant stakeholders in the WHP studies, we used the ladder of citizen participation by Arnstein [27].

**Level of Participation (Ladder of Citizen Participation)**

In this scoping review, we aimed to evaluate the role and level of participation that the end users and stakeholders had in three aspects of the included studies: (1) the conception of the WHP study and intervention, (2) the data collection, and (3) the analysis and reporting of the included studies, as documented a priori in our published protocol [33].

**Conception**

When evaluating the included studies and the degree of participation and power shared with their participants within the conception and design of the WHP studies, 25% (2/8) of the studies were deemed as nonparticipation [38,40]. The studies by Blake et al [38] and Wahlström et al [40] were deemed to be nonparticipation as they stated the following:

---

**a**WHP: workplace health promotion.

**b**Not applicable.

**c**PA: physical activity.

**d**SB: sedentary behavior.
A total of 25% (2/8) of the included studies [16,37] were determined to be tokenistic in how they involved the participants and relevant stakeholders when conceiving and designing their WHP interventions. Gilson et al [37] described that, as researchers, they reached a consensus from the ideas generated by the workers regarding the design and number of strategies in the WHP study. While Tan et al [16] held workshops and participants self-selected their own intervention activities, the individually tailored strategies still needed to meet the prescribed 5 to 10 minutes of exercise breaks as designated by the research team. This level of involvement of the end users could be seen as “consultation” or “placation” on the ladder of citizen participation by Arnstein [27] as the researchers retained the right to judge the end users’ ideas and advice.

In total, 50% (4/8) of the studies were evaluated and determined to have provided citizen power, where the participants discussed and developed the WHP intervention and the planning and decision-making were shared between the researchers and the participants [18,35,36,39]. While the level of participation in 75% (3/4) of these studies was evaluated as “citizen power,” their level of participation was found to be at the lowest rung under “citizen power” of the ladder by Arnstein [27]. They were determined to be at the “partnership” rung as the researcher facilitated the discussions and communicated with team leaders and management concerning the implementation of the WHP study [18]. Kong et al [39] described the intervention as a community-based participation intervention where community (worksite) employees and researchers engaged as equals in a cooperative process; however, the power shared was not clearly documented, and the statement was merely cited as what a community-based participation intervention is, referencing Minkler and Wallerstein [54]. Kong et al [39] did state that they formed employee advisory boards (consisting of 4 to 7 employees), which were established in each worksite from all occupational sectors to “...work closely with the research team to design and implement intervention activities,” whereas Griffiths et al [36] described a “...compromise between the stakeholder input and researchers’ evidence-focused approach” and, furthermore, that they “...were able to include the participants views and opinions within the intervention design.” Finally, Mackenzie et al [35] held a brainstorming session with participants, who identified the strategies to reduce workplace sitting in the WHP study.

Data Collection and Data Analysis and Reporting

When originally planning and proposing our evaluation of the included studies, we had stated that we would evaluate whether participants and stakeholders were involved in data collection and the data analysis and reporting of the study. Upon extracting and charting the data from the included studies, we found that none of the studies included participants or stakeholders within the data collection, analysis, or reporting process. However, participants were included and took part in facilitating and implementing the intervention, and therefore, we retrospectively included a narrative synthesis on how participants were involved in the facilitation of the WHP study.

Facilitation of the Included WHP Studies

The larger components of the WHP interventions can be found in Table S3 in Multimedia Appendix 1. In this section, we provide a narrative synthesis on how end users and stakeholders assisted in facilitating the WHP studies. In this scoping review, the authors use the term “facilitation” to describe methods and processes that are active components of implementation, where individuals, end users, and stakeholders who are defined as facilitators enable and influence the implementation process [55].

One study provided no indication or description of end users assisting with the facilitation [36]. In total, 38% (3/8) of the studies enlisted the workplaces’ assistance in recruiting participants to the study [16,37,38]. Tan et al [16] enlisted their workplace coordinators to facilitate within-cluster recruitment to the study. The workplace coordinators were provided with resources by the research team to aid this recruitment, whereas managers were directly involved and facilitated recruitment by distributing recruitment emails to their administrative and clerical workers in the workplace study conducted by Gilson et al [37]. Management involvement was evident in the remaining 50% (4/8) of the WHP studies [18,35,38,40]. This management involvement ranged from emails [35] to introduce, support, and carry forward the intervention [35] to distributing questionnaires supporting and inspiring a healthy lifestyle, and management was asked to discuss and disseminate messages from the WHP campaigns at workplace meetings [40]. Blake et al [38] described the most involvement from managers, who delivered 30-minute orientations to team leaders on the WHP intervention to demonstrate senior management commitment, with the senior manager randomizing the 2 sites and being involved in the design and development of promotional posters and exercise videos. The management team also provided technical support, including the development and installation of a pop-up window system and web-based exercise log recording system [38]. Though not directly management involvement, the health resources department was responsible for inviting all employees from the worksite to participate in the study via their company intranet web page [38]. Whereas management involvement was not clearly described by Parry et al [18], other than “helped with implementation”.

The most frequent component of facilitation was the empowerment of employees in roles such as workplace champions. This term differed among the studies, with other titles used such as “team leaders” [16,38], and “health inspirers” [18]. Blake et al [38] asked their workplace champions to lead group exercise classes at allocated times, and the champions were involved in the provision of support for PA. This was similarly asked of the workplace champions in the WHP study by Mackenzie et al [35], who asked workplace champions to promote standing and walking meetings or sessions as well as incidental walking (talking and not emailing) and support lunchtime walks. This was further evident in the study conducted
by Wahlström et al [40], who asked the health inspirers to assist with the development of the communication campaign and support and inspire coworkers in attaining a healthy lifestyle. Unfortunately, it was not clear to what extent workplace champions facilitated the study carried out by Parry et al [18] other than that they “helped implementation,” whereas the study by Kong et al [39] was the only one to form an employee advisory board that worked closely to design and implement intervention activities with the research team. However, the extent of involvement of the end users, management, or relevant stakeholders in the implementation of all the intervention activities compared to the research team was not clearly indicated in the study [39], but it was stated that the employees voluntarily set up several exercise teams that monitored and attempted to improve their daily behaviors by recording their daily exercise and reminding each other to exercise daily. These exercise teams would meet regularly to exercise at an agreed time and place together [39].

The final forms of facilitation that we observed in the included studies had to do with the marketing of the interventions. Blake et al [38] formed an organizational committee made up of 2 team leaders and 2 human resources officers, who developed and implemented company policy on the internal marketing of the intervention. In the study by Wahlström et al [40], the organization’s communication department and the internal health strategist distributed all materials via posters, tabletops in meetings, and breakout spaces, along with posts on the workplace intranet.

Impact on PA and SB Levels
Of the 8 included studies, 1 (12%) [35] did not report PA as an outcome variable, but the remaining 7 (88%) studies did so in varying formats and across different intensities. Overall, 86% (6/7) of the studies did report an increase in PA, whereas Griffiths et al [36], who only measured and reported moderate to vigorous PA, found no change in their 2-week pilot study with 5 participants.

When investigating SB, 62% (5/8) of the included studies reported a measure of SB as an outcome variable. A total of 38% (3/5) of the studies reported a reduction in SB, whereas 12% (1/8) of the studies [40] reported no change in sitting time and 12% (1/8) [38] reported an increase in SB. Although Blake et al [38] reported an increase in SB, they did report a significantly lower increase in SB in the intervention group than in the control group.

When reporting the physical behavior outcomes, most studies (5/8, 62%) reported PA or SB during work hours, during which the intervention took place [18,35-37,40], whereas 25% (2/8) of the studies reported PA [16,38] and SB [38] as minutes or hours per week as the intervention could be continued outside of work hours. Furthermore, 12% (1/8) of the studies provided measures of PA daily (steps) and over a week (walking days per week) [39]. Finally, Parry et al [18] reported the physical behaviors as a percentage of accelerometer wear time for work hours and over the workday, and Mackenzie et al [35] provided the SB data during work hours in total and split into morning and afternoon.

The intervention outcomes reported in the included studies related to PA and SB are further outlined in Table S3 in Multimedia Appendix 1.

Reported Benefits of Taking a PR Approach
Many of the included studies (5/8, 62%) provided some context and evaluated how taking a PR approach benefited or hindered the research process and alluded to barriers that were experienced by the participants or researchers. In total, 25% (2/8) of the studies stated that the PR approach used was the key component to the success of the WHP intervention [18,38], whereas 38% (3/8) of the studies stated that taking a PR approach benefited the development of the intervention [18,35,36]. The authors noted that taking a PR approach and including the participants’ views and opinions ensured that the interventions were suitable to the workplace surrounding and, therefore, more likely to be acceptable and feasible for employees [18,35,36]. Mackenzie et al [35] hailed the PR approach for allowing for the development of a “real-world” pragmatic intervention. This point was expanded on further by Griffiths et al [36], who stated that the development group allowed for discussions and refinement of the initial intervention concept, which resulted in an agreed upon evidence-based intervention that was deemed feasible and acceptable by the stakeholders.

Some further benefits of the participatory approach were noted. These were awareness of the intervention [35] and increased communication about PA, suggesting that the information reached the workforce [40]. Managers’ behavior was noted as a motivating factor for employees to sit less and move more [40], and peer support was also identified as an important facilitator of intervention adoption [38]. Participants expressed that they had positive experiences about senior management as they offered great organizational support for the duration of the project and designed and developed the study’s promotional posters, which were regarded as high quality [38].

Reported Barriers of Taking a PR Approach
Despite the several benefits highlighted by the included studies, some barriers were also noted, such as lack of peer support, which reduced engagement [38]; lack of support from management and team leaders [18,38]; and lack of awareness of the intervention components (i.e., workplace champions and Twitter updates) rendering them unhelpful to the participants [35]. Furthermore, Parry et al [18] reported limited success in changing the organizational culture in workplaces even when management and participants were aware of the intervention options, leading to suggestions of stronger external support, such as guidelines.

Griffiths et al [36] presented and reported barriers when taking a PR approach. These were (1) inconsistent stakeholder attendance, which reduced stakeholder input in discussions at meetings; (2) conducting focus group meetings on the web rather than in person as this was thought to lose some of the possible interaction among stakeholders; (3) the fact that attempting to align the research with reality required compromise from both the stakeholders and research team, which did not guarantee resolution for all stakeholders; and (4)
the fact that having a small sample from a few different organizations in the development group and piloting of the intervention facilitated more in-depth discussion and greater input from participants but that whether the intervention translated to other organizations and, subsequently, its applicability and adherence were unknown [36].

A total of 38% (3/8) of the included studies did not evaluate how the PR approach may have benefited or hindered the WHP intervention or its outcomes [16,37,39].

**Discussion**

**Summary of Evidence: End User Engagement**

This scoping review addressed the research question and objectives, which were to identify and understand how previous WHP studies implemented PR with desk-based office workers. The systematic search retrieved 8 studies that met our inclusion criteria. Overall, there is evidence suggesting that taking a PR approach is beneficial to the development of a WHP intervention [18,35,36] and a key component of its success [18,38]. The main reasons cited in the included studies were that, by taking a PR approach, the participants’ opinions and lived experiences were included, which ensured that the intervention was suitable for the workplace and, ultimately, feasible and acceptable to the end users and relevant stakeholders [18,35,36].

A defining characteristic of PR is the degree of engagement and power shared with the end users during the different stages of the research process and decision-making more so than the methods and techniques used [8,56]. A key finding from this scoping review was that, among the 8 identified WHP studies that incorporated end users, the inclusion of the participants ended at the conception of the WHP intervention, with 2 (25%) studies not involving the end users or stakeholders in their conception and only in assisting with the implementation and facilitation of the study [38,40]. An important point to highlight is that, while we use the term conception of the WHP intervention, there was a distinct lack of engagement of end users in the conception of the overall study (ie, none of the included studies commenced due to an end user highlighting their occupational physical behaviors as a health priority and seeking an academic partner). In fact, all the included studies commenced with an academic institute or partner approaching an organization or organizations and recruiting participants and end users to the study before then developing the WHP intervention in collaboration with the end users and relevant stakeholders at varying levels of involvement and shared decision-making. In PR, the research problem should ideally originate from the community or end users [24]; however, in reality, we find that most PR projects are in fact initiated by researchers. Therefore, participatory processes are actively applied to shift ownership and control of the research process toward the end user community [57]. This leads to a more equitable and democratized decision-making process to facilitate the shift in knowledge leadership and support community ownership of the research process and developed intervention [57].

The degree of engagement and power that was shared with end users was evaluated to be low, with 25% (2/8) of the studies determined to be nonparticipation studies [38,40] and 25% (2/8) determined to be tokenistic in how the end users were involved in conceiving and designing the WHP studies [15-34,41-49]. This finding is similar to that of another identified review, which investigated studies that took a participatory action research approach to promote mental health and resilience in youth and adolescents [58]. Raanaas et al [58] identified a distinct lack of authentic involvement of the end users in their included studies, which is similar to our own included studies, with only 38% (3/8) of the studies determined to have provided citizen power at the lowest rung (‘partnership’) [18,35,36]. The similarities, despite the differences in population and outcome measures, can largely be attributed to the PR approach taken by the researchers and the power they shared with the end users.

The evaluation of the included studies via the ladder by Arnstien [27] placed an important emphasis on the level of involvement and shared decision-making, so the evaluation and final “labeling” of the included studies was not impacted by the lack of involvement in the later stages of the research process (data collection, analysis and reporting, and dissemination). However, the evaluation and, therefore, “labeling” was influenced by how the participants were involved in the development of the subsequent WHP intervention, with the included studies labeled as nonparticipatory stating that the research team developed the study without involvement of the end users or relevant stakeholders [38,40] and 25% (2/8) of the studies being labeled as tokenistic as evidenced by the research team tailoring the intervention components to the end users but retaining the overall decision-making power by (1) reaching a consensus from ideas generated by the workers [37] and (2) having a fixed prescribed exercise break of 5 to 10 minutes, which was instructed by the research team [16]. Therefore, 50% (4/8) of the studies demonstrated a lower level of involvement and shared decision-making, as opposed to the 50% (4/8) of the studies that were determined to provide citizen power, which noted a compromise or generated all ideas with the end users. Instead of excluding, tailoring, or retaining the right to draw the conclusions from a consensus of end users’ lived experience.

A goal in PR is that end users should “own” the research process [8], but in this review, we identified a glaring lack of end user involvement in the overall research process. When using the phrase “own the research,” the authors of this scoping review refer to the shift in knowledge leadership and influence from the researcher or academic institute to the community or workplace, which then has ownership and self-determination over the research process and subsequent intervention or interventions [57]. None of the included studies included engagement strategies such as creating a stakeholder board (beyond an advisory board) or council or involving end users as coresearchers in data analysis or research dissemination. These findings are bleak and identify an issue in this emerging field of PR in office-based WHP. One potential reason for the lack of authentic involvement observed in this review may be that 38% (3/8) of the studies were pilots [35-37]. However, authentic involvement was also absent in the remaining 62% (5/8) of the studies, and therefore, future research should strive...
to move beyond using PR to tailor and implement an intervention toward involving end users in the entire research process.

Summary of Evidence: Methods of Involving the End Users and Relevant Stakeholders

In this scoping review, we aimed to discuss how end users and relevant stakeholders were involved and how the participants’ involvement shaped the design of the WHP intervention. To address this aim, we conducted a thematic analysis and charted extracted data on the methods of PR (Table S2 in Multimedia Appendix 1). When conducting the thematic analysis to examine the role of the end users and the level of involvement in the coproduction of the WHP studies, we identified 4 primary themes. The four themes were (1) participant workshops, focus groups, meetings, or brainstorming sessions; (2) manager or management involvement; (3) researcher facilitator; and (4) workplace champions, health inspirers, team leaders, or health strategists. These themes can be compared to the set of themes identified in a previous broader review of end user involvement strategies [59].

The key findings from the thematic analysis in relation to the methods of involving end users were that the included studies invited end users to attend an initial focus group meeting where they were asked to provide their opinions and personal experiences in relation to designing or evaluating a WHP intervention. These focus group meetings were often cited as the conception of strategies that formed or were implemented alongside the WHP intervention. However, this was also often the end of the end user involvement, meaning that the focus groups may have been conducted in a tokenistic manner or tick-box exercise for an intervention that had already been largely planned. Focus groups in and of themselves do not constitute participatory engagement but are simply another form of data collection, which Arnstein [27] rates as tokenistic consultation. One of the included studies noted that compromises were made by the research team, suggesting that power was shared in the decision-making and development of the WHP intervention [36].

Another key finding from the thematic analysis and evaluation of the end users’ level of participation was the use of elevated positions that were created by the research team in the included studies, such as “team leaders,” “health inspirers,” and “workplace champions.” These elevated positions were created and asked end users to take on these positions and lead by example, support, and facilitate the delivery of different intervention components and in some cases provide more in-depth insights or participate in the codevelopment of intervention components [40]. However, barring the latter point, the use of elevated positions is not uncommon in nonparticipatory interventions, which may use members of a research team, intervention facilitators, or participants to assist in the facilitation of an intervention. Whereas a benefit of taking a PR approach is that end users in these elevated positions will remain engaged upon completion of the study and researcher involvement. Thus, these end users in elevated positions have an increased capacity to continue supporting behavior change within the community or workplace.

It is important to note that the lack of reporting on the PR approaches taken in the included studies restricts the conclusions that can be drawn as to whether the end users involved in the conception of the intervention were involved as the intervention facilitators. In some of the included studies (2/8, 25%), it was evident that other or additional employees and not solely the end users involved in intervention design were asked to be the intervention facilitators [38,39]. The potential impact of this is unknown regarding whether it would be a benefit or detriment to the facilitation of the intervention components. The reasons for this were not reported in the included studies but may be related to end user burden, to the end user involved in the PR approach not being an eligible participant, or to increased intervention facilitators being needed when the intervention is rolled out across a large workplace compared to the number of individuals (end users and stakeholders) involved during the commencement of the PR approach and conceptualization.

Summary of Evidence: Effect of PR WHP Interventions on PA and SB

An objective of this scoping review was to discuss the evaluation and outcomes of the included studies. Overall, the PR approach was shown to increases PA in 86% (6/7) and decrease SB in 60% (3/5) of the studies that measured these outcome variables, respectively. These changes were typically reported during occupational hours, which is when the interventions were generally prescribed. Only 12% (1/8) of the studies noted an increase in SB; however, the increase in the intervention group was significantly lower than the increase in the control group, suggesting that the PR approach may have blunted the increase in SB, whereas the final study that investigated SB noted no change. Caution should be taken when interpreting these findings as 38% (3/8) of the included studies [35-37] were pilot studies with no control comparator groups and small sample sizes (ie, not powered to detect a change in physical behavior [PA or SB]). However, the findings of this scoping review and of the included studies do suggest that PR should be investigated further as a research approach within health research due to the numerous benefits identified and the positive results observed overall.

Implications

The included studies and this scoping review provide insights into how PR has been incorporated in office-based PA WHP studies by reporting and evaluating the methods, the degree of engagement and power shared with end users, and the barriers to and benefits of taking a PR approach in the workplace setting. This review identified a lack of authentic and meaningful involvement of the end user, where they were asked to provide input to tailor an intervention but were not provided with the opportunity to “own” the research process. Future research should move beyond piloting PR in WHP interventions and start from the workplace to actively collaborate, identifying the end users’ priorities and developing or tailoring an intervention collaboratively with them to address their priorities. Further recommendations would be to (1) include end users in the published articles, presentations, and reports; and (2) build capacity within the workplace with a stakeholder board, management, or workplace champions that empower the end
users and create a sustainable WHP intervention after researcher involvement. (3) A key recommendation for future WHP studies or any study that wishes to take a PR approach would be to allow adequate time for the coconstruction or collaboration to occur and opportunity for decision-making to be shared and, therefore, for trust to be built [60] between the community or workplace and the research team. Some of the included studies (2/8, 25%) conducted a single “participatory” event, workshop, or focus group, which was often done in a tokenistic manner and to tailor a planned intervention from academia and not from the community. Ultimately, the act of conducting a single “participatory meeting” means that the researchers retain the power or ownership of the research process and decision-making of the research project and do not provide an equitable environment. Researchers should step outside the “normal” confines of research and share power and resources and develop capacity within their relevant community and end users to allow them to appropriately engage in all stages of the research process. This could be in the form of a larger number of participatory workshops (≥3 meetings), earlier engagement with end users to help conceptualize a suitable intervention that addresses their identified health priorities, and training and involvement in all aspects of the research process to increase the capacity of the included end users. These recommendations may facilitate the shift in knowledge leadership and encourage the sharing of power and movement toward end user ownership and self-determination over the research process and subsequent intervention or interventions.

Limitations
While this is the first scoping review to examine how PR is being incorporated in PA WHP studies in desk-based adults working in offices, in which we identified a multitude of common traits regarding how participants and the end users were involved, some limitations should be considered. When evaluating the degree of participation and power shared between participants and researchers in the WHP studies, we used the ladder of citizen participation by Arnstein [27]. Although the ladder by Arnstein [27] is well regarded, our interpretation of the included studies relied solely on self-reported end user involvement details in published literature and associated documents. Reporting standards for patient and public involvement are generally not well followed [61]; therefore, we could only evaluate the studies based on limited reporting, which may have missed details of engagement and shared power. This echoes the limitation of Frankena et al [62], who conducted a structured literature review and found it difficult to evaluate from written text whether the inclusion of their relevant end users (people with intellectual disabilities) was meaningful or tokenistic. Frankena et al [62] stated that more information was often needed regarding the process of inclusion.

Conclusions
In conclusion, the findings from this scoping review provide a foundation of evidence for how PR is currently being implemented in office-based PA WHP studies. We observed that the end user is currently only incorporated in the conception and implementation of the WHP studies and that, largely, the studies were tokenistic or nonparticipative, whereas 50% (4/8) of the studies were evaluated to provide citizen power in the conception of the interventions. Overall, a benefit was observed with positive improvements in PA and reductions in SB across the studies, which was largely attributed to taking a PR approach and involving the end users, which allowed for the design of a WHP intervention that was feasible and acceptable. Future studies should aim to move beyond a pilot and feasibility trial and collaborate with the workplaces, building capacity and empowering the workforce by providing citizen control and letting the end users “own” the research for a sustainable WHP intervention after researcher involvement.

Acknowledgments
This scoping review was funded and supported by the Health Research Institute, University of Limerick, through a PhD Studentship Award. The included sources of evidence were obtained via open-source electronic databases, and the articles were accessed through the University of Limerick library database or open-access publications. The authors would like to acknowledge Liz Dore (Faculty of Education and Health Sciences librarian at the University of Limerick) for the assistance in developing the search string and adapting it for the different databases searched. Generative artificial intelligence was not used in any portion of the manuscript writing or editing process.

Data Availability
Data sharing is not applicable to this paper as no data sets were generated or analyzed during this study.

Authors’ Contributions
All authors made substantial intellectual contributions to the conduct of the scoping review and the development of this manuscript. AJB and JS conceptualized the review approach, and AJB conducted the search, screening of the retrieved articles, and data extraction and drafted the initial manuscript. CKL conducted the search of the electronic databases independently, screened the retrieved articles, and performed checking of the data extraction. BPC and AED contributed to the conceptualization of the review. CKL, BPC, AED, and JS contributed to the writing and editing of the manuscript.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Participatory research methods, intervention components, and reported outcomes of the interventions in the included studies in relation to physical activity and sedentary behavior and completed PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.

References


47. Bree RT, Gallagher G. Using Microsoft Excel to code and thematically analyse qualitative data: a simple, cost-effective approach. All Ireland J High Educ 2016;8(2):28111-28114 [FREE Full text]


Abbreviations

PA: physical activity
PICO: population, intervention, comparison, and outcome
PR: participatory research
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SB: sedentary behavior
WHP: workplace health promotion

©Aidan John Buffey, Christina Kate Langley, Brian P Carson, Alan E Donnelly, Jon Salsberg. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 19.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Impact of Healthy Lifestyle Factors on Life Expectancy and Lifetime Health Care Expenditure: Nationwide Cohort Study

Wei-Cheng Lo1,2, PhD; Tsuey-Hwa Hu3, MSc; Cheng-Yu Shih4, MSc; Hsien-Ho Lin4, MD, ScD; Jing-Shiang Hwang3, PhD

1Master Program in Applied Epidemiology, College of Public Health, Taipei Medical University, New Taipei, Taiwan
2School of Public Health, College of Public Health, Taipei Medical University, Taipei, Taiwan
3Institute of Statistical Science, Academia Sinica, Taipei, Taiwan
4Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Wei-Cheng Lo, PhD
Master Program in Applied Epidemiology
College of Public Health
Taipei Medical University
No. 301, Yuantong Rd., Zhonghe Dist.
New Taipei, 235
Taiwan
Phone: 886 266202589 ext 16011
Email: nicholaslo@tmu.edu.tw

Abstract

Background: The association between lifestyle risk factors and the risk of mortality and chronic diseases has been established, while limited research has explored the impact of healthy lifestyle factors on lifetime health care expenditure using longitudinal individual data.

Objective: We aimed to determine the individual and combined effects of 5 healthy lifestyle factors on life expectancy and lifetime health care expenditure in Taiwan.

Methods: Using data from the National Health Interview Survey cohort, 5 healthy lifestyle behaviors were defined and analyzed: nonsmoking, avoiding excessive alcohol consumption, engaging in sufficient physical activity, ensuring sufficient fruit and vegetable intake, and maintaining a normal weight. We used a rolling extrapolation algorithm that incorporated inverse probability of treatment weighting to estimate the life expectancy and lifetime health care expenditure of the study populations with and without healthy lifestyle factors.

Results: A total of 19,893 participants aged ≥30 (mean age 48.8, SD 13.4) years were included, with 3815 deaths recorded during a median follow-up period of 15.6 years. The life expectancy and per capita estimated lifetime health care expenditures for the overall study population were 35.32 years and US $58,560, respectively. Multivariable-adjusted hazard ratios for all-cause mortality in participants adhering to all 5 healthy lifestyle factors, compared with those adhering to none, were 0.37 (95% CI 0.27-0.49). We found significant increases in life expectancy for nonsmokers (2.31 years; 95% CI 0.04-5.13; P=0.03), those with sufficient physical activity (1.85 years; 95% CI 0.25-4.34; P=0.02), and those with adequate fruit and vegetable intake (3.25 years; 95% CI 1.29-6.81; P=0.01). In addition, nonsmokers experienced a significant reduction in annual health care expenditure (~9.78%; 95% CI ~46.53% to ~1.45%; P=0.03), as did individuals maintaining optimal body weight (~18.36%; 95% CI ~29.66% to ~8.57%; P=0.01). Overall, participants adhering to all 5 healthy lifestyle behaviors exhibited a life gain of 7.13 years (95% CI 4.43%-5.71%; P=0.02) compared with those adhering to one or none, with a life expectancy of 29.19 years (95% CI 25.45-33.62). Furthermore, individuals adopting all 5 healthy lifestyle factors experienced an average annual health care expenditure reduction of 28.12% (95% CI 4.43%-57.61%; P=0.02) compared with those adopting one or none.

Conclusions: Adopting a healthy lifestyle is associated with a longer life expectancy and a reduction of health care expenditure in Taiwanese adults. This contributes to a more comprehensive understanding of the impact of healthy lifestyle factors on the overall health and economic burden.

(JMIR Public Health Surveill 2024;10:e57045) doi:10.2196/57045

https://publichealth.jmir.org/2024/10/e57045

JMIR Public Health Surveill 2024 | vol. 10 | e57045 | p.836

(page number not for citation purposes)
KEYWORDS
healthy lifestyle factors; life expectancy; lifetime health care expenditure; rolling extrapolation algorithm; nationwide cohort study; nonsmoking; moderate alcohol consumption; physical activity; fruit and vegetable intake; optimal body weight

Introduction
Background
Accumulative epidemiologic studies have investigated the relationship between lifestyle factors and the risk of mortality and chronic diseases. These studies have explored both the individual and combined effects of various risk factors on health. A meta-analysis including 531,804 participants from 17 countries indicated that smoking, inactivity, poor diet quality, and heavy alcohol consumption contributed to approximately 60% of all premature deaths [1]. A healthy lifestyle, which is characterized by regular physical activity, normal weight, nonsmoking behavior, moderate alcohol consumption, and healthy diet intake, is associated with an increase in life expectancy (LE) [2-11], with studies reporting increases of 6.6 years for men and 8.1 years for women in Singapore [2]; 8.8 years for men and 8.1 years for women in China [3]; 10.3 years for men and 8.3 years for women in Japan [4]; 12.2 years for men and 14.0 years for women in the United States [5]; 16.8 years for men and 18.9 years for women in Canada [6]; and 7.4 to 15.7 years for 3 European cohorts (RCPH, ESTHER, and Tromsø) [7], indicating that the beneficial effects of a healthy lifestyle may vary across populations and countries. On the other hand, these healthy lifestyle factors have been found to be associated with an increase in disease-free LE as well [12-17].

In recent decades, a significant increase has been noted in the average LE worldwide. However, whether such improvements translate into an overall reduction in health care expenditures remains uncertain. Studies on the impact of a healthy lifestyle on health care costs have focused primarily on the effects of individual risk behavior [18-28]. A regression analysis revealed that a 10% relative decrease in the prevalence of smoking in the United States would lead to a reduction of approximately US $63 billion in health care costs in the year following this decrease [21]. Another study reported that excessive alcohol consumption results in an economic burden of US $223.5 billion in the United States, with 72.2% of this amount being attributable to lost productivity, 11% attributable to health care expenses, 9.4% attributable to criminal justice costs, and 7.5% attributable to other effects [22]. A systematic review recruited 19 cost-of-illness studies on obesity, identifying substantial economic burdens across countries, with direct medical costs ranging from 0.7% to 17.8% of health system expenditure and total costs ranging from 0.05% to 2.42% of the country’s gross domestic product [23]. Studies comparing the effects of obesity, overweight, smoking, and problem drinking on health care use and costs in the United States have indicated that obesity and overweight are associated with the highest health care costs [26,27]. However, these studies had cross-sectional designs and certain analytical limitations, such as problems related to the unit of observation used at the ecological level or the indirect estimation of risk attributable approach. To date, no study has investigated the impact of both individual and combined healthy lifestyle factors on lifetime health care expenditure based on longitudinal individual data.

Objectives
The objective of this study was to investigate the impact of both individual and combined healthy lifestyle factors on LE and lifetime health care expenditure in a contemporary population. Using a nationally representative cohort with over 19,000 participants, we have proposed a rolling extrapolation algorithm that incorporated inverse probability of treatment weighting (IPTW), which can adjust for the effects of potential confounders, to estimate the lifetime survival function for the study cohort. The lifetime survival functions were used to calculate the LE and lifetime health care expenditure for study populations with and without healthy lifestyle factors.

Methods
Study Design, Settings, and Population
A nationwide longitudinal cohort study was performed. We used data from the Taiwan National Health Interview Survey (NHIS) to define study cohorts with or without healthy lifestyle factors. Then, a rolling extrapolation algorithm was introduced to estimate the lifetime survival function of study cohorts [29]. The LE of a study cohort was calculated according to the corresponding extrapolated lifetime survival function. In addition, the total lifetime health care expenditures of a study cohort were estimated by integrating the product of the lifetime survival function and a medical cost function, which was calculated using reimbursement data obtained from Taiwan’s National Health Insurance Research Database (NHIRD) [29]. Finally, we estimated the effect of each healthy lifestyle factor (individual effects) and the combination of the individual effects (combined effects) on LE and lifetime health care expenditure. The study sample comprised individuals who participated in the Taiwan NHIS in 2001 and 2005, with household response rates of 91.1% and 80.6%, respectively. The NHIS is a cross-sectional survey, which adopted a multistage stratified sampling scheme to obtain a nationally representative sample of the Taiwanese population. For the NHIS, baseline data regarding individuals’ sociodemographic and behavioral factors were collected through in-person interviews. Details of the design and sampling scheme of the NHIS have been reported previously [30]. In this study, 27,631 participants aged ≥30 years were included. We excluded participants who refused to link their NHIS data to NHIRD records (n=5680); who had a missing value for smoking, alcohol consumption, BMI, physical activity, or healthy diet (n=1709); who had an extreme value of lifestyle factor (BMI >60 kg/m² or metabolic equivalent of task [MET] >10,000 min/wk; n=229); or who had missing covariate data (n=327). Finally, 19,893 participants were included in the analysis at baseline and were followed till the end of the study period (December 31, 2020) or death, which was confirmed using data from the vital registry of Taiwan (Figure 1). The reporting of this study adhered to the STROBE
Assessment of Healthy Lifestyle Factors

The following 5 lifestyle factors were analyzed in this study: smoking, alcohol consumption, BMI, leisure-time physical activity, and healthy diet (sufficient intake of fruit and vegetables). Information on lifestyle-related factors was collected from the NHIS data, which were collected through in-person interviews conducted at baseline. We categorized smoking status at baseline into 2 groups: never-smokers (never smoked at all or smoked <100 cigarettes in their lifetime) and current or former smokers (smoked more than 100 cigarettes in their lifetime). We defined alcohol consumption at baseline into 2 groups as well: infrequent or nonconsumers (those consuming alcohol less than once a week) and excess consumers (those consuming alcohol at least once a week). Anthropometric data were self-reported. We used the BMI to categorize baseline body weight status into 2 groups: optimal body weight (BMI: 18.5-25 kg/m²) and nonoptimal body weight (underweight: BMI <18.5 kg/m² and overweight: BMI ≥25 kg/m²). To evaluate leisure-time physical activity, we used individuals’ responses to the following NHIS question: Did you participate in any leisure-time physical activity over the past 2 weeks? Respondents could list more than one type of physical activity; they also reported the exercise frequency (ie, the number of times they engaged in the activity in the previous 2 weeks) and duration (ie, how long they exercised). MET intensity levels were assigned to each activity (eg, Tai Chi, walking, jogging, and swimming) on the basis of a relevant study [31]. Total weekly leisure-time physical activity (MET min/wk) was calculated by multiplying the frequency (times per week), duration (minutes), and the MET values of each activity. The physical activity was classified into 2 groups: insufficiently...
active (<500 MET min/wk) and sufficiently active (≥500 MET min/wk). Diet was assessed using the NHIS data, which were collected using a simplified food frequency questionnaire. We defined a healthy diet as having sufficient fruit and vegetable intake: ≥4 servings of fruits and vegetables per week.

**Covariates**

We obtained data regarding the individuals’ sociodemographic characteristics, lifestyle factors, and disease history from the NHIS database. The NHIS data were obtained through in-person interviews and a structured questionnaire survey [30]. The potential confounders and covariates were taken into consideration, including enrollment year, age at baseline, sex, ethnicity, education level, marital status, religion, monthly household income, and medical history of hypertension, hyperlipidemia or raised cholesterol, cardiovascular disease, diabetes mellitus, cancer, chronic lung diseases (asthma or chronic obstructive pulmonary disease), and chronic kidney diseases (Table S1 in Multimedia Appendix 1).

**Statistical Analysis**

Although the participants of the study cohort were followed with a median duration of 15.6 years, a high proportion of participants were still alive by the end of the follow-up. We therefore applied a rolling extrapolation algorithm to obtain the lifetime survival function for the study cohort [29]. The rolling extrapolation algorithm is one of the new methods to improve the accuracy of survival extrapolation using external evidence [32]. The extrapolation method was successfully applied in many studies for estimating loss of LE of cohorts with specific health conditions [33-36] compared with age- and sex-matched reference cohorts generated from life tables of the general population. In an observational study, the estimates of LE differences among comparative study cohorts (eg, with and without specific healthy lifestyle factors) may be biased due to the effects of some confounders in the samples. In this study, we incorporated IPTW to the rolling extrapolation algorithm to reduce the potential confounding effects on the estimate of LE difference between 2 cohorts.

We described the details of the modified rolling extrapolation algorithm in the following 4 major steps. First, each individual in a study cohort was given a weight equal to the inverse probability of the individual being in the cohort, which is estimated by a multinomial logistic regression model with measured confounding covariates; that is, each study cohort is inflated with the weights of the participants to form a pseudocohort in which confounders are equally distributed across the pseudocohorts [37]. The adjusted Kaplan-Meier estimator was applied to the weighted survival data of individuals in the study cohort to obtain the survival function, denoted by \( S(t) \), for the cohort [38]. We may call \( S(t) \) the confounder-adjusted survival function of the study cohort. Second, Monte Carlo methods were used to generate survival times of referents whose age and sex matched with participants in the study cohort using life tables of the general population. The same weights were assigned for the referents matched with the participants in the study cohort. The confounder-adjusted lifetime survival function of the reference population was obtained by applying an adjusted Kaplan-Meier estimator on the weighted survival times of the generated referents and denoted by \( S_r(t) \). We renamed \( S_r(t) \) as \( S_r(t) \times e^{\delta} > S(t) \) for use as a relevant predictor in extrapolating the survival function of the study cohort when \( S_r(t) > S(t) \) during the follow-up period. If \( S_r(t) \leq S(t) \), a proper hazard value \( \delta \) was subtracted from the reference population to ensure the predictor \( S_r(t) = S_r(t) \times e^{\delta} > S(t) \) [39]. The relative survival function, \( \bar{S}(t) \), is then within 0 and 1. Third, the logit transformation of the relative survival would straighten the curve \( \text{logit}[W(t)] \), which is relatively easy to extrapolate. Fourth, we used restricted cubic spline models to fit \( \text{logit}[W(t)] \) for the observed period \( t = 1, 2, ..., F \) months. The fitted curve was used to predict the \( \text{logit}[W(t)] \) 1 month ahead. The predicted \( \text{logit}[W(F+1)] \) was usually quite accurate because of the approximate linearity property of \( \text{logit}[W(t)] \) and treated as an observation at month \( F + 1 \). We then repeated the extrapolation procedures by rolling the same-length observation periods 1 month ahead, \( t = 2, 3, ..., F + 1 \), and refit the restricted cubic spline models for the updated observation periods to predict the value of \( \text{logit}[W(F+2)] \). By repeatedly performing the abovementioned procedures of extrapolating \( \text{logit}[W(t)] \) month-by-month to a time \( L \) beyond which all subjects of the cohort died, we could then invert transformation of the extrapolated \( \text{logit}[W(t)] \) to obtain an estimate of relative survival function \( \bar{S}(t) \) and lifetime survival function \( S(t) \) of the cohort. The validation analysis of the rolling extrapolation algorithm was performed for study cohorts with and without risk lifestyle factors (Table S2 in Multimedia Appendix 1). The confounder-adjusted LE of each study cohort was obtained by summing the extrapolated confounder-adjusted lifetime survival function, \( \bar{S}(t) \). The loss or gain of LE of a cohort is estimated by the difference of confounder-adjusted LE between the cohort and the comparative cohort.

With the extrapolated lifetime survival function \( S(t) \) of a cohort, we can further estimate the lifetime health expenditures of the cohort using the formula \( C(t) = \text{monthly average of the participant’s weighted expenditures in the subset} \), where \( C(t) \) is monthly average expenditure of living participants at time \( t \) in the study cohort [29]. We retrieved all direct health care costs (including inpatient and outpatient care) from the claims database of the National Health Insurance for the participants of each study cohort. Each participant’s costs were also assigned the same weight as the participant in a cohort. All the reimbursement costs, including treatment, examination, and procedures associated with disease outcomes of the participants in a cohort, were weighted and summed up to calculate the monthly average of the cohort. To extrapolate the monthly mean cost function beyond follow-up, we make one reasonable assumption that medical expenditures start increasing from K months before death [40]. Specifically, we classify the participants alive in each month into subsets of participants who died in the current month, within the next K months, and who lived more than K months. We then calculate mean costs in each month by a weighted average of the K+2 mean costs of these subsets in that month, where the weights were estimated by the extrapolated risk of death and the mean cost of each subset was estimated by the average of the participants’ weighted expenditures in the subset [29]. Estimates of LE, difference in LE between study cohorts, lifetime health
expenditures, annual average health expenditure, and their SEs and 95% CIs were obtained using the R package iSQoL2, which can be downloaded [41].

To estimate the magnitude of hazards associated with an unhealthy lifestyle, we used Cox proportional hazard models to calculate adjusted hazard ratios and corresponding 95% CIs for all-cause mortality associated with each individual health risk factor and a composite of health risk factors (0, 1, 2, 3, 4, or 5). The models were adjusted for the effects of the aforementioned covariates. In addition, dose-response relationships between the lifestyle factors and LE or health care expenditure were analyzed to evaluate the robustness of our findings. We classified smoking status into 4 categories based on smoking history and recent smoking behavior: never-smokers, former smokers, current smokers of 1 to 19 cigarettes per day, and current smokers of ≥20 cigarettes per day. Alcohol consumption was classified into 4 categories based on drinking frequency: never drinkers, infrequent drinkers (those consuming alcohol less than once a week), regular drinkers (those consuming alcohol more than once a week but not to the point of getting drunk), and heavy drinkers (those consuming alcohol more than once a week and typically to the point of getting drunk). Leisure-time physical activity was divided into 4 categories: sedentary (0 MET min/wk), insufficiently active (1-499 MET min/wk), active (500-1499 MET min/wk), and highly active (≥1500 MET min/wk). We classified healthy diet into 4 categories based on the frequency of fruit and vegetable intake: ≤1 serving, 2 to 3 servings, 4 to 6 servings, and ≥7 servings per week. Body weight status was divided into 4 categories: underweight (BMI <18.5 kg/m²), normal weight (18.5≤BMI<25 kg/m²), overweight (25≤BMI<30 kg/m²), and obese (BMI ≥30 kg/m²) [42]. Moreover, a series of sensitivity and validation analyses were conducted. We used the World Health Organization (WHO) Asian BMI risk cut points for BMI classification as part of our sensitivity analysis [43]. To investigate the effect of the number of healthy lifestyle factors on health, we initially designated individuals who adhered to none or one of the 5 healthy lifestyle factors as the reference group in our primary analysis. As a supplementary sensitivity analysis, we also designated individuals who embraced none of the healthy lifestyle factors as an alternative reference group. Furthermore, a subgroup analysis was performed based on different numbers of healthy lifestyle factors, including those with 2 factors, 3 factors, or 4 factors, to present estimations of LE and health care expenditure associated with the combinations of healthy lifestyle factors within each subgroup. In addition, a validation analysis was performed to assess the accuracy of survival function extrapolation (further details are provided in Table S2 in Multimedia Appendix 1). The R package programs mnet::multinom and survival::coxph were used for the analysis. Statistical significance was set at a 2-tailed value of P<.05.

Ethical Considerations

In accordance with Article 5 of the Human Subjects Research Act of the Ministry of Health and Welfare, Republic of China (Taiwan) [44] and the guidelines titled “Management Principles for the Application of Health and Welfare Data” announced by the Ministry of Health and Welfare, Taiwan on September 7, 2017, pursuant to Wei-Bu-Tong-Zi (#1062560770)” [45], the data used in this analysis are subject to the management and regulation by the Health and Welfare Data Science Center, Taiwan [46]. Individual health data managed therein (including NHIS, NHIRD, and Death Registry data) undergo deidentification processing in accordance with the Personal Data Protection Act.

This study was approved by the Taipei Medical University Joint Institutional Review Board (TMU-JIRB-N202009047). The Health and Welfare Data Science Center reviewed the study protocol and institutional review board approval and granted access to confidential data. The requirement for written informed consent was therefore waived because anonymous and deidentified information was used for analysis.

Results

Baseline Characteristics

Of 19,893 participants included in the analysis, 10,311 (51.8%) were men; the mean age was 48.8 (SD 13.4) years. During 317,116 person-years of follow-up, 3815 deaths were recorded (1403 women and 2412 men). Table 1 presents the baseline characteristics of study participants by the number of healthy lifestyle behaviors they adhered to. Individuals adhering to a higher number of healthy lifestyle behaviors were more likely to be women, had received higher levels of education, and had higher household income. At baseline, 68.2% of participants were not current or former smokers, 83.2% did not consume excessive amounts of alcohol, 33.8% were sufficiently active, 89.2% had sufficient fruit and vegetable intake, and 62.1% had a normal weight.

https://publichealth.jmir.org/2024/1/e57045

JMIR Public Health Survell 2024 | vol. 10 | e57045 | p.840

(page number not for citation purposes)
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>0 or 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Healthy lifestyle factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (y), mean (SD; range)</td>
<td>45.7 (11.6; 30-90)</td>
<td>47.6 (13.0; 30-96)</td>
<td>49.1 (13.6; 30-98)</td>
<td>48.8 (13.7; 30-98)</td>
<td>50.9 (13.3; 30-93)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1121 (91.89)</td>
<td>2484 (84.81)</td>
<td>3297 (57.36)</td>
<td>2526 (35.17)</td>
<td>883 (31.39)</td>
</tr>
<tr>
<td>Female</td>
<td>99 (8.11)</td>
<td>445 (15.19)</td>
<td>2451 (42.64)</td>
<td>4657 (64.83)</td>
<td>1930 (68.61)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than elementary school</td>
<td>384 (31.48)</td>
<td>947 (32.33)</td>
<td>2114 (36.78)</td>
<td>2351 (32.73)</td>
<td>796 (28.30)</td>
</tr>
<tr>
<td>High school</td>
<td>718 (58.85)</td>
<td>1551 (52.95)</td>
<td>2542 (44.22)</td>
<td>3158 (43.96)</td>
<td>1183 (42.05)</td>
</tr>
<tr>
<td>College or above</td>
<td>118 (9.67)</td>
<td>431 (14.71)</td>
<td>1092 (19.00)</td>
<td>1674 (23.31)</td>
<td>834 (29.65)</td>
</tr>
<tr>
<td><strong>Marriage, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td>856 (70.16)</td>
<td>2157 (73.64)</td>
<td>4335 (75.42)</td>
<td>5594 (77.88)</td>
<td>2192 (77.92)</td>
</tr>
<tr>
<td>Never married</td>
<td>171 (14.02)</td>
<td>323 (11.03)</td>
<td>528 (9.19)</td>
<td>506 (7.04)</td>
<td>199 (7.07)</td>
</tr>
<tr>
<td>Others a</td>
<td>193 (15.82)</td>
<td>449 (15.33)</td>
<td>885 (15.40)</td>
<td>1083 (15.08)</td>
<td>422 (15.00)</td>
</tr>
<tr>
<td><strong>Ethnic group, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minnan</td>
<td>798 (65.41)</td>
<td>1973 (67.36)</td>
<td>4134 (71.92)</td>
<td>5207 (72.49)</td>
<td>2028 (72.09)</td>
</tr>
<tr>
<td>Hakka</td>
<td>135 (11.07)</td>
<td>378 (12.91)</td>
<td>741 (12.89)</td>
<td>940 (13.09)</td>
<td>387 (13.76)</td>
</tr>
<tr>
<td>Mainland Chinese</td>
<td>94 (7.70)</td>
<td>302 (10.31)</td>
<td>537 (9.34)</td>
<td>800 (11.14)</td>
<td>341 (12.12)</td>
</tr>
<tr>
<td>Aborigines</td>
<td>182 (14.92)</td>
<td>252 (8.60)</td>
<td>270 (4.70)</td>
<td>151 (2.10)</td>
<td>29 (1.03)</td>
</tr>
<tr>
<td>Others</td>
<td>11 (0.90)</td>
<td>24 (0.82)</td>
<td>66 (1.15)</td>
<td>85 (1.18)</td>
<td>28 (1.00)</td>
</tr>
<tr>
<td><strong>Household income b, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>307 (25.16)</td>
<td>672 (22.94)</td>
<td>1228 (21.36)</td>
<td>1422 (19.80)</td>
<td>573 (20.37)</td>
</tr>
<tr>
<td>Median</td>
<td>688 (56.39)</td>
<td>1575 (53.77)</td>
<td>3059 (53.22)</td>
<td>3702 (51.54)</td>
<td>1304 (46.36)</td>
</tr>
<tr>
<td>High</td>
<td>225 (18.44)</td>
<td>682 (23.28)</td>
<td>1461 (25.42)</td>
<td>2059 (28.66)</td>
<td>936 (33.27)</td>
</tr>
<tr>
<td><strong>Comorbid diseases, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>240 (19.67)</td>
<td>574 (19.60)</td>
<td>1252 (21.78)</td>
<td>1308 (18.21)</td>
<td>527 (18.73)</td>
</tr>
<tr>
<td>Hyperlipidemia or raised cholesterol</td>
<td>237 (19.43)</td>
<td>530 (18.09)</td>
<td>1083 (18.84)</td>
<td>1162 (16.18)</td>
<td>478 (16.99)</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>109 (8.93)</td>
<td>282 (9.63)</td>
<td>778 (13.54)</td>
<td>927 (12.91)</td>
<td>393 (13.97)</td>
</tr>
<tr>
<td>Stroke</td>
<td>27 (2.21)</td>
<td>87 (2.97)</td>
<td>214 (3.72)</td>
<td>203 (2.83)</td>
<td>84 (2.99)</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>84 (6.89)</td>
<td>237 (8.09)</td>
<td>481 (8.37)</td>
<td>545 (7.59)</td>
<td>210 (7.47)</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>57 (4.67)</td>
<td>162 (5.53)</td>
<td>294 (5.11)</td>
<td>313 (4.36)</td>
<td>103 (3.66)</td>
</tr>
<tr>
<td>Cancer</td>
<td>29 (2.38)</td>
<td>102 (3.48)</td>
<td>290 (5.05)</td>
<td>471 (6.56)</td>
<td>228 (8.11)</td>
</tr>
<tr>
<td>Asthma or chronic obstructive pulmonary disease</td>
<td>108 (8.85)</td>
<td>261 (8.91)</td>
<td>484 (8.42)</td>
<td>539 (7.50)</td>
<td>169 (6.01)</td>
</tr>
<tr>
<td><strong>Lifestyle factors, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never-smokers</td>
<td>59 (4.84)</td>
<td>617 (21.07)</td>
<td>3571 (62.13)</td>
<td>6515 (90.70)</td>
<td>2813 (100)</td>
</tr>
<tr>
<td>Current or former smokers</td>
<td>1161 (95.16)</td>
<td>2312 (78.93)</td>
<td>2177 (37.87)</td>
<td>668 (9.30)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrequent or nonconsumers</td>
<td>213 (17.46)</td>
<td>1541 (52.61)</td>
<td>5002 (87.02)</td>
<td>6978 (97.15)</td>
<td>2813 (100)</td>
</tr>
<tr>
<td>Excess consumers</td>
<td>1007 (82.54)</td>
<td>1388 (47.39)</td>
<td>746 (12.98)</td>
<td>205 (2.85)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
Healthy Lifestyle Factors and Mortality Risk

Each individual healthy lifestyle factor was significantly associated with all-cause mortality (Figure S1 and Table S3 in Multimedia Appendix 1). The risk of mortality decreased with increasing adherence to healthy lifestyle behaviors (in terms of the number of behaviors adopted by the individuals). Individuals adhering to all 5 healthy lifestyle behaviors exhibited a hazard ratio of 0.37 (95% CI 0.27-0.49) for all-cause mortality, compared with those with 0 healthy factors, after adjusting for potential confounding factors.

Impact on LE

Using the rolling extrapolation method, we found that participants with mean ages of 48.8 years who did not smoke had a life gain of 2.31 (95% CI 0.04-5.13) additional years. Not drinking excessively was marginally associated with an additional gain in life years by 1.62 (95% CI -0.15 to 4.59) years. Sufficient physical activity was associated with an additional 1.85 (95% CI 0.25-4.34) life years gain. Sufficient fruit and vegetable intake was associated with a life gain of 3.25 (95% CI 1.29-6.81) years. Finally, having an optimal weight was associated with a nonsignificant increment in life years (0.72 years; 95% CI -1.21 to 2.14) compared with being underweight or overweight. LE increased with the increasing adherence to healthy lifestyle behaviors (Table 2). After covariate adjustment, we note that individuals who adhere to 4 and 5 healthy lifestyle behaviors have a life span increase of 6.58 (95% CI 1.13-10.76) and 7.13 (95% CI 1.33-11.11) years, respectively, compared with LE (29.19 years) for individuals who do not adhere to more than one healthy lifestyle behavior.
Table 2. Life expectancy and years of life gained of study cohorts with and without healthy lifestyle factors.

<table>
<thead>
<tr>
<th>Lifestyle factors</th>
<th>Values, n</th>
<th>Life expectancy (95% CI)</th>
<th>Years of life gained (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cigarette smoking</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current or former smokers</td>
<td>6318</td>
<td>32.13 (29.47 to 34.12)</td>
<td>Reference</td>
</tr>
<tr>
<td>Never-smokers</td>
<td>13,575</td>
<td>34.44 (33.17 to 35.81)</td>
<td>2.31 (0.04 to 5.13)</td>
</tr>
<tr>
<td><strong>Alcohol consumption</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess consumers</td>
<td>3346</td>
<td>33.43 (30.88 to 35.27)</td>
<td>Reference</td>
</tr>
<tr>
<td>Infrequent or nonconsumers</td>
<td>16,547</td>
<td>35.05 (34.17 to 35.86)</td>
<td>1.62 (−0.15 to 4.59)</td>
</tr>
<tr>
<td><strong>Leisure-time physical activity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insufficient active</td>
<td>13,174</td>
<td>33.50 (31.89 to 34.29)</td>
<td>Reference</td>
</tr>
<tr>
<td>Sufficient active</td>
<td>6719</td>
<td>35.35 (33.12 to 36.97)</td>
<td>1.85 (0.25 to 4.34)</td>
</tr>
<tr>
<td><strong>Fruit and vegetable intake</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low or insufficient intake</td>
<td>2154</td>
<td>31.47 (28.72 to 33.56)</td>
<td>Reference</td>
</tr>
<tr>
<td>Sufficient intake</td>
<td>17,739</td>
<td>34.72 (34.33 to 36.16)</td>
<td>3.25 (1.29 to 6.81)</td>
</tr>
<tr>
<td><strong>BMI group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonoptimal body weight(^a)</td>
<td>7540</td>
<td>32.86 (31.58 to 33.85)</td>
<td>Reference</td>
</tr>
<tr>
<td>Optimal body weight</td>
<td>12,353</td>
<td>33.58 (32.14 to 34.22)</td>
<td>0.72 (−1.21 to 2.14)</td>
</tr>
<tr>
<td><strong>Number of low-risk lifestyle factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 or 1</td>
<td>1220</td>
<td>29.19 (25.45 to 33.62)</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>2929</td>
<td>30.69 (26.25 to 32.76)</td>
<td>1.50 (−5.28 to 6.36)</td>
</tr>
<tr>
<td>3</td>
<td>5748</td>
<td>34.42 (33.82 to 36.04)</td>
<td>5.24 (0.98 to 9.70)</td>
</tr>
<tr>
<td>4</td>
<td>7183</td>
<td>35.77 (34.27 to 36.83)</td>
<td>6.58 (1.13 to 10.76)</td>
</tr>
<tr>
<td>5</td>
<td>2813</td>
<td>36.32 (32.76 to 38.42)</td>
<td>7.13 (1.33 to 11.11)</td>
</tr>
</tbody>
</table>

\(^a\)Nonoptimal weight: underweight (BMI <18.5 kg/m\(^2\)) and overweight (BMI ≥25 kg/m\(^2\)).

**Impact on Lifetime Health Care Expenditure**

We evaluated the effects of healthy lifestyle factors on lifetime health care expenditure and found that nonsmoking and a healthy weight were significantly associated with reduced lifetime health care expenditure (Table 3). Per capita estimated lifetime health care expenditures were US $60,395 for smokers and US $58,821 for nonsmokers. Notably, individuals with nonoptimal body weight had a significantly higher per capita estimated lifetime health care expenditure compared with those with optimal body weight (US $62,474 vs US $53,336, respectively). The percentage change in annual health care expenditure was calculated by determining the difference in per capita annual health care expenditure between populations with and without healthy lifestyle factors, which was then divided by the annual average health care expenditure per capita of the overall population. We found that never-smokers had a significant reduction in the percentage change in annual health care expenditure compared with current or former smokers (−9.78%; 95% CI −46.53% to −1.45%). Furthermore, individuals with optimal body weight had notably lower average annual health care expenditures compared with those with nonoptimal body weight, with an 18.36% reduction of annual health care expenditures per capita (95% CI 8.57%-29.66%). Overall, individuals who adopted 2, 3, 4, or 5 healthy lifestyle factors demonstrated a dose-response reduction in the percentage change in annual health care expenditure per capita by 16.01%, 18.90%, 18.61%, and 28.12%, respectively, in comparison to those who adopted 0 or only 1 healthy lifestyle factor.
Table 3. Lifetime health care expenditure and percentage change in annual health care expenditure of study cohorts with and without healthy lifestyle factors.

<table>
<thead>
<tr>
<th>Lifestyle factors</th>
<th>Values, n</th>
<th>Lifetime health care expenditure, US $ (95% CI)</th>
<th>Percentage change in annual health care expenditure, (%; 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cigarette smoking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current or former smokers</td>
<td>6318</td>
<td>60,395 (56,018 to 68,054)</td>
<td>Reference</td>
</tr>
<tr>
<td>Never-smokers</td>
<td>13,575</td>
<td>58,821 (38,595 to 63,525)</td>
<td>−9.78 (−46.53 to −1.45)</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excess consumers</td>
<td>3346</td>
<td>57,849 (51,399 to 65,353)</td>
<td>Reference</td>
</tr>
<tr>
<td>Infrequent or nonconsumers</td>
<td>16,547</td>
<td>57,737 (54,723 to 60,531)</td>
<td>−5.00 (−19.14 to 6.11)</td>
</tr>
<tr>
<td>Leisure-time physical activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insufficient active</td>
<td>13,174</td>
<td>56,733 (53,081 to 59,840)</td>
<td>Reference</td>
</tr>
<tr>
<td>Sufficient active</td>
<td>6719</td>
<td>60,828 (56,098 to 67,518)</td>
<td>1.59 (−7.46 to 11.14)</td>
</tr>
<tr>
<td>Fruit and vegetable intake</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low or insufficient intake</td>
<td>2154</td>
<td>52,815 (46,193 to 59,625)</td>
<td>Reference</td>
</tr>
<tr>
<td>Sufficient intake</td>
<td>17,739</td>
<td>60,217 (58,016 to 63,905)</td>
<td>3.24 (−11.98 to 17.51)</td>
</tr>
<tr>
<td>BMI group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonoptimal body weightc</td>
<td>7540</td>
<td>62,474 (56,549 to 68,844)</td>
<td>Reference</td>
</tr>
<tr>
<td>Optimal body weight</td>
<td>12,353</td>
<td>53,336 (49,394 to 55,884)</td>
<td>−18.36 (−29.66 to −8.57)</td>
</tr>
<tr>
<td>Number of low-risk lifestyle factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 or 1</td>
<td>1220</td>
<td>58,715 (43,229 to 78,470)</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>2929</td>
<td>53,414 (45,675 to 58,903)</td>
<td>−16.01 (−46.06 to 12.47)</td>
</tr>
<tr>
<td>3</td>
<td>5748</td>
<td>58,246 (40,663 to 63,068)</td>
<td>−18.90 (−46.53 to −1.70)</td>
</tr>
<tr>
<td>4</td>
<td>7183</td>
<td>60,667 (40,685 to 67,255)</td>
<td>−18.61 (−47.41 to −1.84)</td>
</tr>
<tr>
<td>5</td>
<td>2813</td>
<td>55,785 (47,896 to 64,346)</td>
<td>−28.12 (−57.61 to −4.43)</td>
</tr>
</tbody>
</table>

aUS $1=New Taiwan $30.65.
bThe percentage change in annual health care expenditure was calculated as the difference in per capita annual health care expenditure between populations with and without healthy lifestyle factors divided by the annual average health care expenditure per capita for the overall population.
cNonoptimal weight: underweight (BMI <18.5 kg/m$^2$) and overweight (BMI ≥25 kg/m$^2$).

Dose-Response Analysis
The analysis of the dose-response effect of each lifestyle factor on LE revealed that nonsmoking, infrequent alcohol consumption (less than once a week), high physical activity levels, healthy diet intake (intake of ≥4 servings of fruits and vegetables per week), and overweight (25≤BMI<30 kg/m$^2$) were associated with a long LE (Figure 2). Regarding lifetime health care expenditure, individuals who were heavy smokers (≥20 cigarettes per day), were regular and heavy drinkers, had a sedentary lifestyle (0 MET min/wk), and were overweight and obese (BMI ≥25 kg/m$^2$) are more likely to have higher average annual health care expenditures (Figure 3). Regarding healthy diet intake, it is important to note that the exposure group with the longest LE may not necessarily be the same as the group with the lowest health care expenses. In addition, we found that overweight, when compared with normal weight, was associated with an additional 0.69 (95% CI −0.54 to 2.26) life years. However, this increased LE comes with the caveat of incurring additional health care expenses, with an increase of percentage change in annual health care expenditure (21.1%; 95% CI 9.98%–31.17%). Alternatively, we depicted the correlation between LE and lifetime health care expenditures for populations with and without healthy lifestyle factors using a scatter plot. We hypothesized a linear relationship, with a steeper slope indicating higher annual average health care expenditures. Figure S2 in Multimedia Appendix 1 illustrates that, in comparison to populations with 0 or only 1 healthy lifestyle factor, individuals with all 5 healthy lifestyle factors not only experience a longer LE but also have lower annual average health care costs. When examining individual factors, generally, each of the healthy lifestyle factors is associated with a longer LE. However, we found that smoking and nonoptimal body weight are linked to considerably higher annual health care expenses, along with reduced LE.
Figure 2. Estimated life expectancy (LE) and years of life gained according to levels of individual lifestyle risk factors. Red color indicates the estimated value was significantly higher or lower than reference group. MET: metabolic equivalent of task. *: cigarettes per day; #: minutes per week; &: servings of fruit and vegetable intake per week.
Sensitivity and Subgroup Analysis

In addition to using the globally common BMI classification criteria, we conducted a sensitivity analysis by incorporating the WHO Asian BMI risk cut points for BMI classification. The results of the sensitivity analysis align with our primary findings. Compared with individuals with a normal weight (18.5≤BMI<23 kg/m²), those classified as overweight (23≤BMI<27.5 kg/m²) demonstrated significantly improved LE, with an additional gain of 3.36 years (95% CI 1.88-5.76; Figure S3 in Multimedia Appendix 1). Moreover, we also designated individuals who adhered to none of the healthy lifestyle factors as an alternative reference group. A discernible dose-response relationship was identified. Overall, individuals who adopted 1, 2, 3, 4, or all 5 healthy lifestyle factors presented LE gains of 4.37, 4.73, 10.22, 11.85, and 12.39 years, respectively, in comparison to those who adopted none of these 5 healthy lifestyle factors (Table S4 in Multimedia Appendix 1). Furthermore, under stratification by the number of healthy lifestyle factors, noticeable heterogeneity emerged across various combinations of these factors.
Discussion

Principal Findings

Using data from a nationally representative cohort, we found that 5 healthy lifestyle factors were associated with significant gains in LE and reductions in health care expenditure for individuals aged ≥30 years in Taiwan. Our findings revealed that adherence to all 5 healthy lifestyle behaviors was associated with a 7.13-year increase in LE and a 28.12% reduction in the percentage of annual health care expenditure per capita compared with the values noted in individuals adhering to 0 or 1 healthy lifestyle behavior. Furthermore, the lifestyle factors did not exert equal effects on LE. Smoking and diet were the most significant risk factors for a short LE, with current and former smokers having an estimated reduction of 2.31 years in LE compared to never-smokers. In addition, individuals consuming sufficient amounts of fruits and vegetables had a life gain of 3.25 years compared with the LE of individuals with insufficient intake. Individuals regularly engaging in sufficient physical activity had a life gain of 1.85 years compared with the LE of individuals with a sedentary or inactive lifestyle. Individuals who were underweight (BMI <18.5 kg/m\(^2\)) had a shorter LE than did those with a normal weight (18.5≤BMI<25 kg/m\(^2\)); in contrast, individuals who were overweight (25≤BMI<30 kg/m\(^2\)) had a longer LE than did those with a normal weight. Heavy drinkers who consumed alcohol more than once a week and typically to the point of getting drunk had a shorter LE than those consuming no alcohol.

Our findings regarding the individual and combined effects of healthy lifestyle factors on LE are generally consistent with those of other studies. Adhering to a healthy lifestyle is associated with a significant increase in estimated LE by 6.6 to 18.9 years across countries [2-8]. Differences in the definitions of healthy lifestyles and the characteristics of study populations might have contributed to the differences in LE increase among the countries. Additional sources of such heterogeneity should be investigated. For example, living environment has also been linked to LEs [47-49]. Residents of low-income areas in wealthy cities, such as New York and San Francisco, had significantly longer LEs than did those of low-income areas in poor cities, such as Gary, Indiana, and Detroit [49]. Further studies are required to better understand how living environment, as well as socioeconomic status and difference between Eastern and Western populations, influence the health benefits of lifestyle factors across countries.

The global average LE has undergone a significant increase in recent decades. However, the LE in Taiwan, one of the wealthiest nations in East and Southeast Asia, falls behind in comparison to other high-income countries in this region. In 2019, LE at birth in Taiwan was 80.3 years, which was lower than that of Singapore (84.9 years), Japan (84.8 years), and South Korea (82.9 years) [50,51]. This discrepancy can be attributed to Taiwan having a health care system that prioritizes disease treatment over prevention. Our revisit analysis of the 2017 Global Burden of Disease study revealed that 44.3% of the total disease burden in Taiwan was associated with modifiable risk factors, whereas the leading risk factors in Japan, Singapore, and South Korea had lower fractions of attributable disability-adjusted life year rates of 38.0%, 35.2%, and 40.3%, respectively [52]. The successful reduction in risk exposure and unhealthy behaviors in these countries has led to significant improvements in health outcomes. Our findings clearly indicate that a healthy lifestyle can significantly extend a person’s life span. Our estimates can serve not only as materials for promoting health among the public but also as key references for the government when it seeks to develop, implement, and evaluate intervention programs.

Our study showed that nonsmoking, avoiding excessive alcohol consumption, healthy diet intake, normal weight, and regular physical activity were all associated with a low risk of premature mortality. Nonsmoking and healthy diet intake considerably affected the LE of Taiwanese adults. The harmful effects of smoking are well known. Our findings highlight the importance of maintaining nonsmoking behavior. Healthy diet intake has been associated with low risks of morbidity and mortality associated with several noncommunicable diseases [53]. We did not investigate the effects of the intake of various foods and nutrients; nevertheless, the quantity of fruit and vegetable intake indicates the healthiness level of a diet, at least partially. A significant dose-response relationship was noted between fruit and vegetable intake and an increase in LE. Nevertheless, this relationship has not been found in the burden of health care expenditure. One plausible interpretation is that individuals who adhere to a healthy diet may also exhibit a heightened health consciousness, potentially resulting in excessive use of health care resources, particularly within the environment of Taiwan’s National Health Insurance system. Regarding BMI, individuals who were normal weight or overweight had longer LEs than those who had underweight or obesity. These findings are generally consistent with those of a population-based cohort study, which reported a J-shaped association between BMI and mortality, with the nadir at a BMI of 25 kg/m\(^2\) [54]. However, the issue of the obesity paradox necessitates thoughtful consideration when using BMI as an indicator of obesity because of its inherent limitations and the potential for misclassification resulting from arbitrary BMI categorization [55-58]. In our
study, consistent results were observed when applying both the WHO global and Asian BMI classification systems. Importantly, it should be noted that overweight and obesity exerted significant negative effects, as indicated by the noteworthy excess health care expenditure associated with these conditions. Furthermore, our findings revealed a positive correlation between the level of physical activity and an increase in LE, yet this did not translate into a reduction in health care expenditure. Although physical activity has been associated with improvements in overall LE, engaging in regular exercises (particularly those involving outdoor activities) may increase the risk of accidental injury and thereby lead to additional health care expenditure. Our findings revealed that infrequent alcohol consumption (less than once a week) exerted a favorable effect on LE, as well as health care expenditure. Although the cardioprotective effects of moderate alcohol consumption have been reported in large cohort studies [59], current guidelines do not endorse initiating alcohol consumption solely for the purpose of preventing cardiovascular disease. Notably, a systematic review and meta-analysis reported that low or moderate alcohol consumption was not significantly associated with a reduced risk of all-cause mortality [60]. Furthermore, relying solely on alcohol consumption frequency rather than the actual intake amount (in grams) in this study may result in an imprecise exposure assessment. To address this limitation, we used NHIS 2009 and 2013 data, which began collecting information on alcohol intake amounts, to investigate the relationship between consumption frequency and intake amounts. Our additional analysis revealed a significant correlation, with a Spearman correlation coefficient of 0.986 (Table S6 in Multimedia Appendix 1). However, further research is warranted to refine exposure assessment methods and comprehensively explore the impact of alcohol consumption on LE and health care costs. In addition, there could be notable diversity in the effects on LE and lifetime health expenditures across various combinations of healthy lifestyle factors. Our subgroup analysis reveals that the combination of nonsmoking and adequate fruit and vegetable consumption stands out as the most crucial factors associated with increased LE while maintaining an optimal body weight emerges as pivotal for reducing lifetime health expenditures. However, the constrained sample sizes within the strata might have impeded the detection of discrepancies owing to diminished statistical power. Thus, further investigation with larger samples is imperative to elucidate these potential heterogeneities more definitively.

To the best of our knowledge, this study is the first to quantify the lifetime health care expenditure associated with the individual and combined effects of healthy lifestyle factors. Our results showed that healthy lifestyle factors were associated with reduced annual average health care expenditure. Previous studies exploring the effects of risk behaviors on health care costs have revealed that smoking, excessive alcohol consumption, obesity, and overweight may lead to significant economic burdens [20-22,24-27]. However, less research to date has explored the impact of individual and combined healthy lifestyle factors on lifetime health care expenditure using longitudinal individual data. Adherence to a low-risk lifestyle can not only prolong an individual’s life span but also significantly postpone disability [7,12-17,61]. Li et al [14] reported that individuals aged 50 years who had a low-risk lifestyle had a longer LE free of major chronic diseases, with an approximate gain of 7.6 years for men and 10 years for women in the United States and 6.9 years for men and 9.4 years for women in the United Kingdom population [13]. Furthermore, a recent Chinese study unveiled that adults aged 40 years, who embraced 5 low-risk lifestyle factors, could attain an additional 6.3 and 4.2 years of life without cardiovascular diseases, cancer, and chronic respiratory diseases for men and women, respectively, compared with those with 0 to 1 low-risk lifestyle factors [12]. These findings indicate that postponing disability can benefit society by ensuring that health care costs associated with disability are not incurred until individuals reach advanced ages and enabling individuals to work longer. On the other hand, our investigation has revealed that specific lifestyle factors exhibit disparate effects on both LE and health care expenditures. For example, our findings show that individuals who are overweight (25≤BMI<30 kg/m²) experience an extension of 0.69 (95% CI 0.54 to 2.26) years in their LE compared with those with a normal weight (18.5≤BMI<25 kg/m²). However, this increased longevity is accompanied by elevated health care costs, resulting in a 21.1% (95% CI 9.98%-31.17%) increase in annual health care expenditure. By carefully considering the trade-offs between the influences of health behaviors on LE and health care expenditure, our outcomes may yield valuable insights for the comprehensive evaluation of cost-effectiveness in primary prevention initiatives. Nonetheless, these findings also emphasize the urgency of future research endeavors to thoroughly investigate the impact of health factors on both nonfatal and fatal disease burdens. For instance, it would be imperative to examine the effects of health factors on composite metrics such as disability-adjusted life years, quality-adjusted life years, or health-adjusted LE at an individual level to gain a comprehensive understanding.

As nations with aging populations face the associated socioeconomic consequences, they must ensure that the health status of older individuals is maintained. Our findings, derived from the Taiwanese population, indicate that significant health and economic gains can be achieved if individuals adopt low-risk behaviors. Taiwan’s Health Promotion Administration has sought to increase the health promotion awareness and the preventive health capabilities of the population. Their efforts have concentrated on the screening of cancer, regulation of tobacco hazards, and prevention of chronic and noncommunicable diseases. These measures may not only reduce disease risks and extend LE but also alleviate economic burdens. For example, consider the tobacco tax revenue in Taiwan for the year 2022, which amounted to approximately 0.97 billion [62]. Approximately 50% (0.49 billion) of this revenue is allocated to subsidize the National Health Insurance expenditure. According to the estimates in this study, combined with a smoking prevalence of 14% in Taiwan [63], we can estimate that approximately an additional 0.36 billion in health care expenditure was contributed by smokers. However, it is important to note that our study does not account for indirect economic impacts, including indirect health care expenditures and productivity losses. In recent decades, Taiwan’s health care expenditures have steadily increased, and the National Health
Insurance system currently grapples with a budget deficit. To improve the health care system in Taiwan, adjustments in health promotion investment strategies are necessary. The findings from this study hold significant implications for the direction of health promotion strategies and policy formulation. Previous literature suggests the effectiveness of fiscal policies in improving health, such as the implementation of taxes on tobacco, alcohol, and sugar-sweetened beverages and foods. Nevertheless, it is not limited to taxing substances with health risks. Adjustments may include increasing investments in health education and preventive health services, as well as enhancing food and physical environments to support healthy behaviors.

Limitations

Our study has several strengths, including the use of a large sample size derived from national representative survey data, a prospective nature, and an almost complete follow-up. By linking these data with National Health Insurance claims and vital registry data and integrating an extrapolated lifetime survival function, we were able to comprehensively evaluate the impacts of health lifestyle factors on LE and health care expenditure in Taiwan. In addition, we integrated IPTW in our survival function extrapolation algorithm to minimize the effects of potential confounders. This innovative approach enhances the robustness and validity of our findings.

This study has several limitations that should be acknowledged. First, all analyzed lifestyle factors were assessed at baseline; therefore, lifestyle changes over time were not accounted for because no repeated measurement data were available. Second, the lifestyle factors data were self-reported, which may have led to misclassification. Third, although we adjusted the statistical models for a wide range of potential confounders, the likelihood of residual and unmeasured confounding effects cannot be ruled out. Fourth, we might have underestimated the proportion of individuals with an unhealthy lifestyle because individuals with poor health may be less likely to participate in surveys or may even die before participating. Fifth, we did not account for other healthy lifestyle factors that may independently affect LE, including factors such as lack of comprehensive dietary intake frequency information (eg, red or processed meat consumption), which prevented us from adequately exploring the impact of various healthy dietary patterns on overall health and economic burden, as well as factors such as sleep quality and duration, and levels of stress. Finally, lifetime health care expenditures were estimated on the basis of reimbursement data obtained from Taiwan’s NHIRD. Therefore, the estimates did not account for out-of-pocket expenses or costs associated with a loss of productivity. Therefore, from a societal perspective, lifetime health care expenditures might have been underestimated in this study.

Conclusions

Our findings revealed dose-response relationships between healthy lifestyle factors and increased LEs and reduced annual health care expenditures. The findings may have implications for primary prevention and resource allocation. They highlight the need for coordinated multisectorial efforts that target modifiable lifestyle factors to reduce the overall burden of disease. By prioritizing interventions that promote healthy behaviors and mitigate risk factors, health outcomes can be improved and health care costs can be minimized.

Acknowledgments

This work was supported by the Ministry of Science and Technology (grants MOST109-2314-B-038-142-MY2 and MOST110-2627-M-006-010), the National Science and Technology Council (grant NSTC112-2314-B-038-073-MY3), and Higher Education Sprout Project of the Ministry of Education in Taiwan (grant NTU-CC-109L891501–109L891507). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The data used in this analysis are not owned by the authors and therefore cannot be shared publicly. To acquire access to the individual-level data, interested researchers must complete an application form, submit a research proposal, and provide documentation of institutional review board approval to the Health and Welfare Data Science Center, Taiwan [46]. The Center will review these materials and grant access to eligible researchers who meet the criteria for accessing confidential data. It should be noted that authorized researchers will be granted access to the data under the same conditions and procedures as the authors.

Authors’ Contributions

WCL, HHL, and JSH conceptualized the study design and objectives. THH and CYS curated the data and did the statistical analysis. WCL, THH, and JSH accessed and verified the data. WCL wrote the original draft of the manuscript. All authors critically revised the manuscript and had final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional methods and results.

[DOCX File , 389 KB - publichealth_v10i1e57045_app1.docx ]


41. iSOQL2 package for Windows, iSOQL2. URL: https://sites.stat.sinica.edu.tw/isql/ [accessed 2023-02-10]


**Abbreviations**

- IPTW: inverse probability of treatment weighting
LE: life expectancy
MET: metabolic equivalent of task
NHIRD: National Health Insurance Research Database
NHIS: National Health Interview Survey
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WHO: World Health Organization

Please cite as:
Lo WC, Hu TH, Shih CY, Lin HH, Hwang JS
Impact of Healthy Lifestyle Factors on Life Expectancy and Lifetime Health Care Expenditure: Nationwide Cohort Study
JMIR Public Health Surveill 2024;10:e57045
URL: https://publichealth.jmir.org/2024/1/e57045
doi:10.2196/57045
PMID:39018094

©Wei-Cheng Lo, Tsuey-Hwa Hu, Cheng-Yu Shih, Hsien-Ho Lin, Jing-Shiang Hwang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Advancing Colorectal Cancer Detection With Blood-Based Tests: Qualitative Study and Discrete Choice Experiment to Elicit Population Preferences

Clarence Ong¹, BSocSci; Alex R Cook¹, PhD; Ker-Kan Tan¹,2, MBBS, PhD; Yi Wang¹, PhD

¹Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
²Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Yi Wang, PhD
Saw Swee Hock School of Public Health
National University of Singapore and National University Health System
12 Science Drive 2
Singapore, 117549
Singapore
Phone: 65 65164988
Email: ephwyi@nus.edu.sg

Abstract

Background: Colorectal cancer (CRC) is the second most deadly form of cancer, inducing an estimated 1.9 million incidence cases and 0.9 million deaths worldwide in 2020. Despite the availability of screening tests, their uptake remains suboptimal. However, blood-based tests that look for signs of cancer-specific markers in the body are increasingly available as an alternative for more invasive tests for cancer. Compared with existing tests, the benefits of blood-based tests for CRC include not needing pretest preparation, stool handling, and dietary or medication restrictions.

Objective: This study aims to explore the population’s preferences for CRC screening tests, with a focus on blood-based tests, and investigate the factors influencing test uptake.

Methods: We used a mixed methods approach, combining semistructured interviews and a discrete choice experiment (DCE) survey. Interviews were analyzed using thematic analysis to identify salient attributes for CRC screening tests. These attributes informed the design of the DCE survey. The DCE data were analyzed using mixed logit and mixed-mixed multinomial logit models.

Results: Qualitative findings from 30 participants revealed that participants preferred blood-based tests due to their perceived low risk, minimal pain, and ease of sample collection. However, concerns about the test’s lower accuracy were also expressed. The DCE survey was completed by 1189 participants. In the mixed logit model, participants demonstrated a stronger preference for blood-based tests over a 2-day stool-based test. The mixed-mixed multinomial logit model identified 2 classes, strong supporters and weak supporters, for CRC screening. Weak supporters, but not strong supporters, had a higher preference for blood-based tests. Women, ethnic Chinese, and people aged 40 to 60 years were more likely to be weak supporters. Both models highlighted the high influence of cost and test sensitivity on participants’ preferences. Transitioning from a 2-day stool-based test to a blood-based test, assuming a national screening program at a base price of Singapore $5 (US $3.75), was estimated to have the potential to increase the relative uptake by 5.9% (95% CI 3.6%-8.2%).

Conclusions: These findings contribute to our understanding of CRC screening preferences and provide insights into the factors driving test uptake. This study highlights the perceived advantages of blood-based tests and identifies areas of concern regarding their accuracy. Further research is needed to determine the actual increase in uptake rate when blood-based tests are made available.

(JMIR Public Health Surveill 2024;10:e53200) doi:10.2196/53200

KEYWORDS

blood-based test; colorectal cancer screening; mixed methods research; discrete choice experiment
**Introduction**

**Background**

Colorectal cancer (CRC) is the second most deadly form of cancer, inducing an estimated 1.9 million incidence cases and 0.9 million deaths worldwide in 2020 [1]. Regular screening for CRC, through methods such as colonoscopy and fecal immunochemical testing (FIT), helps detect CRC earlier, reduces the incidence and mortality of CRC, and brings about cost savings compared with not undergoing any screening [2]. Despite the availability of screening as a preventive intervention for the early detection of CRC, screening rates are suboptimal, even within high-income countries [3-5]. Factors that hamper screening include not just individual characteristics but characteristics of the screening tests as well [6].

The emergence of blood-based, early detection tests for cancer has the potential not only for detecting multiple cancers but also for improving patient compliance and acceptance [7]. The discovery of circulating and cell-free tumor DNA in the blood has ushered in new possibilities for the blood-based detection of CRC as well [8]. Epi proColon—a SEPT9 DNA methylation assay—remains to be the only US Food and Drug Administration–approved test [9]. However, there are at least 5 other blood-based tests in various stages of development, with tests ranging from CRC-specific tests to multi-cancer early detection tests. Some candidate analytic targets include cell-free DNA, methylated circulating tumor DNA, and a combination of methylated DNA and proteins [10]. Compared with existing tests, the benefits of blood-based tests for CRC include not needing pretest preparation, stool handling, and dietary or medication restrictions [9].

Challenges to the implementation of blood-based tests for screening include lower specificity relative to one-time FIT [11] and inferior sensitivity compared with next-generation FIT-DNA tests [12]. As a result, blood-based tests for CRC are not recommended for the general population in the health guidelines of the United States, Europe, China, and Singapore [13-16]. While there exist several clinical disadvantages to blood-based tests, it may serve as an alternative for patients refusing screening by colonoscopy or patients self-excluded from stool-based tests due to bleeding conditions such as hemorrhoids radiation proctitis [17]. In fact, the SEPT9 test was found to be more effective and cost-effective compared with no screening [18]. By making screening easier, blood-based tests have the potential to improve uptake if the benefits outweigh the downside of this screening modality [19]. Studies are required to understand how the population will make trade-offs between different procedures and their attributes.

In a randomized controlled trial (RCT) involving average-risk adults that offered blood-based tests and FIT in a clinical setting, higher screening participation rates were observed in the blood-based test arm [20]. The blood-based test was also found to be effective in increasing screening rates among medically underserved populations [19]. However, another RCT reported no statistically significant improvement in the uptake among the population familiar with FIT if a blood-based test was offered upfront as an option [21]. Conversely, studies offering the blood-based test as a rescue option for those who declined colonoscopy and stool-based tests showed an increase in participant rates [21-23].

**Objective**

Building on the existing literature, at least 4 questions are deserving of further investigation. First, what is the population’s inclination toward blood-based tests if the accuracy of blood-based tests can be improved to satisfactory levels akin to the FIT-DNA test? This insight can help assess the potential value of further investment in the test and inform the design of a target product profile [24,25]. Second, what is the general population’s preference for using the blood-based test in routine CRC screening? Results from RCTs may not be generalizable to the general population given the differences in the characteristics between the study participants and the general population. Third, considering heterogeneous preferences for blood-based tests, can we profile the population based on their preferences? Such profiling efforts can inform the crafting of targeted screening programs to cater to the heterogeneous preferences across different groups. Fourth, many preference studies were done in Western countries and very few were done in Asia [26,27]. Cultural and social norms could influence decision-making and outcomes. Studies understanding the acceptance of blood-based CRC tests in Asia are needed.

Acknowledging these gaps, we have designed a mixed methods study to delve into the population’s preference for blood-based testing modalities in Singapore, a multiracial Asian society, and to understand their decision-making process when choosing between blood-based tests and other existing screening methods. A discrete choice experiment (DCE) was used to construct hypothetical scenarios (eg, higher accuracy for the blood-based test). Furthermore, we intend to undertake subgroup analyses to examine potential variations in the preference for blood-based tests within distinct segments of the population, as highlighted by the mixed results of the RCTs. Our investigation will also delve into whether specific screening methods, such as the blood-based test, might positively impact participation rates, particularly in subpopulations identified with lower anticipated adherence based on prevailing screening recommendations.

**Methods**

This is a mixed methods study with interviews and a survey that incorporated a DCE. The methods for the qualitative and quantitative components will first be outlined, and subsequently, the qualitative and quantitative results will be presented.

**Ethical Considerations**

The study was approved by the National Healthcare Group Domain Specific Review Board (2021/00753) before data collection. Participants of the DCE were from a web-based cohort, and their participation in research was approved by the National University of Singapore (NUS) institutional review board (NUS-IRB: H-18-011).
Qualitative Component

Participant and Sampling

The recruitment and interview of participants for the qualitative component took place between December 2021 and March 2022. Convenience sampling was undertaken to include the Singapore population aged ≥40 years with varied engagement with CRC screening services. Recruitment was conducted via the NUS social media platforms and its email blast services, as well as other participant recruitment channels and word of mouth. Interested potential participants contacted the researchers, who verified their eligibility before taking informed consent.

Conducting the Interview

The interviews took place either on the web via a videoconferencing application or in a quiet room within the NUS that was convenient for recording. The interviews adopted a semistructured format using a topic guide. Each interview lasted approximately 30 to 45 minutes and was conducted in English.

Analysis of Interviews

The interviews were transcribed verbatim, and the data were analyzed using thematic analysis. A preliminary codebook with emerging themes of relevance from the first 5 transcripts was developed upon full familiarization of the transcripts. A deductive and semantic approach was used in the clustering of codes into metacodes and categories of interest. The coding framework was subsequently applied to the remaining transcripts. The identified themes were also further reviewed to ensure their usefulness and accuracy in representing the data.

Quantitative Component

Discrete Choice Experiment

DCEs have increasingly become a popular method for investigating and eliciting patient and population preferences for health care [28]. The method is based on consumer choice theory [29], which posits that respondents make choices between hypothetical products or scenarios comprising of ≥2 alternatives based on the importance they place on the characteristics of these attributes. In a DCE, a product or scenario is described with a fixed number of attributes with varying combinations of levels. Per this paradigm, in choosing the ideal product or scenario, the respondent evaluates the overall desirability of the alternatives and makes trade-offs among the attributes. From the respondents’ choice, their preferences are indirectly revealed, determining the attributes that drive the respondents’ preferences as well as the way variations of the attributes and levels may affect the respective preferences [30].

Selection of Attributes and Levels

The selection of attributes and levels must be relevant to the policy process and the study population, while being consistent with the random utility theoretical foundation of DCEs [31]. An initial set of attributes and levels for the DCE was based on a scoping review of the existing literature, which yielded 13 attributes. During the aforementioned qualitative interviews, participants were asked to rank 3 attributes that they valued the most, and the weighted preferences of all participants helped shortlist the final attribute list for the DCE. Following that, a quantitative survey with eligible health care professionals (n=11), who had at least 1-year working experience with patients with CRC or RC screening, was conducted to ensure the validity of the selected attributes and their corresponding levels. After these iterative processes, six attributes were identified and ultimately used in the DCE: (1) procedure, (2) pain level, (3) sensitivity, (4) recommendation, (5) out-of-pocket cost, and (6) risk of test. Each attribute was assigned various levels based on the best information available. The blood-based test was one level of the procedure attribute.

To optimize the choice sets, a pilot study was conducted with 12 participants. Adjustments were then made to the text for the attributes and levels to improve clarity for the participants. The final set of 6 attributes and levels is presented in Textbox 1. The total number of appearances and selections for each attribute and level may be found in Multimedia Appendix 1.
**Attributes and levels in the discrete choice experiment.**

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Pain level</th>
<th>Sensitivity</th>
<th>Recommendation</th>
<th>Cost</th>
<th>Risk of test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colonoscopy</td>
<td>No pain</td>
<td>100%</td>
<td>Health Promotion Board</td>
<td>Singapore $0</td>
<td>No risk</td>
</tr>
<tr>
<td>Computed tomography colonography</td>
<td>Mild pain</td>
<td>95%</td>
<td>Doctors</td>
<td>Singapore $5 (approximately US $3.75)</td>
<td>1% risk of adverse event</td>
</tr>
<tr>
<td>Stool-based (2 days)</td>
<td></td>
<td>80%</td>
<td>Family or friends</td>
<td>Singapore $30 (approximately US $22.50)</td>
<td></td>
</tr>
<tr>
<td>Stool-based (1 day)</td>
<td></td>
<td>60%</td>
<td>Neither</td>
<td>Singapore $400 (approximately US $300)</td>
<td></td>
</tr>
<tr>
<td>Blood-based</td>
<td></td>
<td></td>
<td></td>
<td>Singapore $1000 (approximately US $750)</td>
<td></td>
</tr>
</tbody>
</table>

**Experimental Design**

The DCE questionnaire was designed using Sawtooth Lighthouse Studio (version 9.13.2), and a 2-stage design was used. For each task, participants first selected the preferred choice from 2 test profiles and were then asked to choose between taking the test or opting out of it in real life. Correspondingly, the parameter labeled “Opt-Out” represents the utility associated with declining the preferred test in the first stage. A negative value thus indicates a participant’s preference to undergo the screening test. The questionnaire was designed using the random task generation method provided by Sawtooth. The DCE included a total of 20 blocks with 10 choice sets each. Each study participant saw 1 block of choice sets, consisting of 10 choice sets, from the 20 blocks. To test for internal validity, 1 fixed choice set offering 2 alternatives is common across all blocks, of which one is intended to be strictly dominant over the other.

The levels of cost were selected to reflect the costs of different procedures in reality. Participants in qualitative interviews also demonstrated a similar perception regarding the cost of the tests. Certain within-concept prohibitions were also specified to provide combinations of attributes that were realistic. This included prohibiting high out-of-pocket payment costs and the presence of mild pain for stool-based tests. However, we allowed the blood-based test to appear together with a higher cost given that commercial companies may set higher prices [32]. The coverage matrix of the DCE design was examined using Sawtooth Lighthouse Studio to ensure all the parameters can be estimated. In addition, considering the large range of costs, we treated cost as discrete variables rather than a continuous.
variable in our analysis. An example of a DCE choice task is presented in Figure 1.

Participants reported any familial history of CRC and if they had attended any type of CRC screening in the past. Various sociodemographic information was also collected. All variables were coded as categorical variables, and some variables were subsequently recoded as binary or ternary variables to form meaningful subgroups for analyses. Psychosocial inventories—the Zimbardo Time Perspective Inventory [33], the Intolerance of Uncertainty Scale [34], and the Duke UNC Functional Social Support Questionnaire [35]—were also included to measure participants’ degree of present orientation, intolerance of uncertainty, and social support, respectively.

Figure 1. An example of a discrete choice experiment choice task. SDG: Singapore Dollar (SGD 1 = approximately US $0.75).

**Participant and Sampling**

This study was conducted as a web-based survey hosted on REDCap from May 19, 2022, to May 28, 2022. The target population for this study was Singapore citizens and permanent residents aged ≥40 years. The survey was sent to the participants from Singapore Population Health Studies web-based panel. This web-based panel consists of a cohort that is broadly representative of the general Singapore population. Participants gave their implicit consent by participating in the survey and were reimbursed with Singapore $10 (approximately US $7.50) for every successful completion of the survey.

In calculating the minimum sample size, we made use of the proposed formula of \( n > \frac{500 \times c}{(t \times a)} \) by Johnson and Orme [36], where 500 is a fixed variable, \( c \) demotes the largest number of levels for a certain attribute, \( a \) indicates the number of DCE choice sets per block of questionnaire, and \( t \) refers to the number of alternatives per DCE choice set (excluding “Opt-out”). Accordingly, the sample size required for this study should be >125 participants (500*5/(2*10)=125). In addition, Lancsar and Louviere [30] suggested 20 responses per block or questionnaire, which led to a minimum sample size of 400. Considering that the DCE may be relatively difficult to answer, we expected a relatively high nonresponse rate of 20% to 30%. Hence, we set the target sample size to be 600. We subsequently tested the sample size using simulation functions with Sawtooth Lighthouse Studio, which deemed the sample size of 600 as sufficient. Owing to the overwhelming response, our final sample size of 1189 participants not only meets the minimum required sample size but also allows for the possibility of conducting flexible subgroup analyses.

**Statistical Analysis**

A total of 2 models were tested: a mixed logit (MXL) model and a mixed-mixed multinomial logit (MMML) model [37]. The MXL model was selected to account for the correlation introduced by the repeated observations from each participant and to relax the assumptions of independence from irrelevant alternatives [38]. This model assumes that the choices made by the same participant are correlated, and preference heterogeneity exists across the population sample. The interpretation of the mean preference weights is made in relation to the chosen base level, and the SD of each level indicates the variability in the mean preference weights. The MMML model was also selected as it incorporates both MXL and the latent class logit model. Unlike latent class logit where a homogeneous fixed preference is assumed within each latent class, a distribution of random coefficients is specified in the MMML model. Within each class, preference weights and the average probability of each demographic within each class can be derived. The number of classes was chosen by examining the Bayesian information criterion.

In addition, the conditional relative importance (CRI) for a given attribute, defined as the difference between the highest preference weight of the attribute level and the lowest preference weight of the attribute level, was reported. A higher CRI indicates the attribute is more important in designing the CRC screening program. Profile-based normalization was then applied to normalize the sum of CRI of all the attributes to 1.

A left-specific constant was included in each regression, with a statistically significant coefficient indicating a left-right bias in the study [39]. Participants in DCE may take shortcuts and use simplifying heuristics when answering DCE questions, which can introduce an unintended source of variability in the...
data. When using reading order heuristic, study participants may tend to choose the option on one side [40]. Incorporating a variable indicating the position of the choice in the regression can disentangle the associated bias [41,42].

All levels were coded as dummy variables. Reference levels were selected based on the current recommendation for CRC screening for the average-risk adult population. However, we intentionally designated “No Recommendation” as the reference level for the “Recommendation” attribute to investigate the nuanced preferences arising from diverse recommendations made by different individuals. Continuous psychological variables were demedian by subtracting the median value from each data point, resulting in a centered distribution. As for the classification of the continuous sociodemographic variables, age bands were set according to the definition of senior citizen locally, which is at the age of ≥60 years. Household income was grouped to ensure a sufficient sample size in both groups.

Statistical significance was set at P<.05. All quantitative data analyses were carried out using the statistical software R (version 4.2.1; The R Foundation) [43].

Results

Qualitative Component

Sample Characteristics

A total of 30 participants completed the interview. A summary of their sociodemographic characteristics is presented in Multimedia Appendix 2. Of these participants, 14 had undergone a colonoscopy, 20 had taken a stool test, and 6 had taken a blood-based test. A total of 5 participants had not undergone any screening for CRC. Five main themes that elucidated the participants’ motivations and important attributes of CRC screening were identified: (1) accuracy, (2) cost, (3) perceived risk and pain, (4) convenience of the test, and (5) the method of sample collection.

Accuracy

When it came to blood-based tests, many participants were uncomfortable with its inferior accuracy relative to other tests. They saw it as “pointless” due to its possibility of giving rise to many false positives and negatives:

...half the chance of being accurate, then why waste my blood? [P6, female]

If [accuracy is] too low, then it defeats the purpose already...you want to know whether you have cancer or not, you see. [P24, female]

For participants who felt that they had no symptoms, the level of accuracy of the stool-based tests provided sufficient assurance that they felt a colonoscopy was unnecessary. Many cited an approach of escalation—to undergo a colonoscopy only in the event of a positive stool-based test:

I’m not in the high-risk category...I’ll go for the FIT first, and then based on [the results], I’ll go for the colonoscopy. [P22, female]

Participants who had undergone invasive tests such as colonoscopy either (1) were diagnosed with hemorrhoids or (2) had discovered blood in their stools. They had opted for the colonoscopy under their physician’s or friend’s advice, seeing that a colonoscopy was a more “comprehensive” or “complete” test. Many felt that the relatively higher accuracy of the colonoscopy provided them greater ease of mind:

[Colonoscopy] gives a more accurate reading, because you’re able to see...what’s inside. [P19, female]

[Colonoscopy] is not so comfortable, but it’s comfortable to the mind. [P30, male]

Perceived Risk and Pain

The stool-based and blood-based tests were more favorable to most participants as the process was “simple” and “pain-free.” Furthermore, many who had undergone a blood-based test did so as part of their annual comprehensive health check-up and saw little extra risk or pain in doing it:

It’s less traumatic to the patient. That kind of [needle] pain is bearable...no issue. [P1, female]

If it’s part of the blood works, might as well, right? Since they are already drawing blood? I don’t mind testing for [colorectal cancer] as well. [P17, female]

Many of these participants who had no prior experience undergoing colonoscopy were more likely to express fear of the risk of colon perforation from the procedure. Due to its invasive nature, many were also “scared” and “uncomfortable” with the pain the procedure might induce:

I don’t know how big the scope is...how difficult is it to insert? Will it damage anything permanently? [P11, male]

The sort that comes with pain...[I] may get cold feet. [P17, female]

Interestingly, most participants who had undergone colonoscopy had little qualms about the risk and pain of the procedure. Instead, many expressed difficulties adhering to the bowel preparation instead:

The agony part was the bowel preparation. I can’t finish the 4 litres...I gave up at 2 litres. [P14, male]

Method of Sample Collection

While being relatively easy to conduct, some participants shared reservations about the collection of stool samples. They saw the process as “dirty,” “disgusting,” or “troublesome,” especially when 2 separate stool samples were required. While some complained about the uncomfortable experience of stool collection, some expressed personal concerns about improper collection:

...because you have to do it on your own, especially when you have to dig the stool, I’m not sure whether we are doing correctly or not. [P28, male]
Nevertheless, participants who have done it across many years expressed little of such concerns, seeing it as a routine exercise that was necessary:

> You’ve done it once and then you make it an annual exercise...it’s not a big deal. [P10, female]

**Convenience of Test**

As colonoscopies and blood-based tests require medical professionals to perform them, some participants felt that the need to arrange a doctor’s appointment was time-consuming. This was especially true for colonoscopies, where a referral from a primary care physician is required to receive a subsidized rate for the colonoscopy:

> ...you have to go to the [primary care] polyclinic and then get a referral, see the specialist and then wait for the appointment...you know, so it’s a bit more cumbersome. [P25, male]

However, participants shared that stool-based tests were relatively more accessible, with kits being easily obtained at pharmacies or mailed to them on request. Even when returning stool samples, some participants shared that mailing services to the laboratory were available, which saved them the shame and hassle of dropping them off at a clinic:

> [It's] kind of a deterrent because you have a book an appointment...compared to FIT test, you can just drop by any of the pharmacies...it's a lot more convenient. [P20, female]

**Cost**

Many participants were aware of the stark difference in cost between a colonoscopy and a stool-based or blood-based test. While many participants, especially those younger and working, had employer and private insurances to subsidize a colonoscopy procedure, they highlighted that out-of-pocket cost was still substantially higher. Many cited that a higher willingness to pay must come in tandem with either higher accuracy and lower frequency of testing:

> If the colonoscopy is 70-80% [accurate], and the other tests are also 70-80% [accurate], of course I will choose the simple one. No point to go for a colonoscopy...pay so much, go through the hassle... [P24, female]

> [If] you do this scope, one time, last you for 10 years, [because them you] don’t have to collect stool sample at the next medical check-up. [P29, male]

**Quantitative Component**

**Sample Statistics**

A total of 1189 participants completed the study. Of these, 127 (10.7%) participants did not have complete sociodemographic information, while 44 (3.7%) participants failed the fixed choice task. This led to 168 (14.1%) participants dropping out, leaving a sample of 1021 participants for analysis. The demographic characteristics of all participants are presented in the Multimedia Appendix 2.

**MXL Model**

The main results of the MXL model are presented in Table 1. All mean coefficients were significant at $P<.05$. On average, participants exhibited a higher preference for blood-based tests relative to a 2-day stool-based test (coefficient=0.40, 95% CI 0.24-0.55). Participants also exhibited a higher preference for a 1-day stool-based test relative to a 2-day stool-based test (coefficient=0.27, 95% CI 0.10-0.45). The preference for these 2 procedures were however not statistically different from each other (coefficient=0.12, 95% CI –0.04 to 0.29).

The profile-based normalized CRI of the 6 attributes based on the MXL model is presented in Figure 2. Ranking the attributes, participants were most concerned with the cost and sensitivity of the screening test. This is followed by the procedure type, the risk level, the pain level, and ultimately the recommendation received for the screening test.
## Table 1. Mixed logit analysis\(^a\)\(^b\).

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Mean Coefficient (95% CI)</th>
<th>SD Mean Coefficient (95% CI)</th>
<th>P value</th>
<th>SD P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>0.13 (0.06 to 0.12)</td>
<td>3.68 (3.67 to 4.05)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Opt-out</td>
<td>-1.74 (-1.93 to -1.55)</td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Colonoscopy</td>
<td>-0.73 (-0.90 to -0.57)</td>
<td>0.98 (0.85 to 1.12)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CT(^d) colonography</td>
<td>-0.75 (-0.91 to -0.60)</td>
<td>0.65 (0.51 to 0.80)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Stool-based (2 days)</td>
<td>0.00 (Reference)</td>
<td>0.00 (Reference)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Stool-based (1 day)</td>
<td>0.27 (0.10 to 0.45)</td>
<td>0.93 (0.71 to 1.16)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Blood-based</td>
<td>0.40 (0.24 to 0.55)</td>
<td>0.91 (0.76 to 1.05)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pain level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pain</td>
<td>0.00 (Reference)</td>
<td>0.00 (Reference)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mild pain</td>
<td>-0.54 (-0.64 to -0.45)</td>
<td>0.26 (0.11 to 0.40)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1.63 (1.52 to 1.75)</td>
<td>0.09 (-0.05 to 0.24)</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>0.70 (0.59 to 0.82)</td>
<td>0.02 (-0.13 to 0.18)</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.00 (Reference)</td>
<td>0.00 (Reference)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>60</td>
<td>-1.26 (-1.43 to -1.08)</td>
<td>1.74 (1.53 to 1.95)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Recommendation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Promotion Board</td>
<td>0.88 (0.77 to 1.00)</td>
<td>0.11 (-0.04 to 0.25)</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Doctors</td>
<td>0.68 (0.56 to 0.80)</td>
<td>0.01 (-0.14 to 0.17)</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>Family or friends</td>
<td>0.35 (0.23 to 0.48)</td>
<td>0.21 (0.05 to 0.36)</td>
<td>.008</td>
<td></td>
</tr>
<tr>
<td>Neither</td>
<td>0.00 (Reference)</td>
<td>0.00 (Reference)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cost (Singapore $(^e))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.00 (Reference)</td>
<td>0.00 (Reference)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>-0.35 (-0.50 to -0.19)</td>
<td>0.38 (0.16 to 0.59)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>-0.81 (-0.93 to -0.69)</td>
<td>0.08 (-0.09 to 0.25)</td>
<td>.35</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>-2.39 (-2.53 to -2.25)</td>
<td>0.46 (0.30 to 0.62)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>-3.88 (-4.07 to -3.68)</td>
<td>1.53 (1.33 to 1.73)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Risk of test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No risk</td>
<td>0.00 (Reference)</td>
<td>0.00 (Reference)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1% risk of adverse event</td>
<td>-0.74 (-0.86 to -0.62)</td>
<td>0.05 (-0.11 to 0.21)</td>
<td>.53</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Log-likelihood: -8353, Akaike information criteria: 16,777, and Bayesian information criteria: 17,054.

\(^b\)Mean refers to the population mean. SD measures the individual preference heterogeneity. A significant value means that the preference for the corresponding level is heterogeneous at the individual levels.

\(^c\)Not applicable.

\(^d\)CT: computed tomography.

\(^e\)Singapore $1=approximately US $0.75.
**MMML Model: Preference Analysis**

The results from the MXL model suggested preference heterogeneity across the participants, with the SDs of several preferences being statistically significant. Thus, we ran an MMML model while also estimating the posterior class membership probabilities. A 2-class MMML model was the most appropriate based on the BIC. The classes were labeled post hoc as (1) strong supporters and (2) weak supporters, based on the coefficient value for “None.” A more negative coefficient value means people are more willing to take the screening test in real life. The class shares for strong supporters and weak supporters are 38.09% (n=339) and 61.91% (n=632), respectively. The main results of the MMML model are presented in Table 2. The full table with SD is presented in Multimedia Appendix 3.

Strong supporters did not exhibit a higher preference for blood-based tests relative to a 2-day stool-based test. However, weak supporters had a higher preference for blood-based tests compared with a 2-day stool-based test (coefficient=0.66, 95% CI 0.44-0.88). Similar to the results in the MXL model, strong and weak supporters were most concerned with the cost and sensitivity of the screening test. Weak supporters were more likely than strong supporters to be concerned with the procedure, pain level, and risk of test. Compared with an existing national screening program that is 2-day stool-based, has no pain, 80% sensitivity, recommended by the government’s Health Promotion Board, costs Singapore $5, and has no risk, the relative uptake rate of a blood-based screening test with all else constant will increase by 0.2% (95% CI 1.2% to 1.6%) for strong supporters and increase by 5.9% (95% CI 3.6% to 8.2%) for weak supporters.
Table 2. Mixed-mixed multinomial logit analysis.

<table>
<thead>
<tr>
<th>Mean coefficient</th>
<th>Class 1: Strong supporters</th>
<th>P value</th>
<th>Class 2: Weak supporters</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (95% CI)</td>
<td></td>
<td>Coefficient (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>0.28 (0.15 to 0.42)</td>
<td>&lt;.001</td>
<td>0.08 (−0.02 to 0.18)</td>
<td>.13</td>
</tr>
<tr>
<td>None</td>
<td>−5.13 (−5.79 to −4.46)</td>
<td>&lt;.001</td>
<td>−0.86 (−1.10 to −0.61)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colonoscopy</td>
<td>−0.35 (−0.71 to 0.02)</td>
<td>.07</td>
<td>−1.05 (−1.28 to −0.81)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CT colonography</td>
<td>−0.53 (−0.87 to −0.18)</td>
<td>&lt;.001</td>
<td>−1.01 (−1.23 to −0.78)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stool-based (2 days)</td>
<td>0.00 (Reference)</td>
<td>&lt;.001</td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>Stool-based (1 day)</td>
<td>−0.20 (−0.57 to 0.16)</td>
<td>.27</td>
<td>0.63 (0.39 to 0.87)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Blood-based</td>
<td>0.06 (−0.27 to 0.49)</td>
<td>.70</td>
<td>0.66 (0.44 to 0.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pain level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No pain</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>Mild pain</td>
<td>−0.25 (−0.46 to −0.04)</td>
<td>.02</td>
<td>−0.84 (−0.98 to −0.70)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2.92 (2.61 to 3.24)</td>
<td>&lt;.001</td>
<td>1.25 (1.10 to 1.41)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>95</td>
<td>1.58 (1.29 to 1.87)</td>
<td>&lt;.001</td>
<td>0.42 (0.25 to 0.58)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>80</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>60</td>
<td>−1.65 (−2.04 to −1.27)</td>
<td>&lt;.001</td>
<td>−0.96 (−1.19 to −0.72)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Recommendation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Promotion Board</td>
<td>1.24 (0.99 to 1.49)</td>
<td>&lt;.001</td>
<td>0.79 (0.62 to 0.96)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Doctors</td>
<td>0.89 (0.64 to 1.14)</td>
<td>&lt;.001</td>
<td>0.67 (0.50 to 0.83)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Family or friends</td>
<td>0.17 (−0.08 to 0.43)</td>
<td>.18</td>
<td>0.40 (0.22 to 0.58)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Neither</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>Cost (Singapore $d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>−0.67 (−1.02 to −0.32)</td>
<td>&lt;.001</td>
<td>−0.33 (−0.55 to −0.11)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>30</td>
<td>−0.97 (−1.24 to −0.70)</td>
<td>&lt;.001</td>
<td>−0.83 (−0.99 to −0.66)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>400</td>
<td>−2.24 (−2.55 to −1.94)</td>
<td>&lt;.001</td>
<td>−2.93 (−3.14 to −2.72)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1000</td>
<td>−4.00 (−4.41 to −3.59)</td>
<td>&lt;.001</td>
<td>−4.46 (−4.75 to −4.18)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Risk of test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No risk</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>1% risk of adverse event</td>
<td>−0.60 (−0.85 to −0.35)</td>
<td>&lt;.001</td>
<td>−1.04 (−1.21 to −0.87)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Class membership</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>—</td>
<td></td>
<td>0.19 (0.11 to 0.27)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>—</td>
<td></td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>—</td>
<td></td>
<td>0.52 (0.41 to 0.63)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Non-Chinese</td>
<td>—</td>
<td></td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-60</td>
<td>—</td>
<td></td>
<td>0.00 (Reference)</td>
<td>—</td>
</tr>
<tr>
<td>Variable</td>
<td>Class 1: Strong supporters</td>
<td>P value</td>
<td>Class 2: Weak supporters</td>
<td>P value</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-----------------------------</td>
<td>---------</td>
<td>--------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Household income level (Singapore $)</td>
<td>0.00 (Reference)</td>
<td></td>
<td>-0.52 (-0.61 to -0.42)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High income (≥60000)</td>
<td>0.00 (Reference)</td>
<td></td>
<td>-0.13 (-0.22 to -0.05)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lower income (≤59999)</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td>0.00 (Reference)</td>
<td></td>
<td>-0.06 (-0.16 to 0.04)</td>
<td>.21</td>
</tr>
<tr>
<td>Married</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Single or divorced or widowed or separated</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td>0.08 (-0.02 to 0.19)</td>
<td>.13</td>
<td>0.02 (-0.22 to -0.05)</td>
<td>.61</td>
</tr>
<tr>
<td>Primary and secondary</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Preuniversity</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>University and above</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Housing type</td>
<td>0.00 (Reference)</td>
<td></td>
<td>-0.47 (-0.60 to -0.35)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Public housing</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Private housing</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Working status</td>
<td>-0.24 (-0.34 to -0.14)</td>
<td>&lt;.001</td>
<td>-0.07 (-0.22 to -0.05)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Currently working</td>
<td>-0.24 (-0.34 to -0.14)</td>
<td>&lt;.001</td>
<td>-0.07 (-0.22 to -0.05)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not working or retired or student</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Family history of CRC&lt;sup&gt;e&lt;/sup&gt;</td>
<td>-0.74 (-0.85 to -0.63)</td>
<td>&lt;.001</td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>CRC screening history</td>
<td>-0.52 (-0.60 to -0.44)</td>
<td>&lt;.001</td>
<td>-0.16 (-0.17 to -0.14)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.00 (Reference)</td>
<td></td>
<td>0.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Perceived safety of test score</td>
<td>-0.16 (-0.17 to -0.14)</td>
<td>&lt;.001</td>
<td>-0.03 (-0.05 to -0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Social support score</td>
<td>-0.03 (-0.05 to -0.02)</td>
<td>&lt;.001</td>
<td>-0.03 (-0.05 to -0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Present orientation</td>
<td>-0.03 (-0.05 to -0.02)</td>
<td>&lt;.001</td>
<td>-0.03 (-0.05 to -0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Intolerance of uncertainty</td>
<td>-0.01 (-0.02 to -0.01)</td>
<td>&lt;.001</td>
<td>-0.01 (-0.02 to -0.01)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>Log-likelihood: −8041, Akaike information criteria: 16,257, Bayesian information criteria: 16,955.
<sup>b</sup>CT: computed tomography.
<sup>c</sup>Not applicable.
<sup>d</sup>Singapore $1=approximately US $0.75.
<sup>e</sup>CRC: colorectal cancer.

**MMML Model: Analysis of Demographic Information**

Weak supporters were more likely to be women, ethnic Chinese, and people aged 40 to 60 years. However, strong supporters were more likely to be working, have higher income levels, and live in private housing. Strong supporters were also more likely to have a family history of CRC and to have opted for CRC screening in the past. Compared with weak supporters, strong supporters were also found to have higher scores for the perceived safety of CRC screening tests, social support, present orientation, and intolerance of uncertainty.

**Discussion**

**Principal Findings**

For every screening program to be successful, it is important to identify patterns in the population’s choice for CRC screening test to encourage uptake. By using a sequential exploratory mixed methods design and using a web-based cohort that is designed to be representative of the general Singapore population, our study is able to identify salient attributes of screening among participants to inform our understanding of the population’s preference. Furthermore, controlling for some of the salient attributes in our DCE allows for a better...
interpretation of the preference of specific procedures, which is likely to encompass the value of convenience and method of sample collection that is otherwise not considered in the DCE.

Our DCE results suggest that most participants preferred a blood-based test over a 2-day stool-based test and colonoscopy after accounting for the other attributes (eg, sensitivity). A blood-based test was perceived to be pain free, with a method of sample collection that was relatively simpler, compared with a colonoscopy. This pattern of aversion to aspects of pain and risk that comes with colonoscopy is supported by several studies [44,45]. Furthermore, studies have also supported the convenience and ease of the blood-based tests as compared with colonoscopy, as no preparation is involved [44,46]. Stool-based tests were regarded as unpleasant and disgusting to some participants in line with studies from the United States [47], Australia [48], and Germany [49], wherein participants expressed a preference for blood over stool sampling. However, comparing a blood-based test with a 1-day stool-based test, the utility for a blood-based test was not statistically significantly higher. This suggests that the requirement of testing twice was a key contributor to the perceived inconvenience and disutility of a stool-based test.

One advantage of DCE is its ability to profile the population and understand the preference and estimate the uptake in different subpopulations. We profiled the population into 2 classes, the strong supporters for CRC screening who were indifferent between the 2-day stool-based test and blood-based test, and the weak supporters for CRC screening who preferred the blood-based test over the 2-day stool-based test. Our results suggest that the sensitivity of tests was a key consideration of participants. For example, a 2-day stool-based test with a sensitivity of 80% was preferred by both strong supporters and weak supporters compared with a blood-based test with a sensitivity of 60%. Unfortunately, the shortcoming of existing blood-based tests lies in the relatively inferior sensitivity and specificity [11,12], which was a concern for participants based on the qualitative interviews as well. In reality, while the procedure of blood-based tests itself is preferred, the low sensitivity and specificity rate is unlikely to appeal to the masses. However, if blood-based tests can achieve the same sensitivity as 2-day stool-based tests, by offering blood-based tests to weak supporters, there is a potential to increase the relative CRC screening uptake by approximately 6%. The impact is likely to be significant as weak supporters make up approximately 62% of the population. Hence, further investment in research and development to improve the accuracy of blood-based tests could be beneficial to society.

The weak supporters identified in the study were less likely to do CRC screening in real life, but showed a higher preference for blood-based tests. Several observable demographic factors were associated with being weak supporters, including being female, ethnic Chinese, and younger and having lower income. The government may thus use targeted information campaigns to strong supporters should focus on better accuracy (eg, a blood-based test with high accuracy). Both strong supporters and weak supporters valued the recommendation from the Health Promotion Board [50] most, the governmental agency driving preventive care under the Ministry of Health in Singapore.

While DCEs can inform patients’ preferences on specific tests, a proper health technology assessment should be performed on whether blood-based tests are appropriate in each country. Unfortunately, cost-effectiveness studies on SEPT9 were rarely included in systematic reviews of the cost-effectiveness of CRC screening strategies [51,52]. All found studies that included SEPT9 as a screening strategy discovered that an annual SEPT9 screening was more cost-effective than no screening [18,53,54]. A total of 2 studies that relied on test characteristics of the earlier version of SEPT9 found that annual screening with FIT dominated the SEPT9 strategy [18,54], meaning that using FIT provided superior outcomes at a lower cost compared with using SEPT9. However, a more recent study using test characteristics of the improved version of SEPT9 with higher sensitivity and specificity found that the strategy resulted in higher quality-adjusted life years gained, CRC cases adverted, and CRC deaths adverted compared with other screening strategies [53]. Nevertheless, the strategy of using SEPT9 remained more costly than FIT as it resulted in a 63% higher referral for colonoscopy than FIT, increasing the cost by 26%. As a result, the strategy of FIT was still more cost-effective than SEPT9. However, these conclusions were based on perfect adherence of strategies. On the basis of our findings, the blood-based tests like SEPT9 are more likely to have higher adherence than stool-based tests such as FIT if similar accuracy can be achieved. Indeed, 2 studies that considered imperfect uptake found that when the uptake rate of FIT fell below 70% relative to that of SEPT9, FIT was no longer more cost-effective than SEPT9 [18,54]. Thus, the possibility of SEPT9 being more cost-effective than FIT likely hinges on (1) an improved version of SEPT9 with higher sensitivity and specificity and (2) a significantly higher uptake for SEPT9 over FIT. 

Policy Implications

It is inevitable that the next frontier of cancer screening will be the adoption of blood-based tests [8]. Multiple such tests are being evaluated or studied currently [7]. Hence, it is no longer far-fetched that policy makers need to decide whether and how to include blood-based tests in the national cancer screening program. In the overall landscape of cancer screening, different stakeholders have different views on the matter. Patients and clinicians alike will want any patient with cancers to be diagnosed earlier, while developers and manufacturers of the test will ultimately focus on monetary returns as a primary consideration. Policy makers and government agencies have to determine the cost-effectiveness of such tests and be mindful of all the additional subsequent more invasive and expensive tests that would be performed in the presence of a positive test result, all of which can compound health care costs. Aside from
the financial aspect, the other downsides of screening, such as patient anxiety and lead-time bias, will all need to be considered seriously. Most of the existing studies examining the attitude and preference for blood-based CRC tests were conducted in Western societies [26,27]. Preference of the technology in Asia is understudied. Therefore, studying preferences in Singapore, with its mix of East, South, and Southeast Asian cultures, provides a valuable addition to the literature.

Screening test only serves as the first step—follow-up diagnostic tests are required to complete the process. The gaps between stool-based tests and follow-up colonoscopies have been documented in the literature [55,56], which compromises the benefit of the screening program. Nonetheless, blood-based tests give policy makers another option to improve CRC screening. While blood-based tests in themselves may result in higher colonoscopy referral rates, blood-based tests may be combined with the existing screening methods rather than replacing them. This potentially improves on current CRC screening program by reducing the burden of colonoscopy through a 2-step screening approach that triages positive stool-based test patients [37] and serves as an alternative for people who reject stool-based tests [22]. Additional research is required to address these practical issues and understand the value brought by blood-based tests before advocating for the inclusion of blood-based cancer screening tests into the national guidelines.

**Limitations**

We acknowledge several limitations in this study. First, the DCE could not encompass the entirety of decision attributes pertinent to CRC screening, potentially limiting the comprehensive representation of individual decision matrices. However, we prioritized the most salient attributes of the population through qualitative interviews. Moreover, the qualitative interviews helped furnish and provide supplementary perspectives beyond the finalized attributes. Second, it is imperative to note that our study was conducted within a relatively affluent nation, thereby limiting the generalizability of economic considerations, such as income sensitivity and trade-offs to settings with lower income levels. In addition, the availability, accessibility, and quality perception of essential infrastructure, services, and resources may be influenced by local contexts, and their differential manifestations in various settings could yield disparate research conclusions.

Nevertheless, this study advances our understanding of the preferences of the population for CRC screening tests with respect to the type of procedure. The quantified value of the population’s preferences can help in the design of more targeted policies to promote optimal screening behavior and improve screening rates. Given the constrained available resources, more resources can be allocated in the short term to (1) increase the awareness of noninvasive tests and (2) the accessibility of noninvasive tests. Given that stool-based tests are nationally recommended in Singapore, efforts addressing the emotional barriers of embarrassment and disgust of stool collection should be promoted to encourage the collection of stool as something fundamental and shameless. In the long-run, policy makers should consider investing in research and development to improve the accuracy of blood-based tests, as they are generally preferred over invasive tests and may lead to greater uptake. With an improved blood-based test that yields higher sensitivity and specificity rates, institutionalizing CRC screening alongside other routine blood works is likely to be widely acceptable.

**Acknowledgments**

YW, KKT, and ARC were responsible for the concept and design of the study and obtaining the funding for this study. CO, KKT, and ARC were responsible for the recruitment of participants and data collection. CO and YW were responsible for the analysis of data. YW, KKT, and ARC provided supervision and administrative support. CO and YW wrote the first draft of the manuscript. All authors were involved in the interpretation of the data and worked on critical revisions of the manuscript. All authors read and approved the final manuscript.

This research is supported by the Singapore Ministry of Health’s (MOH) National Medical Research Council under its Health Services Research New Investigator Grant (MOH-000740-00). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of MOH and National Medical Research Council.

**Data Availability**

Data are available upon request, subject to the approval of the review boards.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Attributes and levels of the discrete choice experiment with the total number of appearances and selections.

[DOCX File, 17 KB - publichealth_v10i1e53200_app1.docx ]

Multimedia Appendix 2

Sociodemographic of participants from the qualitative survey and the discrete choice experiment.
Multimedia Appendix 3
Full mixed-mixed multinomial logit analysis table.

References


Abbreviations

CRC: colorectal cancer
CRI: conditional relative importance
DCE: discrete choice experiment
FIT: fecal immunochemical testing
MMML: mixed-mixed multinomial logit
MXL: mixed logit
NUS: National University of Singapore
RCT: randomized controlled trial
©Clarence Ong, Alex R Cook, Ker-Kan Tan, Yi Wang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Original Paper

Newspaper Coverage of Hospitals During a Prolonged Health Crisis: Longitudinal Mixed Methods Study

Frank van de Baan¹, MSc; Rachel Gifford¹, PhD; Dirk Ruwaard¹, MD, PhD; Bram Fleuren², PhD; Daan Westra¹, PhD

¹Department of Health Services Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands
²Department of Work and Social Psychology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Frank van de Baan, MSc
Department of Health Services Research
Faculty of Health, Medicine and Life Sciences
Maastricht University
Duboisdomein 30
Maastricht, 6229GT
Netherlands
Phone: 31 043 38 81 553
Email: f.vandebaan@maastrichtuniversity.nl

Abstract

Background: It is important for health organizations to communicate with the public through newspapers during health crises. Although hospitals were a main source of information for the public during the COVID-19 pandemic, little is known about how this information was presented to the public through (web-based) newspaper articles.

Objective: This study aims to examine newspaper reporting on the situation in hospitals during the first year of the COVID-19 pandemic in the Netherlands and to assess the degree to which the reporting in newspapers aligned with what occurred in practice.

Methods: We used a mixed methods longitudinal design to compare internal data from all hospitals (n=5) located in one of the most heavily affected regions of the Netherlands with the information reported by a newspaper covering the same region. The internal data comprised 763 pages of crisis meeting documents and 635 minutes of video communications. A total of 14,401 newspaper articles were retrieved from the LexisNexis Academic (RELX Group) database, of which 194 (1.3%) articles were included for data analysis. For qualitative analysis, we used content and thematic analyses. For quantitative analysis, we used chi-square tests.

Results: The content of the internal data was categorized into 12 themes: COVID-19 capacity; regular care capacity; regional, national, and international collaboration; human resources; well-being; public support; material resources; innovation; policies and protocols; finance; preparedness; and ethics. Compared with the internal documents, the newspaper articles focused significantly more on the themes COVID-19 capacity ($P<.001$), regular care capacity ($P<.001$), and public support ($P<.001$) during the first year of the pandemic, whereas they focused significantly less on the themes material resources ($P=.004$) and policies and protocols ($P<.001$). Differences in attention toward themes were mainly observed between the first and second waves of the pandemic and at the end of the third wave. For some themes, the attention in the newspaper articles preceded the attention given to these themes in the internal documents. Reporting was done through various forms, including diary articles written from the perspective of the hospital staff. No indication of the presence of misinformation was found in the newspaper articles.

Conclusions: Throughout the first year of the pandemic, newspaper articles provided coverage on the situation of hospitals and experiences of staff. The focus on themes within newspaper articles compared with internal hospital data differed significantly for 5 (42%) of the 12 identified themes. The discrepancies between newspapers and hospitals in their focus on themes could be attributed to their gatekeeping roles. Both parties should be aware of their gatekeeping role and how this may affect information distribution. During health crises, newspapers can be a credible source of information for the public. The information can also be valuable for hospitals themselves, as it allows them to anticipate internal and external developments.

(JMIR Public Health Surveill 2024;10:e48134) doi:10.2196/48134

https://publichealth.jmir.org/2024/1/e48134
KEYWORDS

health communication; news coverage; media; misinformation; COVID-19; dissemination; communication

Introduction

It is important for health organizations to communicate with the public during health crises [1-4]. Newspapers, both print and on the web, form an important mode of communication, given their relatively high perceived credibility, frequent use, and agenda-setting role [5-8]. Newspapers can inform the public of recent developments, including the extent of the crisis and threats, and preventive measures [3,9,10]. This enables the public to be well informed, can reduce fear, and stimulates the public to adopt the behavior needed for crisis mitigation [11-14]. However, communication can also induce public anxiety and make people refrain from appropriate behavior [14-16]; misinformation, an overabundance of information, or a wrongful depiction of certain issues or phenomena can be contributing factors [15,17,18]. Therefore, it is essential that both health organizations and newspapers facilitate the distribution of information that is accurate, timely, and valid [9,19,20].

Traditionally, governments and public health organizations have been the main organizations involved in communication during a health crisis [21-23]. However, the COVID-19 pandemic has been recognized by health communication scholars as the first large-scale health crisis to occur in a new media context, characterized by, among other elements, a deluge of information and misinformation [15,23,24], relatively high rates of skepticism toward media reporting [23,25], and new organizations involved in sharing information [23]. Hospitals emerged as such a new organization because of the central role they played during the crisis in ensuring the treatment of patients with COVID-19 [26-28], often being on the front pages of newspapers [29]. Furthermore, hospitals experienced firsthand the effects of misinformation and skepticism toward media reporting, including continued pressure on hospital functioning owing to low vaccination rates [30,31] and increased aggression toward hospital workers [32]. In light of the new media context and the negative effects associated with flawed communication, scholars have emphasized the importance of gaining a better understanding of information distribution through media [19,33].

Studies analyzing communication through newspapers so far have shown mixed results. Some studies have suggested adequate information distribution, for example, during the SARS-CoV-1 outbreak [34] or the H1N1 influenza pandemic [35,36]. However, other studies have been indecisive or deemed that communication could be improved, including those focusing on the West Nile virus outbreak [37], the H1N1 influenza pandemic [9,38,39], and the COVID-19 pandemic [8,40,41]. Although these studies provide valuable insights, some aspects remain understudied because most studies focused on epidemiological information [34-37,39] or included a short time frame (ie, <6 months) [8,40,41]. Furthermore, in light of the new media context, it remains unclear how information from new key stakeholders, such as hospitals, is being presented to the public. Overall, there is a need for further insights into how information from health organizations is presented to the public by newspapers during health crises [19,33,42]; therefore, scholars have specifically called for studies using both qualitative and quantitative components to understand this phenomenon [19].

Given this need, this paper presents a longitudinal mixed methods study examining the representation of hospital information in (web-based) newspaper reporting during 12 months of a health crisis in the Netherlands. Specifically, this study aimed to (1) explore the situation of hospitals during the first year of the COVID-19 pandemic, (2) determine the degree to which the representation of the identified themes in internal documents aligned with what was presented to the public in newspaper articles, (3) examine the extent to which hospital actors were included in newspaper articles reporting on hospitals, and (4) assess whether newspaper articles contained misinformation. The results of this study deepen our understanding of the portrayal of organizational information in the media and the role of both media and health care organizations in facilitating effective communication, especially during health crises [19]. Moreover, these insights can help optimize communication strategies and facilitate improved information distribution during a health crisis [19]. As such, this study contributes to health emergency preparedness [43].

Methods

Overview

This longitudinal mixed methods study compared internal hospital data with newspaper coverage over time using thematic, quantitative, and manifest content analyses. We focused on (web-based) newspaper reporting because in the Netherlands and many other European countries, both regional and national newspapers still form an important source of information for the public, especially during health crises [44]. This is contrary to some other countries, such as the United States or Argentina, possibly because of lower trust in newspaper reporting in these countries [44]. We focused on the first year of the pandemic to encapsulate the most pressing period in terms of hospital admissions and deaths within the Netherlands [45,46]. This study was part of a larger research project investigating hospitals’ responses to the COVID-19 crisis.

Ethics Approval

Ethics approval was obtained for this study by the Faculty of Health, Medicine and Life Sciences Research Ethics Committee of Maastricht University (FHML-REC/2020/110).

Setting

This study focused on the Dutch province of Limburg, one of the most heavily affected provinces in the country during the first year of the COVID-19 pandemic [47]. The province had 1,115,872 inhabitants as of January 1, 2021 [48]. During the first year of the pandemic (February 27, 2020, to February 28, 2021), the province experienced 71,348 confirmed COVID-19 cases and 1438 COVID-19–related deaths [46,49,53].
There are 5 hospitals located in the province of Limburg. These 5 hospitals vary in size, ranging from one of the smallest in the country to one of the largest in the country in terms of the number of available beds, the number of patients treated, and the number of staff. One of the hospitals is an academic hospital, whereas the other 4 are general hospitals, of which 2 are top clinical centers. Geographically, the hospitals are equally spread across the province. All included hospitals provide both inpatient and outpatient care and have a 24-hour emergency ward. All hospitals are members of the same regional acute care network; regional acute care networks are networks through which Dutch healthcare organizations coordinate crisis responses [54].

The province of Limburg has 1 major regional newspaper covering the entire province, namely De Limburger, with a daily circulation of >120,000 newspapers [55]. The newspaper has no specific political leaning and is considered neutral [56,57]. The newspaper is available both web-based and in print.

Data Collection

This study used 2 data sources: internal data from all 5 hospitals and all related media coverage from the main newspaper of the province. On March 4, 2020, the province had the first confirmed COVID-19 case [58]; therefore, the time frame was set from March 1, 2020, to February 28, 2021.

Regarding the hospital data, each of the 5 hospitals provided us with in-depth organizational data from the first year of the COVID-19 pandemic in the Netherlands. Specifically, we were given crisis meeting minutes (action and decision lists) from all meetings held between March 1, 2020, and February 28, 2021, and internal communication videos posted during the same time frame. Ultimately, 763 pages of written documents were collected, together with 635 minutes of video communications. The documents and videos contained a variety of content, including detailed information on discussions held within the crisis teams, purchases made, and department-level issues.

The regional newspaper De Limburger was selected as a data source, as it is a large regional print newspaper, also available on the web, that covers the entire province where all the 5 hospitals are located. We included a regional newspaper as opposed to a national newspaper because, in light of gatekeeping theory, the regional newspaper would likely contain more articles about the hospitals [59,60]. We used the LexisNexis Academic (RELX Group) database to collect articles published about the hospitals [59,60]. To retrieve relevant articles, the keywords used were as follows: “COVID-19,” “corona,” “care” (zorg), “hospitals” (ziekenhuizen), or “hospitals” (ziekenhuis), and “COVID-19,” “corona,” “care” (zorg), “hospitals” (ziekenhuis), or “hospitals” (ziekenhuizen). In total, 14,401 newspaper articles were identified and screened for relevance based on title and preview. Articles were deemed relevant when they included information on ≥1 of the 5 hospitals and contained information on anything related to COVID-19 in these hospitals. When in doubt of relevance, the articles were fully read and included if relevant. Of the 14,401 articles, 194 (1.3%) were included in the data analysis. The reasons for the exclusion of a large number of articles include the fact that many did not relate to COVID-19 or any of the hospitals. In part, this has to do with the fact that we used a broad search string (using “OR” instead of “AND”), as we did not want to miss any relevant articles. Furthermore, the keyword “care” (zorg) also means “ensure” in Dutch, leading to the inclusion of articles with this keyword but not related to COVID-19 and hospitals.

Data Analysis

To assess newspaper reporting of the situation in hospitals, content and thematic analyses were performed [61,62]. All coding was done by hand, as opposed to computer-aided searches for certain terms, as this is more likely to lead to a more accurate interpretation of data [63].

Regarding the internal data (ie, the action and decision lists and communication videos), FvdB used thematic inductive coding to identify themes that characterized organizational reality during the pandemic [64]. This approach was deemed most suitable because through this technique, common themes revolving around a certain phenomenon (ie, the COVID-19 pandemic) can be identified [65]. It involved pattern recognition within the data, requiring constant careful rereading of the data and construction of themes based on the characteristics of the data [64]. The coding continued until no new themes emerged. Throughout the process, FvdB engaged in ongoing discussions with RG and DW regarding the emerging codes and identified themes and presented identified codes, grouped themes, and other updates. Through these discussions, the authors, for example, distilled public support as a main theme. This theme emerged through codes related to signs of public support and appreciation (eg, bringing gifts, care bonuses, and support through social media) and codes related to public aggression and frustration (eg, verbal assault of employees, impatient patients, and downplay of the severity of the crisis on social media). In total, 12 main themes were identified, which were used to create a codebook. Multimedia Appendix 1 includes the codebook with the definition of each theme and examples of quotes. This study was part of a larger project aimed at identifying hospitals’ challenges during the COVID-19 pandemic and related responses. For the project, interviews were conducted by FvdB and RG with hospital employees from all layers of the organization (eg, medical staff, managers, and chief executive officers [66,67]). This provided a certain safeguard that the main themes identified through the coding resonated with what was experienced in practice.

Quantitative thematic content analysis was used to determine the degree to which the representation of themes in the internal documents aligned with that in the newspaper [61]. The created codebook was used to deductively code the 194 newspaper articles and 763 pages of internal documents. For each newspaper article, sentences, paragraphs, or passages that encapsulated a certain theme from the codebook were coded [64]. For example, in one of the articles, the code human resources was assigned to the following paragraph:

Bottleneck in all hospitals remains the availability of people. Who also drop out. “Absenteeism is high,” says the hospital spokesperson. Many employees have [been] infected, many more than in the first wave. Moreover, the second wave lasts longer than the first.

Articles not covering any of the themes from the developed codebooks were assigned the code “miscellaneous” (6/194,
If an article covered multiple themes, it was given multiple codes. The same approach was used for the internal documents (i.e., each hospital’s action and decision lists per day). That is, sentences, paragraphs, or passages were deductively coded. For example, in one of the documents, the code ethics was assigned to the following passage:

When the corona law takes effect on Dec. 1, wearing a mouth mask will be mandatory in all public places. ...[But] we will not enforce this law because we have a care duty.

FvdB, RG, and DW engaged in initial deductive coding of the newspaper articles and internal documents. Differences in coding were discussed to reach a consensus on the coding procedure. FvdB coded the rest of the data and consulted with RG and DW in case of doubt. After the proportions of themes for the internal documents and newspaper articles were identified, chi-square tests were conducted to calculate differences between theme proportions of the internal documents and newspaper articles. Chi-square tests were performed using SPSS Statistics for Windows (version 27; IBM Corp). \( P \) values <.01 were considered statistically significant.

To assess the presence of misinformation in the newspaper articles, manifest content analysis was performed, which is a form of qualitative content analysis [62,68]. This method of content analysis regards the direct examination of text with a low level of abstraction and a low degree of interpretation [68]. This was considered an appropriate approach because we wanted to directly compare the text in newspaper articles (e.g., statements on absenteeism rates, intensive care unit capacity, or change in visitor policies) with the text of the internal documents to assess whether the information in the newspaper articles aligned with the information in the internal documents.

### Results

#### Identified Themes and Quantitative Differences in Coverage

The internal data revealed 12 themes regarding hospitals’ situation during the first year of the COVID-19 pandemic. The themes included COVID-19 capacity; regular care capacity; regional, national, and international collaboration; human resources; well-being; public support; material resources; innovation; policies and protocols; finance; preparedness; and ethics (refer to Multimedia Appendix 1 for a full description of themes and exemplary quotes). Table 1 shows the number of times the code of a theme was found in the internal documents and newspaper articles, the proportions of themes compared with the total amount of codes, and whether the proportions of the internal documents differed significantly from those of the newspaper articles. As can be seen, over the course of the year, the newspaper articles covered significantly more about COVID-19 capacity (\( P < .001 \)), regular care capacity (\( P < .001 \)), and public support (\( P < .001 \)). The articles covered significantly less about material resources (\( P = .004 \)) and policies and protocols (\( P < .001 \)). No significant differences were found in the other themes.

#### Changes in Reporting Over Time

During the first year of the pandemic, the attention toward themes in both the internal documents and newspaper articles fluctuated and showed both similarities and differences (Figures 1-3). During the first wave of the pandemic (March to May 2020), the attention given to the themes was largely similar. The content of both the internal documents and newspaper articles mainly related to COVID-19 and regular care capacity, the change in policies and protocols, forms of collaboration to accommodate the large influx of patients with COVID-19, and human resources needed to handle the crisis.
Between July and September 2020, some differences in attention toward themes arose. Specifically, the internal documents mainly covered the themes of policies and protocols and preparedness. The documents revealed that the hospitals reflected on the experiences of the first wave and the lessons learned. This was done to be able to better accommodate a potentially expected new wave after the summer, whereby policies and protocols were adjusted accordingly. The hospitals seemed to have had difficulty in making decisions on certain issues, given the uncertainty of a potential new wave and the lack of clarity on the effectiveness of certain policies. For example, the internal documents contained discussions on whether policies and protocols should be adjusted for staff and patients coming from countries with a high prevalence of infections. Suggestions were made to impose a 2-week quarantine for individuals returning from holidays in these countries. Ultimately, it was decided to follow the national recommendations. The newspaper reported on the themes of policies and protocols and preparedness at a more aggregate level. Instead, the articles focused largely on regular care capacity. The articles described the large backlash of regular care that occurred owing to most care practices being canceled during the first wave. Moreover, the articles described how hospitals aimed to catch up with regular care and the difficulties associated with that owing to, for example, staff being on holidays.

As depicted in Figure 2, a large proportion of the newspaper articles published between July and September 2020 also focused on staff well-being. This was ahead of the attention given to that theme in the internal documents. The newspaper articles described the physical and emotional impacts of the first wave on hospital staff, the insufficiency of a 2-week holiday for full recovery, and staff concerns regarding a potential new wave after the summer period. Similarly, public support became a focal point in the newspaper articles from August onward, before receiving attention in the internal documents. The articles highlighted the decline in public support after the first wave, with the hospital staff increasingly experiencing (verbal) aggression. Moreover, the articles reported on government plans to offer a bonus to health care workers as a sign of appreciation and the ensuing backlash owing to strict eligibility criteria.
From October 2020 onward, the attention toward themes largely aligned again. The main exception was coverage of the themes well-being and public support during January and February 2021, when the third wave of admissions of patients with COVID-19 declined. During this period, the internal documents frequently covered the well-being of staff. The documents showed that the hospitals sought ways to show appreciation and deemed it important to pay attention to the mental well-being of staff, as the work started to take its toll. Instead, the newspaper articles increasingly focused on public support. The articles described the growing tension in society, whereby skeptics, for example, called hospitals to check whether intensive care units were indeed occupied with patients with COVID-19. Furthermore, the newspaper articles described how hospital staff had to deal with increasing aggression.
Types of Articles

We distinguished 3 types of newspaper articles (refer to Table 2 for an overview of the types and number of newspaper articles). The first type regarded articles that included stakeholders as a source in the reporting. This included, for example, interviews with the chief executive officers or other types of hospital staff or reporting on specific issues such as having to deal with a large influx of patients with COVID-19 and the perspectives of stakeholders on this. The second type regarded articles that did not use stakeholders as a source or where this at least was not stated. This mainly entailed more general reporting, such as the use of eHealth services or changes in hospital visitor policies. As a third type, the newspaper published diary articles in which staff from the hospitals wrote about their personal experiences of working during the crisis. Staff, for example, wrote about the struggle of having to deal with increased emotional and work pressure whereas part of the general public believed COVID-19 to be comparable with the flu. The theme well-being was covered predominantly in diary articles in the newspaper (10/27, 37%).

Table 2. The types and number of articles published in the newspaper related to the hospitals’ situation during the first year of the COVID-19 pandemic (March 1, 2020, to February 28, 2021).

<table>
<thead>
<tr>
<th>Type of newspaper article</th>
<th>Values (N=194), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article including hospital as a source</td>
<td>132 (68)</td>
</tr>
<tr>
<td>Article not including hospital as a source</td>
<td>39 (20.1)</td>
</tr>
<tr>
<td>Diary article</td>
<td>23 (11.9)</td>
</tr>
</tbody>
</table>

Comparison of Content

All the newspaper articles were assessed for the presence of misinformation. For 7 (3.6%) of the 194 newspaper articles, the content could not be compared with the content of the internal documents because the internal documents did not contain related information. This was, for example, the case for an article published on April 4, 2020, stating that the hospitals in the studied province had the most hospital admissions in the entire country. This could not be verified, as the internal documents did not contain data on admissions in other regions. For all other newspaper articles (187/194, 96.4%), the content could be compared with the internal documents. No indication of the presence of misinformation was found. Multimedia Appendix 2 provides an example for every theme of information described in the newspaper compared with the internal documents. No indication of the presence of misinformation was found in any of the newspaper articles.

The discrepancies between newspapers and hospitals in their focus on themes can be attributed to their gatekeeping roles [59,60]. Media are faced with a plethora of events that could be covered and, therefore, have to decide which events will eventually be reported on and how [60]. During this process, several factors are considered, including timeliness, importance, proximity, and novelty [60]. Regarding our findings, the consideration of importance for readers (ie, the general public) could explain why newspaper articles covered significantly more about COVID-19 capacity, regular care capacity, and public support. Especially during the summer of 2020, the focus on regular care capacity in the newspaper deviated from the hospitals’ focus on policies and protocols and preparedness. Considering the growing need for regular care to be scaled up again after being largely canceled during the first wave [28], it could be that the newspaper deemed it necessary to frequently report on this issue, given the implications it had for the general population. The newspaper’s increased focus on public support from September 2020 onward may have been driven by societal debates on the appropriateness of public health measures and the related decrease in public support over time [69,70]. Moreover, the lower emphasis on policies and protocols and material resources could reflect the newspaper’s perception that these topics were of less importance to the readers, as they mainly had implications for hospitals and staff. However, the role of hospitals as frontline gatekeepers could also have contributed to this disparity [59,60]. As the primary source of information, organizations can decide which information is being shared with the media [60]. The hospitals may have provided less information regarding material resources and changes in policies and protocols because they deemed it irrelevant or wanted to keep it private.

Discussion

Principal Findings

This study aimed to examine newspaper coverage of the situation in hospitals in the Netherlands during the initial 12 months of the COVID-19 pandemic. Previous research found that hospital information was frequently included in media reporting during the COVID-19 pandemic [8,29]. However, our study specifically examined how hospitals navigated the first year of the pandemic and how this reality was communicated to the public through (web-based) newspaper articles. In total, we identified 12 themes that characterized the situation. We observed significant differences in the extent of coverage between internal organizational documents and newspaper articles for 5 (42%) of these 12 themes, namely COVID-19 capacity, regular care capacity, public support, material resources, and policies and protocols. An analysis of changes over time mainly revealed discrepancies in the attention paid to these themes between the first and second waves (May to September 2020) and at the end of the third wave (January and February 2021). Three types of articles could be distinguished, namely those that included stakeholders as a source, those that did not include stakeholders as a source, and diary articles. No
actors, the public, and news media are coproducing and responding to messaging [2,71]. In contrast to the more unidirectional flow of information, where public health actors disseminate information to the public through media, the current context involves a greater exchange of information between these entities. Studies have, for example, shown how public health actors monitor (social) media to assess the response to and the potential effectiveness of communication [2,71]. Our results suggest that the multidirectional nature of information should also be considered for hospitals. That is, not only do hospitals form an important source of information to promptly inform the public about recent developments, but newspaper articles can also yield relevant information for hospitals. For example, the themes well-being and public support gained attention in newspaper articles before gaining attention in the internal documents of hospitals, which shows that newspapers can be a means for those leading the crisis response to become aware of issues that might affect the organization. Moreover, the type of information found in articles, including details on available capacity, absenteeism rates, and experiences of frontline staff working during the crisis, can both meet the information needs of staff [72] and help those leading the crisis to assess and compare their internal situation with the situation of others. Although the extent to which newspapers provide new information and act as agenda setters might vary depending on the context [41], our results thus suggest that taking stock of media reporting can be valuable for hospital staff, and organizational leaders in particular, to be aware of changes occurring in both the external and internal environments and how these might impact the organization.

Contrary to what the public or stakeholders might perceive [5,6,73] and what some studies during prior health crises found [37,74], we did not find any indication of the presence of misinformation in any of the newspaper articles. Therefore, in line with previous research, our findings suggest that newspapers can be a credible source for those seeking valid information during a health crisis [35,36]. One explanation for not finding misinformation compared with previous studies could be that the study focused on hospital information instead of risk information, the latter being potentially characterized by more uncertainty [18,37]. Another explanation is that the studied newspaper was a regional newspaper. Previous studies found that compared with national newspapers, regional newspapers tend to have higher quality in reporting a local event because of audience considerations (i.e., a local audience prefers more details of a local event than national audiences) and better access to local sources [75,76]. Given the frequent inclusion of local sources, including diary articles written by hospital staff, our findings seem to support these studies. The large inclusion of sources supports the notion that during health crises, health systems would benefit from good relationships between stakeholders and traditional media [42,73,74].

**Practical Implications**

Our findings have several implications for crisis communication and health emergency preparedness. First, in light of the ongoing debate surrounding fake news and media distrust, our findings suggest that newspapers can be a valuable source of valid information during health crises [35,36]. Second, health care organizations should reflect upon and be aware of their role in informing the public and how this can be done through the media. Furthermore, health care organizations should be aware that newspapers can provide valuable information to them, allowing them to anticipate the developments occurring in both their external and internal environments. For newspapers, the inclusion of sources forms an important aspect for providing coverage that is reflective of organizational reality [76]. The use of nontraditional forms of reporting, such as diary articles, can be valuable in this regard. Furthermore, using a variety of reporting styles enables newspapers to cater to a diverse audience. The general public, as recipients of news, encompasses groups with diverse characteristics, including differences in sex, education level, and socioeconomic status [77]. These differences can influence the perceived relevance of information and subsequent behavioral responses [78]. By diversifying communications toward different groups of recipients, the effectiveness of communications can be enhanced [77]. Furthermore, newspapers need to be aware of their gatekeeping role, as this can explain deviances in the attention given to certain aspects. Although relevance to readers can justify these deviances, a decreased focus on relevance could make reporting more reflective of organizational reality. Ultimately, the public benefits from health organizations and the media collaborating in crisis communication [42,73]. By fostering relationships before a crisis occurs, both parties can ensure the rapid spread of valid information to the public [42,73].

**Limitations**

This study is subject to several limitations. First, the hospital theme proportions were constructed based on the content found in the action and decision lists. These sources reflect the outcome of a larger process, whereby it could be that verbally or via other means, such as email, more attention toward certain themes was given than was observed in the action and decision lists. Therefore, the actual focus of hospitals on themes throughout the year could differ from what we identified through our analysis. Nevertheless, considering the extensiveness of these documents (763 pages from 5 hospitals) and that these documents captured all aspects on which the hospitals needed to decide or take action, the focus derived from the action and decision lists would likely reflect the overall attention to a large extent. A second limitation regards the fact that in our study, we focused on a regional print newspaper, which could differ in reporting from national newspapers and other types of traditional media, including (local) radio and television. However, for local issues, the local newspaper can often be the agenda setter that other media follow, meaning that how the local newspaper reports on an issue can influence how it is presented in other media [76]. Moreover, considering that a substantial part of the population still accesses local newspapers as a source of information [44], studying this type of media itself could be considered still of value to health communication research.

**Conclusions**

The COVID-19 pandemic constituted a health crisis in which hospitals were a main actor and source of information. The situation of hospitals could be categorized into 12 themes, and
there was a significant difference in focus between the internal hospital documents and newspaper articles for 5 (42%) of the 12 identified themes. Newspaper articles focused significantly more on COVID-19 capacity, regular care capacity, and public support and significantly less on material resources and policies and protocols. The deviations seem to have mainly occurred between the first and the second COVID-19 waves and at the end of the third wave. The gatekeeping roles of the media and stakeholders could explain the deviations in focus. Some themes gained attention in the newspaper articles before gaining attention in the internal hospital data, supporting the notion that the flow of information can be multidirectional. Regarding newspaper reporting on hospital reality, no indication of the presence of misinformation was found. Overall, our findings suggest that newspapers can be a valuable source of information for both the public and health organizations during a health crisis.

Acknowledgments
The authors would like to thank Niels Hameleers for his support in data visualization. Data collection was part of a grant project funded by the Netherlands Organization for Health Research and Development (ZonMW) (grant 10430022010011). The funders had no role in the design of the study; collection, analyses, or interpretation of data; writing of the manuscript; or decision to publish the results. The authors did not use generative artificial intelligence in any part of the manuscript writing.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Codebook.
[DOCX File, 21 KB - publichealth_v10i1e48134_app1.docx ]

Multimedia Appendix 2
Comparison of the information found in the newspaper articles with that found in the internal documents of the 5 Dutch hospitals.
[DOCX File, 19 KB - publichealth_v10i1e48134_app2.docx ]

References


54. Over NAZL. Netwerk Acute Zorg Limburg. URL: https://www.nazl.nl/over-nazl [accessed 2021-09-24]


Web-Based Information on Spinal Cord Stimulation: Qualitative Assessment of Publicly Accessible Online Resources

Tiev Miller¹,², PhD; Ali Hosseinzadeh¹*, BSc; Thomas Thordarson¹, BSc; Tamila Kalimullina¹, BSc; Soshi Samejima³, DPT, PhD; Claire Shackleton¹,², PhD; Raza Malik¹,², PhD; Martín Calderón-Juárez¹,², MD, PhD; Rahul Sachdeva¹,², PhD; Andrei Krassioukov¹,²,⁴, MD, PhD

¹International Collaboration on Repair Discoveries, The University of British Columbia, Vancouver, BC, Canada
²Division of Physical Medicine and Rehabilitation, Department of Medicine, The University of British Columbia, Vancouver, BC, Canada
³Department of Rehabilitation Medicine, University of Washington, Seattle, WA, United States
⁴Spinal Cord Program, GF Strong Rehabilitation Centre, Vancouver Coastal Health, Vancouver, BC, Canada
*these authors contributed equally

Corresponding Author:
Andrei Krassioukov, MD, PhD
International Collaboration on Repair Discoveries
The University of British Columbia
818 West 10th Avenue
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 604 675 8810
Email: andrei.krassioukov@vch.ca

Abstract

Background: Despite the growing accessibility of web-based information related to spinal cord stimulation (SCS), the content and quality of commonly encountered websites remain unknown.

Objective: This study aimed to assess the content and quality of web-based information on SCS.

Methods: This qualitative study was prospectively registered in Open Science Framework. Google Trends was used to identify the top trending, SCS-related search queries from 2012 to 2022. Top queried terms were then entered into separate search engines. Information found on websites within the first 2 pages of results was extracted and assessed for quality using the DISCERN instrument, the Journal of the American Medical Association benchmark criteria, and the Health on the Net Foundation code of conduct certification. Website readability and SCS-related information were also assessed.

Results: After exclusions, 42 unique sites were identified (scientific resources: n=6, nonprofit: n=12, for-profit: n=20, news or media: n=2, and personal or blog: n=2). Overall, information quality was moderate (DISCERN). Few sites met all the Journal of the American Medical Association benchmark criteria (n=3, 7%) or had Health on the Net Foundation certification (n=7, 16%). On average, information was difficult to read, requiring a 9th- to 10th-grade level of reading comprehension. Sites described SCS subcategories (n=14, 33%), indications (n=38, 90%), contraindications (n=14, 33%), side effects or risks (n=28, 66%), device considerations (n=25, 59%), follow-up (n=22, 52%), expected outcomes (n=31, 73%), provided authorship details (n=20, 47%), and publication dates (n=19, 45%). The proportion of for-profit sites reporting authorship information was comparatively less than other site types (n=3, 15%). Almost all sites focused on surgically implanted SCS (n=37, 88%). On average, nonprofit sites contained the greatest number of peer-reviewed reference citations (n=6, 50%). For-profit sites showed the highest proportion of physician or clinical referrals among site types (n=17, 85%) indicating implicit bias (ie, auto-referral).

Conclusions: Overall, our findings suggest the public may be exposed to incomplete or dated information from unidentifiable sources that could put consumers and patient groups at risk.

(JMIR Public Health Surveill 2024;10:e50031) doi:10.2196/50031

KEYWORDS

access to information; consumer health information; internet; spinal cord stimulation; web-based information; communication; quality; readability; Google Trends; misinformation; synthesis
Introduction

Background
Spinal cord stimulation (SCS) is an emerging therapeutic approach that has been used as an intervention to address chronic pain [1,2], paralysis, and autonomic dysfunctions [3,4] resulting from injury and disease. SCS delivers electrical stimulation to the spinal cord using invasive or noninvasive interfaces. Groundbreaking pilot studies involving individuals with paralysis after spinal cord injury have demonstrated the potential benefits of SCS for improving functional recovery [5,6], resulting in a growing demand for information related to SCS among various consumer or patient groups, their caregivers, and the general public [7]. In the absence of evidence-based guidelines and standardized treatment options, people living with long-term disease and disability are likely to turn to other sources of information (eg, web-based search results), and potentially alternative treatment options (eg, advertised devices and surgical procedures), to help manage their condition. However, web-based information on SCS has not been evaluated in depth.

Currently, the US Food and Drug Administration [8] has only approved the use of epidural spinal cord stimulation (ESCS; ie, surgically implanted) to manage failed back surgery syndrome, refractory angina pectoris, peripheral arterial disease, complex regional pain syndrome, painful diabetic neuropathy as well as nonsurgical low back pain. Whether publicly accessible websites curating information on SCS provide clear indications (or contraindications) for the intended use (or potential risk) of this therapeutic modality remains unknown. Over the next decade, medical consumerism is projected to grow as people seeking treatment become more knowledgeable and active in their care through the expansion of web-based resources and the availability of medical information [9]. As the marketing of SCS to treat or manage various conditions expands concomitantly with the fragmentation of web-based medical information [10] and the gradual erosion of routine patient-to–health care provider interactions [11], patients may become more vulnerable to predatory marketing strategies that inflate the benefits of SCS while obscuring or underreporting its potential harms [12]. This market-driven shift by which medical information is being provided and subsequently accessed may engender patient consumerism and potentially alternative treatment options (eg, advertised devices and surgical procedures), to help manage their condition. However, web-based information on SCS has not been evaluated in depth.

The objectives of this qualitative study were (1) to systematically map web-based resources containing SCS-related information using common search methods that are openly and freely available to the general public and (2) to assess the content and quality of information regarding the use of SCS for the treatment or management of health conditions or symptoms.

Methods

Ethical Considerations
Ethics approval was not required for this study, as it did not involve human participants.

Search Strategy and Inclusion Criteria
This study was prospectively registered in Open Science Framework [28]. A pilot search was conducted to identify the most popular search terms (worldwide) within the last decade (2012 to 2022) for the topic “spinal cord stimulator” using Google Trends (Figure 1). All searches were performed using a United States IP designation to standardize the country-specific origin of returned results for each search engine [17]. A detailed description of the search strategy using the highest trending search queries (ie, “spinal stimulator” and “spinal cord stimulator”) is provided in the Search Strategy section in Multimedia Appendix 1 [17,28-77].
Figure 1. Search term identification. (A) Search terms with relevance to spinal cord stimulation were identified using Google Trends. Topic search using the term “spinal cord stimulator” was used to identify the top related queries: 1=spinal stimulator, 2=spinal cord stimulator, 3=spinal stimulation, and 4=spinal cord stimulation. Among the top queried terms, 1=spinal stimulator and 2=spinal cord stimulator ranked the highest in terms of relative search term popularity. (B) Search interest for the topic term “spinal cord stimulator” by global region indicated the United States (n=100), followed by Ireland (n=53), Laos (n=45), Australia (n=45), and United Kingdom (n=30) had the highest proportion of all queries (ie, not the highest number of queries) for this topic term within the specified timeframe (2012-2022).

This study considered publicly accessible websites providing information on SCS. Websites were categorized as scientific resources (ie, academic institutions and government organizations), nonprofit or foundations or advocacy organizations, for-profit or private sector or industry, news or media, or independent sources or personal blogs in accordance with subcategorizations previously outlined by Fisher et al [17]. Sites were not excluded on the basis of origin or date. Peer-reviewed papers or book chapters hosted by journal publisher websites were not included in the data extraction. Evidence suggests the general public may be unable to interpret or understand all scientific content found within peer-reviewed literature (ie, accessibility or readability of information) [78]. As this content has already been subjected to professional scrutiny from subject-area experts and editorial revision of considerable rigor or is altogether inaccessible (eg, paywall restricted), this information was not evaluated. Advertisement sites that did not include information relevant to SCS were also excluded.

Data Extraction
Data were extracted by 2 independent authors, and discrepancies were resolved through discussion or consensus with a third author. Extracted data included website type [17], characteristics (ie, site title, address [URL], publication date or most recent update, author names and credentials, country of origin or geographical designation, alternate accessibility options [format], and number of peer-reviewed reference citations), and SCS-related information provided (ie, SCS definition or summary, indications, contraindications, side effects or risks, device considerations, follow-up, outcomes, intended audience [consumer or caregiver, researcher or clinician, or both], and referral to a health care professional [eg, physician and therapist] or clinic).

Quality and Readability
In accordance with methods described in previous studies [17,18], the quality of the information provided on each website was assessed by 2 independent authors using the DISCERN instrument [29], the Journal of the American Medical Association (JAMA) benchmark criteria [30], and Health on the Net Foundation (HON) code of conduct certification [31]. Readability for each site was determined using the Flesch-Kincaid indices [32]. These assessments are described in detail in the Quality and Readability Assessments section in Multimedia Appendix 1.

Data Synthesis
Data extracted from websites were tabulated and aggregated according to site type. Websites were also explored for
high-frequency keywords. Webpages were imported as PDF to NVivo software (version 12; QRS International Pty Ltd) and archived for further analysis using the NCapture extension in Google Chrome. Website text from imported PDF files was then used to generate word clouds and frequency tables. Descriptive analyses were performed using SPSS software (version 28; IBM Corp). Average scores and ratings for each site type are expressed as mean and SD, median and IQR, or frequencies (n) and proportions (%) for continuous, ordinal, and dichotomous levels of data, respectively.

**Results**

**Screening and Search Term Frequency**

Using the term “spinal stimulator” in the first search and “spinal cord stimulator” in the second search (Figure 1), a total of 202 results were returned (Google: n=56, Baidu: n=39, Yahoo: n=58, and Bing: n=49). After exclusions (ie, duplicates, advertisements, and sites unrelated to SCS), a total of 57 unique websites were identified. Of these, 15 were host sites for peer-reviewed journal papers (n=8), book chapters (n=4), or patent applications (n=3) and were thus excluded from the main synthesis. Of the remaining 42 sites, the majority were categorized as for-profit (n=20) followed by nonprofit (n=12), scientific resource (n=6), news or media (n=2), and personal or blog site types (n=2). A flow diagram outlining the screening procedure is provided in Table 1.

Word frequencies for each search term used (ie, spinal, cord, and stimulator) are provided in Table 2 and Figure S1 in Multimedia Appendix 1. Word cloud and word frequency summaries are also provided in Figure S1 in Multimedia Appendix 1. Word frequency summaries showed that the occurrence of individual words comprising the search topic (ie, “spinal,” “cord,” and “stimulator”) was among the most frequent. This suggests the returned search results are likely to be an accurate representation of the web-based resources and SCS-related information that consumers would encounter irrespective of the search engine used [79,80]. Average occurrence (ie, frequency) of all 3 search terms was highest for scientific resource sites (spinal: mean 25.7, SD 16; cord: mean 23.7, SD 14.5; and stimulator: mean 13.8, SD 13.3), while proportional usage (ie, relative to all other words) was highest among personal or blog sites for the terms “spinal” (mean 4.4, SD 2.7) and “stimulator” (mean 3.5, SD 3.2) and news or media sites for the term "cord" (mean 5.4, SD 6.8).

**Table 1.** Search results and website selection.

<table>
<thead>
<tr>
<th>Screening procedure</th>
<th>Search engine</th>
<th>Baidu, n (%)</th>
<th>Google, n (%)</th>
<th>Yahoo, n (%)</th>
<th>Bing, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial search results (n=202)</td>
<td>56 (28)</td>
<td>39 (19)</td>
<td>58 (29)</td>
<td>49 (24)</td>
<td></td>
</tr>
<tr>
<td>Unique sites after duplicates (n=57)</td>
<td>23 (40)</td>
<td>15 (26)</td>
<td>6 (11)</td>
<td>13 (23)</td>
<td></td>
</tr>
<tr>
<td>Eligibility criteria not met</td>
<td>3 (37)</td>
<td>5 (63)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Journal papers (n=8)</td>
<td>2 (50)</td>
<td>—</td>
<td>1 (25)</td>
<td>1 (25)</td>
<td></td>
</tr>
<tr>
<td>Book chapters (n=4)</td>
<td>—</td>
<td>3 (100)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Patent applications (n=3)</td>
<td>18 (43)</td>
<td>7 (17)</td>
<td>5 (12)</td>
<td>12 (28)</td>
<td></td>
</tr>
<tr>
<td>Included websites (n=42)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

aNot available.
Table 2. Quality, readability, and search term frequency.

<table>
<thead>
<tr>
<th>Website category</th>
<th>DISCERN, median (IQR)</th>
<th>JAMA benchmark criteria, median (IQR)</th>
<th>Readability, mean (SD)</th>
<th>Search term word frequency, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Item 1 (clear aims)</td>
<td>Benchmark 1 (authorship)</td>
<td>Flesch reading ease score</td>
<td>“Spinal” frequency</td>
</tr>
<tr>
<td></td>
<td>Item 2 (achieved aims)</td>
<td>Benchmark 2 (attribution)</td>
<td>47.13 (11.54)</td>
<td>25.67 (16.02)</td>
</tr>
<tr>
<td></td>
<td>Item 3 (relevant)</td>
<td>Benchmark 3 (currency)</td>
<td>52.19 (7.84)</td>
<td>14.33 (13.39)</td>
</tr>
<tr>
<td></td>
<td>Item 4 (information source)</td>
<td>Benchmark 4 (disclosure)</td>
<td>45.73 (10.66)</td>
<td>2.79 (2.07)</td>
</tr>
<tr>
<td></td>
<td>Item 5 (publication date)</td>
<td>Benchmark (total)</td>
<td>43.80 (16.69)</td>
<td>13.92 (13.03)</td>
</tr>
<tr>
<td></td>
<td>Item 6 (balanced and unbiased)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>18.65 (19.22)</td>
</tr>
<tr>
<td></td>
<td>Item 7 (additional information sources)</td>
<td></td>
<td>40.00 (19.4)</td>
<td>12.00 (15.56)</td>
</tr>
<tr>
<td></td>
<td>Item 8 (areas of uncertainty)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>11.00 (12.73)</td>
</tr>
<tr>
<td></td>
<td>Item 9 (defines treatment)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>4.07 (2.18)</td>
</tr>
<tr>
<td></td>
<td>Item 10 (describes benefits)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>23.67 (14.51)</td>
</tr>
<tr>
<td></td>
<td>Item 11 (describes risks)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>3.94 (2.26)</td>
</tr>
<tr>
<td></td>
<td>Item 12 (no treatment or usual care outcomes)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>“Stimulator” frequency</td>
</tr>
<tr>
<td></td>
<td>Item 13 (effect on quality of life)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>13.83 (13.27)</td>
</tr>
<tr>
<td></td>
<td>Item 14 (alternative treatment choices)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>“Stimulator” (%)</td>
</tr>
<tr>
<td></td>
<td>Item 15 (shared decision-making support)</td>
<td></td>
<td>60.00 (19.94)</td>
<td>1.70 (1.48)</td>
</tr>
<tr>
<td></td>
<td>Item 16 (overall)</td>
<td>HON code of conduct certificate, n (%)</td>
<td>60.00 (19.94)</td>
<td>“Spinal” (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 (33)</td>
<td>60.00 (19.94)</td>
<td>25.67 (16.02)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 (33)</td>
<td>60.00 (19.94)</td>
<td>14.33 (13.39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 (5)</td>
<td>60.00 (19.94)</td>
<td>2.79 (2.07)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0 (0)</td>
<td>60.00 (19.94)</td>
<td>13.92 (13.03)</td>
</tr>
</tbody>
</table>

**Readability**

A summary of readability is provided in Table 2 and Table S1 in Multimedia Appendix 1. Readability indices across all sites indicated that information was difficult on average and required a 9th- to 10th-grade level of reading comprehension (reading ease: mean 48.4, SD 10.8 and grade: mean 9, SD 2.7). News or media required the highest level of reading comprehension (mean 15.2, SD 9.3), and personal or blog sites required the lowest level of reading comprehension (mean 6.7, SD 3.5).

---

*a* JAMA: Journal of the American Medical Association.

*b* HON: Health on the Net Foundation.
Quality
A summary of information quality (ie, DISCERN, JAMA scores, and HON code of conduct certification) is provided in Table 2 and Figure S2 in Multimedia Appendix 1. Based on DISCERN rating criteria, just over one-third of all sites were rated as having high overall quality (n=16, 38%), with the remainder rated as having moderate (n=19, 45%) or low (n=7, 16%) overall quality. Scientific resources (median 3.5, IQR 2-3) and news or media sites (median 2, IQR 2-2) showed the highest and lowest overall DISCERN scores, respectively. Few sites met all 4 JAMA benchmark criteria (n=3, 7%). Nonprofit (median 2.75, IQR 2-3) and for-profit sites (median 1.25, IQR 1-1.9) showed the highest and lowest benchmark totals, respectively. HON code of conduct certification was present in less than a quarter of all sites (n=7, 16%), with the majority attributed to nonprofit sites (n=4, 9%).

Website Characteristics
Tabulated website characteristics are summarized in Tables S3 and S4 in Multimedia Appendix 1. Publication dates (or latest revision date) were reported for less than half (19/42, 45%) of all websites and was proportionately lowest among for-profit sites (5/20, 25%) compared to other site types. Author names, aliases, and credentials were provided in less than half of all sites (20/42, 47%), and authorship disclosure was also proportionately lowest among for-profit sites (3/20, 15%) compared to others. Site addresses included designation in the United States (30/42, 71%), Canada (1/42, 2%), and Australia (1/42, 2%) or did not provide a location (10/42, 23%). Relative to other site types, the proportion of sites that provided a geographical designation was highest among for-profit sites (19/42, 45%). Just under half of all websites provided alternate content accessibility options (ie, PDF, audio, and video; 20/42, 47%). Some sites also featured pop-up notifications and advertisements (9/42, 21%). On average, nonprofit sites contained the greatest number of reference citations (mean 11.1, SD 27.1) followed by scientific resources (mean 1.2, SD 2.9).

Website Information
Website information is summarized in Tables S3 and S5 in Multimedia Appendix 1. Most websites were intended for consumers and caregivers (n=27, 64%), with the remainder intended to serve as resources for clinicians and researchers (n=1, 2%) or both (n=14, 33%). Site information pertaining to the use of ESCS (n=37, 88%), both transcutaneous SCS and ESCS (n=2, 4%), or did not define SCS type (n=3, 7%). Sites also provided summary information or a definition of SCS (n=37, 88%) and described SCS subcategories (n=14, 33%), indications (n=38, 90%), contraindications (n=14, 33%), side effects or risks (n=28, 66%), device considerations (n=25, 59%), follow-up (n=22, 52%), and outcome-related information (n=31, 73%). Most sites featured physician or clinical referral (n=26, 61%). Relative to other site types, for-profit sites showed the highest proportion of referrals (17/20, 85%).

Discussion
Principal Findings
This study provides a novel qualitative synthesis of publicly available web-based information on the topic of SCS. For-profit websites were the most commonly encountered. Website quality ratings varied between site types suggesting that the reporting (or omission) of important quality criteria (eg, authorship, information source, and publication date) was site-type dependent. For-profit sites were rated as having relatively lower quality than nonprofit or scientific resource sites based on quality criteria (ie, JAMA benchmarks and HON code of conduct certification). The proportion of for-profit sites reporting authorship was comparatively less than nonprofit and scientific resource sites. The average number of peer-reviewed reference citations identified within for-profit sites was also comparably less frequent than that of nonprofit and scientific resource sites. These findings are largely consistent with other studies of web-based information that report similar discrepancies in the provision of authorship and information sources [17,18]. This suggests that consumers may be exposed to web-based information on SCS that originates from unidentified authors and sources. While two-thirds of all websites mentioned potential risks or side effects of SCS, only one-third described possible contraindications for SCS. Moreover, the majority of websites specifically targeted consumers or their caregivers and featured clinical or physician referral for SCS. Relative to other site types, for-profit sites constituted the largest proportion of referrals indicating potential information bias and consumer-directed marketing of SCS [13]. This is a cause for concern, as the selective provision of SCS-related information from undisclosed sources may potentially place consumers and patient groups at unnecessary risk and consequently diminish their capacity to make informed health care decisions.

Limitations
This study has several limitations. First, returned search results were limited by geographical location (ie, IP designation within the United States based on geographical trends for the search topic “spinal cord stimulator”) and may lack generalizability across regions and therefore introduce bias unintentionally. This study also incorporated the use of Baidu, the second-most commonly used search engine by international ranking at the time the search was conducted, in conjunction with other common search engines (ie, Google, Bing, and Yahoo). Although Baidu is the most commonly used web-based resource for medical information seekers following the withdrawal of Google from China in 2010 [81], the additive benefit of using this search engine and associated indices (eg, Baidu Index) to include a broader user demographic and potentially enhance search result generalizability remains inconclusive. For-profit websites were the most commonly encountered site type, representing approximately half of all included sites (n=20, 48%). However, it is unclear whether the developers of these sites allocated funding and resources to increase user traffic and enhance website visibility through the use of strategic search engine optimization techniques (eg, content management systems; schematic data structuring; and technical optimizations.
to improve URL structure, loading, navigation, and internal linking). Although the results of the word frequency summary potentially suggest the use of word frequency–based ranking algorithms, the way these search engines ultimately select which resources to display in a given list of returned search results remains unknown. Future studies may investigate whether academic institutions and government and nongovernment organizations (ie, scientific resources) that curate high-quality, evidence-based medical information actually benefit from devoting funds and organizational resources to increase website user traffic and visibility via technical optimization (eg, site content and search engine compatibility) or by other means (eg, endorsement by influential people and use of slogans, catchphrases, logographic designs, and multimedia to enhance site awareness and recognition and to demonstrate expertise, authority, and trustworthiness).

Several tools were used to provide a robust quality assessment of the information encountered on various websites (ie, DISCERN, JAMA, and HON code of conduct). Although this approach is largely consistent with previous studies of web-based medical information, among these assessments, there are both redundant components (eg, authorship, attribution, and disclosure) as well as quality criteria that are unique to each of the tools used (eg, financial disclosure and advertising). Future studies may consider the use of more recent assessments comprising these and other essential criteria for evaluating the quality of information encountered on websites and web-based resources.

Additionally, the data extracted reflects publicly accessible information from the first 2 pages of returned results for 2 top trending search terms. Though not exhaustive (ie, all possible websites providing information on SCS), the search methodology used was intended to simulate scenario-specific consumer behavior (ie, seeking SCS-related information). Future studies may consider an alternate methodological approach involving divergent user-behavior scenarios.

Conclusions
Publicly available web-based information on SCS mainly targets consumers and their caregivers. We found that almost all sites focused on surgically implanted SCS with less than half providing authorship details and publication dates. Most websites mentioned risks or side effects of SCS, yet few websites described possible contraindications. Sites curated by for-profit entities featured clinical or physician referral for SCS more often than other site types indicating implicit bias. This suggests that the public is likely to be exposed to incomplete, potentially misleading, or out-of-date information from unidentified sources regarding the use of SCS. For-profit sites were also encountered more frequently than nonprofit and scientific resource sites, suggesting the content and technical format of sites featuring higher-quality medical information are not optimized to enhance user traffic and search engine visibility. To avoid the spread of misinformation that could potentially harm certain consumer or patient groups, sites featuring evidence-based information on the current use of SCS should declare authorship, ensure the evidence is frequently updated, and consider using search engine optimization techniques that enable preferential site positioning for commonly used search terms.

Acknowledgments
TM is supported by postdoctoral fellowship awards jointly funded by Michael Smith Health Research BC, the Canadian Institutes for Health Research (CIHR), and the Rick Hansen Foundation (RHF). TK is supported by a Canada Graduate Scholarship through the CIHR and a WorkSafeBC research trainee award. MC-J is supported by a Paralyzed Veterans of America (PVA) postdoctoral fellowship award and the RHF. SS is supported by a PVA postdoctoral fellowship award, the RHF, and the Wings for Life Spinal Cord Research Foundation (WFLSCRF). RM is supported by a PVA postdoctoral fellowship award, a CIHR-funded fellowship award, and the RHF. CS is supported by a PVA postdoctoral fellowship award, a CIHR postdoctoral fellowship award, and the RHF. RS is supported by the WFLSCRF and the US Department of Defense. AK is the endowed chair in the Division of Physical Medicine and Rehabilitation within the Faculty of Medicine at the University of British Columbia and is supported through grant funding provided by WFLSCRF, International Spinal Research Trust, Praxis, US Department of Defense, RHF, Onward, SpineX, Convatec, and Coloplast. Laboratory resources used for conducting the study were purchased with prior funding support from the Canadian Foundation for Innovation and the British Columbia Knowledge Translation Foundation.

Authors’ Contributions
TM, RS, and AK conceptualized the study; TM, AH, TT, TK, and MC-J conducted the analyses; TM drafted the paper; and all authors reviewed and approved the final version.

Conflicts of Interest
RS is a paid consultant at SpineX Inc, a medtech startup building spinal cord stimulation devices. This author contributed to the conception and final review of the study and had no role in developing the study methodology nor in the identification, extraction, and interpretation of the study data. All other authors have no potential conflicts of interest to declare.

Multimedia Appendix 1
Supplementary figures and tables.
References


75. Miller et al. JMIR Public Health and Surveillance 2024 | vol. 10 | e50031 | p.893


**Abbreviations**

- **ESCS**: epidural spinal cord stimulation
- **HON**: Health on the Net Foundation
- **JAMA**: Journal of the American Medical Association
- **SCS**: spinal cord stimulation
Impacts of Smoking Ban Policies on Billiard Hall Sales in South Korea Using Objective Sales Information of a Credit Card Company: Quasi-Experimental Study

Jin-Won Noh, PhD; Jooyoung Cheon, PhD; Hohyun Seong, PhD; Young Dae Kwon, MD, PhD; Ki-Bong Yoo, PhD

1Division of Health Administration, College of Software and Digital Healthcare Convergence, Yonsei University, Wonju, Republic of Korea
2Department of Nursing Science, Sungshin Women's University, Seoul, Republic of Korea
3College of Nursing, Keimyung University, Daegu, Republic of Korea
4Department of Humanities and Social Medicine, College of Medicine and Catholic Institute for Public Health and Healthcare Management, The Catholic University of Korea, Seoul, Republic of Korea

Corresponding Author:
Ki-Bong Yoo, PhD
Division of Health Administration
College of Software and Digital Healthcare Convergence
Yonsei University
Yonseidae gil 1
Wonju, 26493
Republic of Korea
Phone: 82 33 760 2418
Email: ykbong@yonsei.ac.kr

Abstract

Background: Smoking ban policies (SBPs) are potent health interventions and offer the potential to influence antismoking behavior. The Korean government completely prohibited smoking in indoor sports facilities, including billiard halls, since the government revised the National Health Promotion Act in December 2017.

Objective: This study aimed to examine the impact of the SBP on the economic outcomes of indoor sports facilities, particularly billiard halls.

Methods: This study used credit card sales data from the largest card company in South Korea. Data are from January 2017 to December 2018. Monthly sales data were examined across 23 administrative neighborhoods in Seoul, the capital city of South Korea. We conducted the interrupted time series model using the fixed effects model and the linear regression with panel-corrected standard errors (PCSE).

Results: The sales and transactions of billiard halls were not significantly changed after the introduction of the SBP in the full PCSE models. The R2 of the full PCSE model was 0.967 for sales and 0.981 for transactions.

Conclusions: The introduction of the SBP did not result in substantial economic gains or losses in the sales of billiard halls. In addition to existing price-based policies, the enhanced SBP in public-use facilities, such as billiard halls, can have a positive synergistic effect on reducing smoking prevalence and preventing secondhand smoke. Health policy makers can actively expand the application of SBPs and make an effort to enhance social awareness regarding the necessity and benefits of public SBPs for both smokers and the owners of hospitality facilities.

(JMIR Public Health Surveill 2024;10:e50466) doi:10.2196/50466

KEYWORDS
smoking ban policy; indoor sports facility; South Korea
Introduction

Background

It is well-known that exposure to secondhand smoke causes illness and death. The World Health Organization has estimated that tobacco smoking kills 7 million people per year globally, of which 890,000 are due to secondhand smoke [1]. In the United States, the prevalence of secondhand smoke exposure among nonsmokers diminished between 1988 and 2014, from 87.5% to 25.2%. However, there was no change in exposure between 2011-2012 and 2013-1014 periods, and about 1 in 4 nonsmokers were still exposed to secondhand smoke during the 2013-2014 period [2].

Previous studies have reported that South Korea has a high prevalence of tobacco smoking [3,4], but the prevalence of smokers had markedly decreased by 2021 [5]. However, the prevalence is still high compared to other Organisation for Economic Co-operation and Development member countries [6], which means that people are considerably exposed to health threats and risks of secondhand smoke.

Between 2007 and 2018, the exposure rate to secondhand smoke among Korean adults decreased by 10.7%, and the exposure rate to secondhand smoke in indoor working areas decreased by 34.5% [6]. The reduction in the exposure rates to secondhand smoke in indoor working areas and public regions was especially prominent after 2012, most probably due to the continuous expansion of nonsmoking zones [7].

The smoking ban policy (SBP) is a potent health intervention, offering the potential to influence antismoking behavior. There has been an increase in the number of SBPs in countries globally, including Australia, England, and the United States, aligning with an increase in knowledge about the risk of secondhand smoke [8]. Indeed, the introduction of SBP leads to a decrease in exposure to secondhand smoke, improves indoor air quality, protects workers, reduces adult and youth smoking levels, decreases hospitalizations for acute myocardial infarctions, and promotes respiratory health [9-11]. Notwithstanding the benefits of SBP, owners of hospitality facilities, including restaurants, bars, and billiard halls, have vigorously opposed the policy to curb smoking in these places, arguing that SBPs will result in economic hardship for them. This argument suggests that a complete ban on smoking in these places would discourage people from dining out, potentially negatively affecting sales. However, there is much evidence from the United States, Korea, Australia, and European countries indicating that economic performance was not affected by SBPs [12-15].

In accordance with the global trends of implementing SBP, indoor sports facilities (eg, billiard halls) in South Korea were regulated by the SBP as completely nonsmoking areas since the Korean government revised the National Health Promotion Act to prohibit smoking in all indoor spaces in December 2017 [16]. This change in SBP has led owners of indoor sports facilities to feel that the policy may negatively influence their economic profit, despite no significant change in sales. The effectiveness of SBPs has been evaluated by studies in other countries, demonstrating whether SBP affects economic profit. Previous works have consistently highlighted the impacts of SBPs on sales in various indoor places, such as restaurants and bars. For example, a previous study that included a sample of all 88 counties in the state of Ohio demonstrated that there was no significant difference in bar and restaurant sales following a statewide SBP between border regions in Ohio and nonborder areas [17]. The SBP in Ohio did not differentially influence the sales revenue for bars and restaurants located in counties where the border is shared with 5 other non–smoke-free states, compared to those in nonborder counties. Another study [18] supported the evidence that the SBP did not significantly affect facility sales, as the overall impact on sales in bars was negligible. The SBP was related to an increase in sales in medium to large bars in the rural region of Ireland and a small reduction in sales among large bars in the urban areas. These findings from previous works support the evidence for justification of continued use of SBPs to prevent the general public from exposure to secondhand smoke. However, there is a lack of evidence of the effects of SBP on business revenues in indoor sports facilities, even though many studies have been steadily involved in such research based on other indoor places. Moreover, it is crucial to assess the effect of the SBP, determining whether it resulted in a positive or negative economic impact. The results of the economic impact are important to provide evidence to visitors and owners of indoor sports facilities.

Objective and Hypotheses

To date, there has been little study on changes in business revenues of indoor sports facilities, especially billiards halls since the introduction of the SBP in South Korea. This study aimed to examine the impact of the SBP on the economic outcomes of indoor sports facilities, specifically billiard halls, using actual revenue data from the largest card company in South Korea. Based on the evidence that hospitality facilities’ sales were not affected by SBPs [14,19,20], we hypothesized that the introduction of the SBP does not significantly affect billiard halls’ sales.

Methods

Data

This study used sales data from the Shinhan Card Big Data Center. The data included Shinhan’s credit, debit, and check card sales information from January 2017 to December 2018. Shinhan Card holders were 12 million in 2015, representing 44.6% of the economically active population in South Korea [21,22]. Shinhan Card has the largest market share (21.7%) in South Korea as of 2017 [23]. In 2016, a total of 80% of all private consumption in South Korea was made through card payments. A payment method survey in South Korea reported that 94% of Seoul citizens had 1 or more debit or check cards in 2013 [24]. The cash transactions were not included in our data, but the correlation between sales information provided by Shinhan Card data and the retail sales information of Statistics Korea was 0.92 [25]. The correlation between the sales information from all card companies in South Korea and the data of Shinhan
Card was 0.97 [25]. Our data are suitable for assessing the effect of the policy.

In total, 3 districts in Seoul were selected for this analysis; Nowon district (533,498 population in 2019; 35.44 km²), Secho district (430,697 population in 2019; 46.98 km²), and Songpa district (675,843 population in 2019; 33.88 km²). The population in Seoul was 9,729,107 in 2019, and the 3 districts selected for this study account for 16.9% of Seoul’s population and cover 19.2% of Seoul’s total area. Regional experts at the Seoul Institute selected the districts considering various factors, including the percentage of the population aged 20-59 years, health behaviors (eg, smoking, drinking, and obesity rates), income level, environmental factors (eg, park space per capita and the number of designated smoking areas), and the similarity of z score for each factor with the average values for Seoul.

The unit of analysis for this study was neighborhood-month. Neighborhood in South Korea is referred to as “dong,” which is a submunicipal-level administrative unit of a city. We aggregated the individual billiard hall data into the neighborhood level. As the unit of analysis does not involve human subjects, this study does not require approval from an institutional review board.

Ethical Considerations
This research used aggregated sales data from billiard halls by region. It is not subject to ethical considerations.

Variables
The dependent variables were sales per neighborhood-month and transactions per neighborhood-month. The total sales information was aggregated from credit, debit, and check card use. A total of 1100 Korean won (KRW) was exchanged for US $ 1.

Based on previous studies on retail sales [14,26,27], the following factors were considered as the independent variables: socioeconomic factors of customers and region, seasonal factors, weather factors, employee factors, and overall economic status. The data consisted of neighborhood-month, making it impossible to consider the characteristics of individual customers and specifics of the store. Therefore, regional socioeconomic factors, seasonal factors, and economic factors were included as the independent variables in our study. The research model and control variables are shown in Figure 1.

Neighborhood total sales size represented the overall economic size of the neighborhood. It summarized all monthly credit, debit, and check card use in all business types except for online shopping, university tuition, insurance fees, taxes, and utility bills. As revenue is largely influenced by the total sales size of the neighborhood, it is an important covariate to analyze the billiard halls’ sales. Season and number of holidays per month were included as seasonal factors. Seasons were classified into spring, summer, fall, and winter because customers prefer to visit indoor facilities during summer and winter. Some billiard halls close on holidays, while others are crowded on holidays. The Composite Index of Business Indicators was included to adjust the overall economic condition in South Korea [28].

Figure 1. Research model and control variables of this study.

Statistical Analysis
The Wilcoxon signed rank test was used to assess the mean difference between before and after introducing the policy. Our data were balanced panel data. Data constituted time-series–cross-section (TSCS) data, consisting of 24 months and 23 panels. Although the fixed effects (FE) model is commonly applied to analyze TSCS data, the assumptions of independence and identical distribution are prone to violation due to panel heteroskedasticity, contemporaneous correlation, serial correlation, and nonstationarity [29]. To identify these violations, we used various tests, including the Wooldridge test for serial correlation, the Pesaran cross-sectional dependence test for contemporaneous correlation, and the likelihood ratio
test using Wiggins and Poi’s method [30] for panel heteroskedasticity. With the results of these tests, panel heteroskedasticity, serial correlation, and contemporaneous correlation were observed to be significant.

Therefore, a regression model with panel-corrected standard errors (PCSE) was the most suitable approach for our data [31]. The PCSE model is known to provide robust estimation for TSCS data when \( T \geq 15 \). Since the number of panels and the number of time points are almost the same, we conducted an FE model with robust standard errors to assess the robustness although there was a contemporaneous correlation.

The interrupted time series model, a quasi-experimental analysis, was used in the analysis [32]. Interrupted time series is a well-known method to analyze the effects of policies. It provides a policy effect by comparing the actual outcome with the potential outcome assuming that the baseline trend would be extended if the policy were not introduced [33].

The full regression model is as follows:

\[
Y_{it} = \beta_0 + \beta_1 t + \beta_2 \text{policy} + \beta_3 D_{it} + \epsilon_{it}
\]

where \( Y \) is the dependent variable; \( t \) is the time period (month); \( \text{policy} \) is an indicator for the introduction of the policy introduced (0: before the introduction of the policy; 1: after the introduction of the policy); \( D \) is the dummy variables for neighborhood FEs; and \( \epsilon \) is the error term. \( \beta_2 \) and \( \beta_3 \) represent the effects of the policy. \( \beta_2 \) represents the level change due to the policy, and \( \beta_3 \) represents the trend change after the policy was introduced, compared to the baseline time trend (\( \beta_1 \)). The effects of the policy can be calculated by considering both \( \beta_2 \) and \( \beta_3 \) after the time point when the policy was introduced. For example, the 1-year effect of SBP is calculated as follows: 

\[
\beta_2 + \beta_3 \times 12.
\]

### Results

The general characteristics of the study data are shown in Table 1. The changes in sales information and neighborhood total sales size of billiard halls were insignificant in all 3 districts. Only the transactions of billiard halls in the Secho district significantly decreased, implied in the number of card payments. Since the sales in the Secho district did not change significantly, customers might be paying more per visit (Table 1).

Figure 2 shows the monthly sales trends of districts. The trends of billiard halls’ monthly sales in the 3 districts were almost flat (Figure 2).

#### Table 1. General characteristics of the study data by districts before and after the smoking ban policy.

<table>
<thead>
<tr>
<th>Variables (in billiard halls)</th>
<th>Districts</th>
<th>Variables (in billiard halls)</th>
<th>Districts</th>
<th>Variables (in billiard halls)</th>
<th>Districts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Songpa (n=12a)</td>
<td></td>
<td>Songpa (n=12a)</td>
<td></td>
<td>Songpa (n=12a)</td>
</tr>
<tr>
<td>Log (sales per neighborhood-month: US $100), mean (SD)</td>
<td>Before</td>
<td>After</td>
<td>( P ) value</td>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>Log (transactions per neighborhood-month), mean (SD)</td>
<td>7.3 (0.7)</td>
<td>7.2 (0.7)</td>
<td>.44</td>
<td>7.7 (1.3)</td>
<td>7.6 (1.7)</td>
</tr>
<tr>
<td>Log (neighborhood total sales size: US $1 million), mean (SD)</td>
<td>5.7 (1.1)</td>
<td>5.7 (1.2)</td>
<td>&gt;.99</td>
<td>5.9 (1.3)</td>
<td>5.7 (1.7)</td>
</tr>
</tbody>
</table>

\( ^a \)The number of neighborhoods.
Figure 2. The trends of the monthly sales of billiard halls from January 2017 to December 2018.

Table 2 shows the results of FE and PCSE regressions. PCSE regression with FE terms and control variables showed the highest $R^2$ results in both the sales and transactions models. Our variables of interest were SBP and the month after the policy. In model 5, with the highest $R^2$, the SBP’s coefficient ($\beta_2$) was 0.0767. It shows that the dependent variable increased by 0.0767 constantly after the SBP implementation. The coefficient for the month after the policy ($\beta_3$) was –0.0123. It showed that the dependent variables decreased by 0.0123 every month after the policy was introduced. However, neither variable was significant in all models for sales. The month after the policy, variables in transaction models were significant in models 1 and 2 (FE) as well as model 4 (PCSE), but they were not significant in model 5 (full PCSE). There was little evidence that the sales were affected by the SBP in billiard halls.

The Composite Index of Business Indicator was not significant for both sales and transactions, but it showed a positive relationship. This suggests that there may have been no significant macroeconomic issues from 2017 to 2018. The log of the neighborhood total sales size was significant for both sales and transactions; it showed that the billiard hall business was strongly affected by the economic status of its location. Compared to spring, only sales in winter were significantly higher. However, transactions were not significant in winter, compared to spring. It meant that people were likely to visit billiard halls and stay longer in the winter season. Transactions in summer were significantly higher than in spring ($P=.049$). The number of holidays did not significantly impact both sales and transactions.
Table 2. Effects of the smoking ban policy on the log of monthly sales and the log of monthly transactions of billiard halls (N=552).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Log (monthly sales)</th>
<th>Log (monthly transactions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month</td>
<td>β 1</td>
<td>β 2</td>
</tr>
<tr>
<td></td>
<td>.0038</td>
<td>.0289</td>
</tr>
<tr>
<td></td>
<td>.46</td>
<td>.54</td>
</tr>
<tr>
<td>Smoking ban policy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>0.54</td>
</tr>
<tr>
<td>Month after the policy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.07</td>
</tr>
<tr>
<td>Composite Index of Business Indicator</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.77</td>
</tr>
<tr>
<td>Log (neighborhood total sales size)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Summer</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.07</td>
</tr>
<tr>
<td>Fall</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.30</td>
</tr>
<tr>
<td>Winter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.02</td>
</tr>
<tr>
<td>Number of holidays</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>.32</td>
</tr>
<tr>
<td>Adjusting regional FE</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>.0595</td>
<td>.081</td>
</tr>
</tbody>
</table>

aFE: fixed effects.
bPCSE: panel-corrected standard errors.
cNot applicable.

Discussion

**Principal Findings**

Despite the concerns of many people about the negative impacts of the SBPs on sales of indoor working areas [14,34], this study found that the sales and transactions in billiard halls were not affected by the SBP introduced in 2017 in South Korea. This finding supports previous research demonstrating that SBPs had no negative economic impact on sales of restaurants and bars in South Korea and other countries [14,34,35].

The first of the 3 reasons for no negative economic impacts on sales of billiard halls is that the social awareness of the need for public SBPs to prevent the harms of secondhand smoke has been increased due to mass media campaigns among both smokers and nonsmokers [36-39]. As smoking in public places becomes increasingly stigmatized, smokers may increasingly...
become aware that nonsmokers have the right to object to exposure to harmful passive smoking [36,40,41]. The second reason may be due to changes in the smoking population and increased preferences for no-smoking areas. The smoking prevalence among Korean adults aged 219 years decreased from 27.5% in 2010 to 20.6% in 2020 [42]. Smoking prevalence among men aged 30-50 years, who were the dominant population of smokers, decreased especially after 2015, when tobacco prices were raised from KRW 2500 (US $2.1) to KRW 4500 (US $3.8) and indoor smoking was banned in all businesses and restaurants [42-44]. A study found that Korean smokers in 2016 reported more positive perceptions of the effectiveness of expanded smoking bans and smoke-free policies compared to smokers in 2010 [40]. Therefore, the number of smokers who may complain regarding SBPs in billiard halls has decreased, and smokers who prefer smoke-free environments may still visit the billiard halls despite knowing that they are nonsmoking areas.

The third reason may be related to the indoor smoking room. According to a study conducted between 2018 and 2019, a total of 87% of billiard halls have indoor smoking rooms [41]. Based on the National Health Promotion Act in Korea, smoking rooms can be installed inside and outside of facilities, even if the facilities are smoking-free areas. Most smokers could use indoor smoking rooms despite the SBP in billiard halls, which may have resulted in no change in the sales of billiard halls. If there are no indoor smoking rooms or if rooms are far away from the playing area, smokers may decide not to smoke and focus on playing. The Ministry of Health and Welfare in South Korea reported that nonsmokers are more likely to be exposed to secondhand smoke in indoor public places with indoor smoking rooms and recommends closing indoor smoking rooms in all public facilities by 2025 [45]. Future research should examine the economic impact on indoor facilities and the consequences of secondhand smoke following the closure of indoor smoking rooms.

There are several additional benefits related to the SBP in billiard halls. First, SBPs reduce exposure to secondhand smoke, improve health outcomes, and reduce mortality due to smoking-related illnesses for both smokers and nonsmokers [36,46]. Second, smoking restrictions may lead to changes in smoking behavior among smokers, as they should spend additional time smoking due to SBPs, leading to an increase in quit attempts [46]. Third, the SBPs in billiard halls can lead to enhanced positive perceptions of the effectiveness of expanded smoking-free areas. A study found that past smokers and nonsmokers among owners, workers, and users in billiard halls and indoor golf clubs were more in favor of smoke-free areas after the SBP in 2017 compared to before the SBP was implemented [41].

Previous studies stressed price-based policy as the most effective means of reducing the consumption of tobacco [38,44], but the level of price increase in 2015 was insufficient to lead to a noticeable difference in South Korea [44,47]. Therefore, SBPs in public facilities, such as billiard halls, together with price-based policies, have positive synergistic effects in reducing smoking prevalence and preventing secondhand smoke [35,38,40,41]. This study has the strength of examining the impact of the SBP on billiard halls’ economic outcomes using actual revenue data from the largest card company in South Korea to provide a basis for enhancing the SBP.

However, there are some concerns in interpreting the findings of this study. First, this study could not adjust for the presence of indoor smoking rooms in billiard halls, which may be related to sales and transactions in these establishments. Future studies could compare sales between the billiard halls with and without indoor smoking rooms, or, in the case of closing indoor smoking rooms, compare sales before and after closure to provide a more robust evidence base for smoke-free policies. In addition, the inability to control for individual preferences and accessibility to indoor smoking rooms is one of the limitations of this study. Smokers might prefer to play pool rather than smoke, even though they are aware of the ban in billiard halls. Alternatively, smokers may not be aware of the existence of an indoor smoking area, or even if they are, they may choose not to use it while playing pool. Therefore, future studies should include individual preferences and adherence to smoking and smoke-free areas as well as environmental constraints, such as the presence and accessibility of indoor smoking rooms, in their analyses to determine if smoke-free policies have an impact on sales.

Conclusions
This study examines the effects of the SBP in indoor sports facilities on billiard halls’ economic outcomes. Despite the worries of the owners of hospitality facilities, the SBP does not affect the sales of billiard halls. In addition to existing price-based policies, enhancing SBP in public use facilities, such as billiard halls, can have a positive synergistic effect on reducing smoking prevalence and preventing exposure to secondhand smoke. Based on this finding, health policy makers can actively expand the application of SBPs and make efforts to enhance social awareness of the need and benefits of public SBPs among both smokers and owners of hospitality facilities.

Acknowledgments
This research was supported by a grant of the Patient-Centered Clinical Research Coordinating Center (PACEN) funded by the Ministry of Health and Welfare, Republic of Korea (grant HC21C0059).

We used the generative AI tool, ChatGPT, only to correct English grammar.

Data Availability
The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.
References


5. Prevalence of tobacco use. Korea Health Promotion Institute. URL: https://nosmk.khepi.or.kr/ntcc/subIndex/547.do?text=Enforcement%20of%20strong%20tobacco%20controls,other%20OECD%20member%20ranking&ifth=1 [accessed 2024-03-28]


42. The 8th Korea national health and nutrition examination survey. KCDC. URL: http://www.mohw.go.kr/board. es?mid=a10503010100&bid=0027&act=view&list_no=351193&tag=&nP

43. Noh et alJMIR PUBLIC HEALTH AND SURVEILLANCE


Abbreviations

FE: fixed effects
KRW: Korean won
PCSE: panel-corrected standard errors

https://publichealth.jmir.org/2024/1/e50466

JMIIR Public Health Surveill 2024 | vol. 10 | e50466 | p.903

(page number not for citation purposes)
SBP: smoking ban policy
TCS: time-series–cross-section

©Jin-Won Noh, Jooyoung Cheon, Hohyun Seong, Young Dae Kwon, Ki-Bong Yoo. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Religion, Geography, and Risky Sexual Behaviors Among International Immigrants Living in China: Cross-Sectional Study

Yuyin Zhou1, MD; Feng Cheng2,3, PhD; Junfang Xu1, PhD

Corresponding Author:
Junfang Xu, PhD

Abstract

Background: Behavioral differences exist between countries, regions, and religions. With rapid development in recent decades, an increasing number of international immigrants from different regions with different religions have settled in China. The degrees to which sexual behaviors—particularly risky sexual behaviors—differ by religion and geographical areas are not known.

Objective: We aim to estimate the associations of religion and geographical areas with sexual behaviors of international immigrants and provide evidence for promoting the sexual health of international immigrants.

Methods: A cross-sectional study was conducted via the internet with a snowball sampling method among international immigrants in China. In our study, risky sexual behaviors included having multiple sexual partners and engaging in unprotected sex. Descriptive analysis was used to analyze the basic characteristics of international immigrants as well as their sexual behaviors, religious affiliations, and geographical regions of origin. Multivariate binary logistic regression analyses with multiplicative and additive interactions were used to identify aspects of religion and geography that were associated with risky sexual behaviors among international immigrants.

Results: A total of 1433 international immigrants were included in the study. South Americans and nonreligious immigrants were more likely to engage in risky sexual behaviors, and Asian and Buddhist immigrants were less likely to engage in risky sexual behaviors. The majority of the Muslims had sexually transmitted infection and HIV testing experiences; however, Muslims had a low willingness to do these tests in the future. The multivariate analysis showed that Muslim (adjusted odds ratio [AOR] 0.453, 95% CI 0.228 - 0.897), Hindu (AOR 0.280, 95% CI 0.082 - 0.961), and Buddhist (AOR 0.097, 95% CI 0.012 - 0.811) immigrants were less likely to report engaging in unprotected sexual behaviors. Buddhist immigrants (AOR 0.292, 95% CI 0.086 - 0.990) were also less likely to have multiple sexual partners. With regard to geography, compared to Asians, South Americans (AOR 2.642, 95% CI 1.034 - 6.755), Europeans (AOR 2.310, 95% CI 1.022 - 5.221), and North Africans (AOR 3.524, 95% CI 1.104 - 11.248) had a higher probability of having multiple sexual partners.

Conclusions: The rates of risky sexual behaviors among international immigrants living in China differed depending on their religions and geographical areas of origin. South Americans and nonreligious immigrants were more likely to engage in risky sexual behaviors. It is necessary to promote measures, including HIV self-testing, pre-exposure prophylaxis implementation, and targeted sexual health education, among international immigrants in China.

(JMIR Public Health Surveill 2024;10:e44616) doi:10.2196/44616

KEYWORDS

religion; geography; risky sexual behaviors; international immigrants; China; sexual behavior; immigrant

Introduction

Culture is a complicated phenomenon, and there are important cultural differences between different countries, regions, and religions. International immigrants undergo an acculturation process, which influences how well they can live a new life in their new country and may result in differences in individual sexual behaviors [1]. Traditionally, in most countries, religious individuals have conservative attitudes toward sexual behavior [2,3]. For example, Christianity prohibits adolescents and unmarried ones from having sexual intercourse [4,5]. Similarly, Muslims are only permitted to have sex within marriage [6]. Some international immigrants have strong and deep-rooted religious beliefs derived from their motherland, and some of these beliefs may influence their sexual behaviors. Consequently, some international immigrants may be less likely to engage in risky sexual behaviors compared to local nonreligious people. However, some studies have found opposite results, indicating that religion may be a risk factor for engaging in risky sexual behaviors. Many world religions promote doctrines that include negative views toward homosexual
behavior, multiple sexual partners, and premarital sexual behaviors. These doctrines may cause sexual guilt in individuals engaging in nonmonogamous sexual activity, which is associated with reduced contraceptive use [7,8]. In addition, a meta-analysis [9] focused on youths identified that religion was a protective factor for age at sexual debut and the number of sexual partners but had no association with contraceptive use. Individuals’ geographical region of origin also has associations with their sexual behaviors. It was reported that polygamy was one of the main factors contributing to risky sexual behaviors and HIV infection among Haitian immigrant women living in America [10]. Additionally, single Latina women were likely to have multiple sexual partners in America [11]. Moreover, Huang [12] indicated that the migration process, especially from middle- and low-income countries to high-income countries may increase the possibility of risky sexual behaviors.

With rapid development in recent decades, an increasing number of international immigrants have settled in China. The National Bureau of Statistics reported that approximately 845,697 international immigrants lived in China in 2020, representing an increase of 251,865 compared to 2010 [13,14]. However, international immigrants tend to be exposed to sexual vulnerability [15] for several reasons. The totally new living environment, including historical and cultural differences, huge language barriers, and lack of solid social networks among international immigrants can lead to social isolation, so they may engage in risky sexual behaviors to relax and spend spare time [16]. Meanwhile, the lack of community supervision brings sexual freedom among these immigrants, which may also increase their risky sexual behaviors as they seek mental and physical pleasure [17]. For instance, a previous study found that risky sexual behaviors among homosexual/bisexual male migrants from Central and Eastern Europe living in London had been significantly influenced by the process of migration. Specifically, disentangled from the conventional frameworks of social regulation prevalent in their countries of origin and with expanded access to gay venues in London led to a notable escalation in their sexual engagement [18]. Therefore, it has also been reported that migrants are an at-risk population vulnerable to contracting sexually transmitted infections (STIs) or HIV [19-21]. However, rates of engagement in risky sexual behaviors may differ by religion and geographical region of origin. This study investigated the associations of religion and geographical region of origin with the sexual behaviors of international immigrants living in China. Findings from this study can generate evidence to assist the promotion of sexual health and implement programs for target populations at high risk of sexual behaviors.

**Methods**

**Participants**

A cross-sectional study was conducted via the internet during the COVID-19 pandemic between January 2021 and September 2021. Considering the cultural differences and sensibility of sexual issues, the respondent-driven sampling method was used to access international immigrants living in China. First, 3 - 6 international immigrants with plenty of social contacts living in different cities in China were chosen as the initial seed participants, and these seed participants had different occupations, ages, geographical regions of origin, and genders. After introducing the inclusion criteria of participants to every seed participant, they were required to recommend 1 - 3 international immigrants. If a seed participant did not recruit any other participants within 2 weeks or recruited fewer than 5 participants over 2 months, a new seed participant corresponding to its sociodemographic characteristics would be developed to ensure survey continuity and seed diversity.

Guangdong province, Zhejiang province, and Beijing were chosen as the study sites. We selected these sites because the Seventh Population Census of China in 2022 showed that these provinces or municipalities were included in the top 10 provinces or municipalities in terms of the size of international immigrants in China [22]. Guangdong is the province with the largest number of international immigrants in China, with more than 410,000 international immigrants residing there. Guangdong’s open policy and diverse social environment provide good living and working conditions for international immigrants. Zhejiang Province has a developed economy, convenient transportation, and many private enterprises, with more than 46,000 international immigrants living there. Moreover, Yiwu in Zhejiang province is one of the largest commodity trading cities, attracting nearly half a million international immigrants every year [23]. Beijing is the capital and the political, economic, and cultural center of China, with over 62,000 international immigrants. Foreign embassies, offices of major foreign companies in China, commercial and trade organizations, and foreign students are all gathered in Beijing.

The inclusion criteria of the participants were the following: (1) their home country is not China, but they currently live in China for more than 3 months; (2) aged ≥18 years; (3) able to read in English; and (4) willing to participate in the study. The exclusion criteria were the following: (1) short stay in China for business, travel, and other purposes and (2) failure to pass the “attention check” in the questionnaire, which was set to identify the careless respondents and improve the data quality. PASS 2021 (NCSS LLC) software was used to calculate the sample size. Previous studies [24,25] showed that the prevalence of HIV among international immigrants was 1.70% - 62.0%, so the prevalence of HIV was selected as 31.85% in this study with an α error set at .05, implying a 3% margin of error. To ensure the validity of the questionnaire, we increased the sample size by 20%, which required 1150 participants. Finally, 1460 participants were surveyed, and 1433 eligible participants were included in this study.

**Data Collection**

All participants completed a web-based questionnaire in English, which included sociodemographic information (eg, home country, gender, age, marital status, religious beliefs, and education level), sexual behavior–related items, and the willingness to undergo testing.

The geographical regions of origin included sub-Saharan Africa, North America, South America, Europe, North Africa, Asia, and “others.” The religious beliefs included Christianity, Islam,
Hinduism, Buddhism, and “others,” and a group of participants who identified as nonreligious were also included.

Sexual behavior–related items included whether one had participated in sexual behavior before, sexual orientation, number of sexual partners, and the frequency of condom use. Risky sexual behaviors included having multiple sexual partners and unprotected sexual behaviors. Multiple sexual partners referred to engaging in sexual behaviors with at least 1 other person in addition to their spouse or stable partner within the last year. Unprotected sexual behaviors referred to ever having sex without a condom. The questions regarding sexual behaviors were designed based on the guidelines for the prevention of HIV/AIDS issued by the China Center for Disease Control and our previous studies [26-28].

The willingness to undergo testing included STI/HIV test experiences (ie, whether they had had an HIV test in the past), perceived risk of being infected with STIs/HIV, and willingness for STI/HIV testing in the future.

Statistical Analysis
The basic characteristics of international immigrants, and their sexual behaviors, religious beliefs, and geographical regions of origin were analyzed using descriptive statistics with frequencies and percentages. Multivariate binary logistic regression analyses were used to identify religion-related and geographical origin-related factors significantly associated with risky sexual behaviors among international immigrants. Additionally, in logistic regression analysis, we took having multiple sexual partners and unprotected sexual behaviors as the dependent variables. We considered age, gender, marital status, education level, employment status, annual disposable income, STI/HIV testing experience, perceived risk of being infected with HIV, and willingness for STI/HIV testing in the future variables as covariates for conducting a multiplicative and additive interaction analysis to further explore the interaction between religion and geography. The statistical software R (version 4.2; R Foundation for Statistical Computing) and SPSS (version 23.0; IBM Corp) were used to analyze all data, and \( P < .05 \) was considered to indicate statistical significance.

Ethical Considerations
The study protocol and consent procedure were approved by the Ethics Review Committee, School of Public Health, Zhejiang University (#2019 - 064). Informed consent information was provided before the questions; participants had the option to exit the survey after reading the informed consent information or to provide consent to continue. The confidentiality of individuals was properly protected in the management of the investigation and the processing of data.

Results

Basic Characteristics of International Immigrants
As shown in Table 1, a total of 1433 international immigrants living in China were included in the study, with the mean age of 24.97 (SD 4.57) years. The majority were male (n=973, 67.90%), and more than half were unmarried (n=1263, 88.14%). Participants who had completed over 12 years of education (n=831, 57.99%) and their annual disposable income was at most 50,000 yuan (n=1040, 72.58%) accounted for the largest percentage in all groups. The participants’ religions included Christianity (n=641, 44.73%), Islam (n=565, 39.43%), Hinduism (n=73, 5.09%), Buddhism (n=47, 3.28%), and others (n=11, 0.77%), and 95 (6.63%) had no religion. The geographical regions of origin included sub-Sahara Africa (n=665, 46.41%), North America (n=44, 3.07%), South America (n=23, 1.61%), Europe (n=40, 2.79%), North Africa (n=14, 0.98%), Asia (n=610, 42.57%), and others (n=6, 0.42%). Moreover, 37.68% (n=540) of the participants had risky sexual behaviors before.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>24.97 (4.57)</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td></td>
</tr>
<tr>
<td>≤20</td>
<td>90 (6.28)</td>
</tr>
<tr>
<td>21 - 30</td>
<td>903 (63.01)</td>
</tr>
<tr>
<td>31 - 40</td>
<td>306 (21.35)</td>
</tr>
<tr>
<td>&gt;40</td>
<td>134 (9.35)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>973 (67.90)</td>
</tr>
<tr>
<td>Female</td>
<td>460 (32.10)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>1263 (88.14)</td>
</tr>
<tr>
<td>Married</td>
<td>110 (7.68)</td>
</tr>
<tr>
<td>Widowed</td>
<td>4 (0.28)</td>
</tr>
<tr>
<td>Divorced</td>
<td>9 (0.63)</td>
</tr>
<tr>
<td>Others</td>
<td>47 (3.28)</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
</tr>
<tr>
<td>Illiteracy</td>
<td>49 (3.42)</td>
</tr>
<tr>
<td>1 - 5 years</td>
<td>299 (20.87)</td>
</tr>
<tr>
<td>6 - 10 years</td>
<td>97 (6.77)</td>
</tr>
<tr>
<td>11 - 12 years</td>
<td>157 (10.96)</td>
</tr>
<tr>
<td>&gt;12 years</td>
<td>831 (57.99)</td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>231 (16.12)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1202 (83.88)</td>
</tr>
<tr>
<td>Annual disposable income (¥), n (%)</td>
<td></td>
</tr>
<tr>
<td>≤50,000</td>
<td>1040 (72.58)</td>
</tr>
<tr>
<td>50,001 - 100,000</td>
<td>212 (14.79)</td>
</tr>
<tr>
<td>100,001 - 150,000</td>
<td>79 (5.51)</td>
</tr>
<tr>
<td>&gt;150,000</td>
<td>102 (7.12)</td>
</tr>
<tr>
<td>Religion, n (%)</td>
<td></td>
</tr>
<tr>
<td>Christianity</td>
<td>641 (44.73)</td>
</tr>
<tr>
<td>Islam</td>
<td>565 (39.43)</td>
</tr>
<tr>
<td>Hinduism</td>
<td>73 (5.09)</td>
</tr>
<tr>
<td>Buddhism</td>
<td>47 (3.28)</td>
</tr>
<tr>
<td>Others</td>
<td>11 (0.77)</td>
</tr>
<tr>
<td>No religion</td>
<td>95 (6.63)</td>
</tr>
<tr>
<td>Geography, n (%)</td>
<td></td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>665 (46.41)</td>
</tr>
<tr>
<td>North America</td>
<td>44 (3.07)</td>
</tr>
<tr>
<td>South America</td>
<td>23 (1.61)</td>
</tr>
<tr>
<td>Europe</td>
<td>40 (2.79)</td>
</tr>
<tr>
<td>North Africa</td>
<td>14 (0.98)</td>
</tr>
</tbody>
</table>
Sexual Behaviors Among International Immigrants by Religion and Geography

Table 2 shows the distributions of risky sexual behaviors among international immigrants by religion. The rates of participants with multiple sexual partners were as follows: 30.53% (n=29) among nonreligious individuals, 27.93% (n=179) among Christians, 16.64% (n=94) among Muslims, 16.64% (n=12) among Hindus, and 8.51% (n=4) among Buddhists. The rates of unprotected sexual behaviors were as follows: 22.11% (n=21) among participants without religion, 17.63% (n=113) among Christians, 8.32% (n=47) among Muslims, 5.48% (n=4) among Hindus, and 4.26% (n=2) among Buddhists. Table 3 shows the distributions of risky sexual behaviors among international immigrants by geographical region of origin. The rates of participants who reported having multiple sexual partners were as follows: 52.17% (n=12) from South America, 35.71% (n=5) from North Africa, 32.50% (n=13) from Europe, 22.73% (n=10) from North America, 27.22% (n=181) from sub-Saharan Africa, and 15.74% (n=96) from Asia. Engaging in unprotected sexual behaviors was reported as follows: 30.43% (n=7) in those from South America, 20% (n=8) in those from Europe, 16.99% (n=113) in those from sub-Saharan Africa, 15.91% (n=7) in those from North America, 14.29% (n=2) in those from North Africa, and 7.87% (n=48) in those from Asia.

Table 2. Risky sexual behaviors of international immigrants by religion (N=1432a).

<table>
<thead>
<tr>
<th>Religion</th>
<th>Multiple sexual partners, n (%)</th>
<th>Unprotected sexual behaviors, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christianity (n=641)</td>
<td>179 (27.93)</td>
<td>113 (17.63)</td>
</tr>
<tr>
<td>Islam (n=565)</td>
<td>94 (16.64)</td>
<td>47 (8.32)</td>
</tr>
<tr>
<td>Hinduism (n=73)</td>
<td>12 (16.64)</td>
<td>4 (5.48)</td>
</tr>
<tr>
<td>Buddhism (n=47)</td>
<td>4 (8.51)</td>
<td>2 (4.26)</td>
</tr>
<tr>
<td>Others (n=11)</td>
<td>3 (27.27)</td>
<td>2 (18.18)</td>
</tr>
<tr>
<td>No religion (n=95)</td>
<td>29 (30.53)</td>
<td>21 (22.11)</td>
</tr>
</tbody>
</table>

aOne person was missed in the reporting.

Table 3. Risky sexual behaviors of international immigrants by geography.

<table>
<thead>
<tr>
<th>Geography</th>
<th>Multiple sexual partners, n (%)</th>
<th>Unprotected sexual behaviors, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-Saharan Africa (n=665)</td>
<td>181 (27.22)</td>
<td>113 (16.99)</td>
</tr>
<tr>
<td>North America (n=44)</td>
<td>10 (22.73)</td>
<td>7 (15.91)</td>
</tr>
<tr>
<td>South America (n=23)</td>
<td>12 (52.17)</td>
<td>7 (30.43)</td>
</tr>
<tr>
<td>Europe (n=40)</td>
<td>13 (32.50)</td>
<td>8 (20)</td>
</tr>
<tr>
<td>North Africa (n=14)</td>
<td>5 (35.71)</td>
<td>2 (14.29)</td>
</tr>
<tr>
<td>Asia (n=610)</td>
<td>96 (15.74)</td>
<td>48 (7.87)</td>
</tr>
<tr>
<td>Others (n=6)</td>
<td>2 (33.33)</td>
<td>2 (33.33)</td>
</tr>
</tbody>
</table>

The Willingness to Undergo Testing Among International Immigrants by Religion and Geography

Table 4 shows that 378 (58.97%) Christians, 469 (83.01%) Muslims, 58 (79.45%) Hindus, 3 (76.60%) Buddhists, and 65 (68.42%) nonreligious people had STIs/HIV test experiences. There were 348 (61.59%) Muslims and 46 (63.01%) Hindus who believed that it was impossible for them to be infected with STIs/HIV, which accounted for the largest percentages. As for the willingness to undergo STIs/HIV testing in the future, the willingness of Muslims was relatively low and the willingness of Christians was relatively high.
Table 5 shows that a total of 50 (83.11%) Asians, 399 (60%) sub-Saharan Africans, 20 (45.45%) North Americans, 14 (60.87%) South Americans, 29 (72.50%) Europeans, and 12 (85.71%) North Africans had STIs/HIV test experiences. There were 3 (13.04%) South Americans who believed that they were extremely likely to be infected with STIs/HIV, and 11 (78.57%) North Africans believed that it was impossible for them to be infected with STIs/HIV, which accounted for the largest percentages. As for the willingness to undergo STIs/HIV testing in the future, the willingness of Asians was relatively low and the willingness of sub-Saharan Africans was relatively high.
**Religion and Geography Factors Associated With Risky Sexual Behaviors in International Immigrants**

The results of binary logistic regression modeling performed to identify factors significantly associated with risky sexual behaviors are shown in Table 6. In multivariate analyses, Muslim (adjusted odds ratio [AOR] 0.453, 95% CI 0.228 - 0.897), Hindu (AOR 0.280, 95% CI 0.082 - 0.961), and Buddhist (AOR 0.097, 95% CI 0.012 - 0.811) immigrants were less likely to report engaging in unprotected sexual behaviors. Moreover, Buddhist (AOR 0.292, 95% CI 0.086 - 0.990) were also less likely to have multiple sexual partners. With regard to geography, compared to Asians, South Americans (AOR 2.642, 95% CI 1.034 - 6.755), Europeans (AOR 2.310, 95% CI 1.022 - 5.221), and North Africans (AOR 3.524, 95% CI 1.104 - 11.248) had a higher probability of having multiple sexual partners. shows the multiplicative interaction between religion and geography in relation to risky sexual behaviors, indicating that there was no significant multiplicative interaction. Additionally, with regard to having multiple sexual behaviors, the additive interaction analysis indicated that the relative excess risk due to interaction was −0.731 (−4.957 to 3.495), the attributable proportion due to interaction was −0.526 (−3.732 to 2.680), and the synergy index 0.347 (0.002-63.644). Similarly, with regard to having unprotected sexual behaviors, the relative excess risk due to interaction was 0.197 (-3.320-3.715), the attributable proportion due to interaction was 0.116 (–1.899 to 2.131), and the synergy index was 1.390 (0.002-1041.066). Therefore there was no significant interaction between religion and geography in relation to risky sexual behaviors among international immigrants in this study (Table 7).

**Table 6.** Logistic regression results of the impact of religion and geography on risky sexual behaviors. The italicized *P* values are considered significant.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Having multiple sexual partners</th>
<th>Having unprotected sexual behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AOR(^a) (95% CI)</td>
<td><em>P</em> value</td>
</tr>
<tr>
<td><strong>Religion</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Christianity</td>
<td>0.870 (0.510-1.510)</td>
<td>.62</td>
</tr>
<tr>
<td>Islam</td>
<td>0.627 (0.346-1.139)</td>
<td>.13</td>
</tr>
<tr>
<td>Hinduism</td>
<td>0.638 (0.266-1.526)</td>
<td>.31</td>
</tr>
<tr>
<td>Buddhism</td>
<td>0.292 (0.086-0.990)</td>
<td>.048</td>
</tr>
<tr>
<td>Others</td>
<td>0.810 (0.182-3.604)</td>
<td>.78</td>
</tr>
<tr>
<td>No religion</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td><strong>Geography</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>1.394 (0.897-2.167)</td>
<td>.14</td>
</tr>
<tr>
<td>North America</td>
<td>0.751 (0.316-1.783)</td>
<td>.52</td>
</tr>
<tr>
<td>South America</td>
<td>3.684 (1.379-9.840)</td>
<td>.009</td>
</tr>
<tr>
<td>Europe</td>
<td>2.310 (1.022-5.221)</td>
<td>.04</td>
</tr>
<tr>
<td>North Africa</td>
<td>3.524 (1.104-11.248)</td>
<td>.03</td>
</tr>
<tr>
<td>Others</td>
<td>1.471 (0.214-10.110)</td>
<td>.70</td>
</tr>
</tbody>
</table>

\(^a\)AOR: adjusted odds ratio.

\(^b\)Not applicable.
No sexual misconduct means abstaining from deception and killing, no stealing, no sexual misconduct, and no lying [35].

These precepts state the following: no harming (no sexual behaviors and multiple sexual behaviors than nonreligious participants (P < 0.05). In a previous study, African-American women reported that religious affiliation was associated with unprotected sex, and increased religiosity predicted a decrease of condomless sex [31]. Another prior study demonstrated that sexual standards among Muslims were reportedly paradoxical, with two opposing attitudes toward sex resulting in the back-and-forth swing in modern society; however, these standards have become increasingly restrictive now [32]. In a study conducted on men who have sex with men in China, Muslims were reportedly less likely to engage in unprotected sex, which aligns with the findings of our study [33]. The majority of the Muslims had STIs/HIV testing experiences, which was consistent with a prior study among Muslim Americans [34]. However, Muslims also had a low willingness to undergo STIs/HIV testing in the future. This may be because Muslims felt reassured after completing initial testing and believed that they were at low risk of contracting STIs/HIV, so their future willingness to undergo testing was relatively reduced. Notably, Buddhists were less likely to have unprotected sexual behaviors and multiple sexual behaviors than nonreligious people. Compared with other religions, Buddhism has less strict rules about sexual behaviors, but Buddhists are also taught to observe four ethical precepts that may limit their sexual behaviors. These precepts state the following: no harming (no killing), no stealing, no sexual misconduct, and no lying [35]. No sexual misconduct means abstaining from deception and betrayal within intimate or committed relationships. This precept aligns with the virtues of fidelity and purity of heart [36].

In this study, South Americans, Europeans, and North Africans were more likely to have multiple sexual partners and engage in unprotected sexual behaviors compared to Asians. The rate of engagement in unprotected sexual behaviors was highest in immigrants from South America, and the rate of having multiple sexual partners was also highest in immigrants from South America. In a study published in 2020, the prevalence of risky sexual behaviors among undergraduate students in Brazil was high [37]. The number of female sex workers is also high in South America, which may influence sexual attitudes, increasing rates of engagement in commercial sex with multiple partners and having unprotected sex [38,39]. Polygamy still exists in some South American and North African countries [40], and it may have influenced the sexual behaviors of people who migrate from these countries to other places, including acceptance of multiple sexual partners. Notably, global HIV data indicate that African regions had the highest number of people living with HIV in 2022 [41]. In 2022, the incidence rate of HIV in Southeast Asia was 6/100,000, compared to 57/100,000 in the African region, 20/100,000 in the European region, and 18/100,000 in the Americas [41]. In our study, compared with other geographical regions, immigrants from Asian regions had the lowest incidence of having multiple sexual partners and engaging in unprotected sex. This may be strongly influenced by Confucian culture, which compels a large number of Asians to maintain traditional conservative attitudes toward sexual behavior. In many Asian cultures, particularly Muslim communities, the supervision of female and male circumcision also limits the spread of STIs [42].

As a marginal group, immigrants have to experience cross-cultural adaptation. China is a nonreligious country; there may be few religious activities for religious international immigrants. Besides the differences in religious customs, they also face isolation and loneliness in the new living environment.
by the disruption of the social networks; the relatively poor living space; or the unfamiliarity with the language, the customs, and even the weather, which may result in their high rates of unethical sexual behaviors [43,44]. Moreover, China has a special cultural environment, such as the “condom culture” in terms of safe sex, which in the context of Chinese culture, can be considered as distrust of other parties and is often associated with “nightclubs,” commercial sexual behaviors, and stigmatization [12]. A previous study demonstrated that religious social capital was associated with higher levels of social support among international immigrants, and social support could reduce immigration stress effectively [45]. Additionally, another study in Nigeria [46] indicated that religion was beneficial to the quality of life of people living with HIV. Specifically, intrinsic religiosity, extrinsic religiosity, and positive religious coping had a positive correlation with psychological and physical health as well as the quality of life among people living with HIV. In such situations, religion is beneficial to health-promoting behaviors, as some religious communities are inclined to preach health education, which includes information delivered through faith ministers, religious leaders, and clergy. Furthermore, social support and encouragement from congregants are likely to foster acceptance of undergoing HIV testing [47]. Therefore, it may be a good way to organize religious activities in immigrant communities appropriately.

The reported rates of STIs/HIV testing experience and willingness to take these tests in the future were not high. This may be because many international immigrants are fearful of positive results or are reluctant to disclose private issues (eg, diagnosis of an STI) in a foreign country [48]. However, unawareness of HIV infection may result in severe consequences, contributing to the HIV/AIDS epidemic [49]. Additionally, international immigrants have proven to be risk groups for HIV [21]. Therefore, as an innovation, HIV self-testing should be promoted to increase the uptake and frequency of testing among international immigrants who are missed by existing services [50]. In addition, pre-exposure prophylaxis (PrEP) is widely recognized to reduce the risk of HIV acquisition among HIV-negative people; however, the awareness of PrEP among international immigrants is low [51]. A previous study indicated that language, cost, and medical system are the main barriers to acquiring HIV prevention, such as PrEP, when enhancing the reach of these services to international immigrants [52]. China is still developing the necessary legal regulations, implementation guidance, and standard operating procedures, which are needed for improving the coverage of PrEP. For instance, some researchers indicated that community-driven social media and web-based strategies are effective ways to promote PrEP use among Latino immigrant men [53]. Furthermore, sexual health education is still an effective and feasible way to enhance people’s awareness about the severe consequences of risky sexual behaviors. For example, health workshops could be held regularly in immigrant communities; carrying out STI/HIV-related knowledge contests with prizes may also have noticeable effects.

Limitations
First, the questionnaire was distributed via the internet, which entails a number of extensively previously described limitations, such as the accuracy of the results. However, it should be noted that web-based surveys also have advantages in numerous contexts, for example, avoidance of potential embarrassment during face-to-face surveys containing sex-related items. Second, it is possible that some international immigrants were compelled by social and cultural norms to hide their sexual behaviors. Third, given the very large numbers of international immigrants now relocating to China, the generalizability of the study results to the entire population may be limited. Fourth, the questionnaire was designed based on previous work and was amended according to our research participants and objectives. Thus, the validity of the questionnaire in the survey needs to be further tested.

Conclusions
Sex-related behaviors differ among international immigrants with different geographical regions of origin and religions. South Americans and nonreligious people were more likely to engage in risky sexual behaviors, such as having multiple partners and unprotected sexual behaviors. Given that international immigrants face a myriad of sexual health challenges, it is necessary to promote measures, including HIV self-testing, PrEP implementation, and targeted sexual health education to promote the physical and mental health of international immigrants in China.

Acknowledgments
We are grateful to all study participants for their participation. This work was supported in part by the National Natural Science Foundation of China (project number 71904171) and the Fundamental Research Funds for the Central Universities.

Data Availability
All of the main data have been included in the results. Additional materials with details may be obtained from the corresponding author upon reasonable request.

Authors’ Contributions
All authors were responsible for the structure of the manuscript. YZ conducted the data analysis and drafted the manuscript. JX contributed to the study’s conception and design, interpretation of the data, and critical revisions of the manuscript. FC contributed...
to the interpretation of the data and provided critical feedback on the manuscript. All authors approved the final version for submission.

Conflicts of Interest
None declared.

References


49. HIV/AIDS. World Health Organization. 2022. URL: https://www.who.int/health-topics/hiv-aids#tab=tab_1 [accessed 2024-06-14]


Abbreviations

AOR: adjusted odds ratio
PrEP: pre-exposure prophylaxis
STI: sexually transmitted infection

© Yuyin Zhou, Feng Cheng, Junfang Xu. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 28.6.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
CD4/CD8 Ratio Recovered as a Predictor of Decreased Liver Damage in Adults Infected With HIV: 16-Year Observational Cohort Study

Bingyu Liang1,2*, PhD; Rujing Sun1*, MD; Yanyan Liao1,2*, PhD; Aidan Nong3, BD; Jinfeng He1, BD; Fengxiang Qin1, PhD; Yanyun Ou3, BD; Jianhua Che1, BD; Zhenxian Wu3, BD; Yuan Yang1,2, BD; Jiao Qin3, BD; Jie Cai3, BD; Lijuan Bao3, BD; Li Ye1,2, PhD; Hao Liang1,2, PhD

1Guangxi Key Laboratory of AIDS Prevention and Treatment, School of Public Health, Guangxi Medical University, Nanning, China
2Biosafety III Laboratory, Life Science Institute, Guangxi Medical University, Nanning, China
3HIV/AIDS prevention department, Chongzuo Center for Disease Control and Prevention, Chongzuo, China

*these authors contributed equally

Corresponding Author:
Hao Liang, PhD
Biosafety III Laboratory, Life Science Institute
Guangxi Medical University
22 Shuangyong Road
Nanning, 530021
China
Phone: 86 13737089380
Fax: 86 07715323713
Email: lianghao@gxmu.edu.cn

Abstract

**Background:** As the life expectancy of individuals infected with HIV continues to increase, vigilant monitoring of non–AIDS-related events becomes imperative, particularly those pertaining to liver diseases. In comparison to the general population, patients infected with HIV experience a higher frequency of liver-related deaths. The CD4/CD8 ratio is emerging as a potential biomarker for non–AIDS-related events. However, few existing studies have been specially designed to explore the relationship between the CD4/CD8 ratio and specific types of non–AIDS-related events, notably liver damage.

**Objective:** This study aimed to investigate the potential association between the CD4/CD8 ratio and the development of liver damage in a sizable cohort of patients infected with HIV receiving antiretroviral treatment (ART). Additionally, the study sought to assess the effectiveness of 3 antiretroviral drugs in recovering the CD4/CD8 ratio and reducing the occurrence of liver damage in this population.

**Methods:** We conducted an observational cohort study among adults infected with HIV receiving ART from 2004 to 2020 in Guangxi, China. Propensity score matching, multivariable Cox proportional hazard, and Fine-Gray competing risk regression models were used to determine the relationship between the CD4/CD8 ratio recovered and liver damage.

**Results:** The incidence of liver damage was 20.12% among 2440 eligible individuals during a median follow-up period of 4 person-years. Patients whose CD4/CD8 ratio did not recover to 1.0 exhibited a higher incidence of liver damage compared to patients with a CD4/CD8 ratio recovered (adjusted hazard ratio 7.90, 95% CI 4.39-14.21; P<.001; subdistribution hazard ratio 6.80, 95% CI 3.83-12.11; P<.001), findings consistent with the propensity score matching analysis (adjusted hazard ratio 6.94, 95% CI 3.41-14.12; P<.001; subdistribution hazard ratio 5.67, 95% CI 2.74-11.73; P<.001). The Efavirenz-based regimen exhibited the shortest time for CD4/CD8 ratio recovery (median 71, IQR 49-88 months) and demonstrated a lower prevalence of liver damage (4.18/100 person-years).

**Conclusions:** Recovery of the CD4/CD8 ratio was associated with a decreased risk of liver damage in patients infected with HIV receiving ART, adding evidence for considering the CD4/CD8 ratio as a potential marker for identifying individuals at risk of non–AIDS-related diseases. An efavirenz-based regimen emerged as a recommended choice for recovering the CD4/CD8 ratio and mitigating the risk of liver damage.
Additionally, we assessed a potential link between 3 distinct cohort of patients infected with HIV, receiving ART. The aim of this study was to determine whether the CD4/CD8 ratio remains inadequately investigated [13].

Drugs class might lead to differing impacts on the CD4/CD8 ratio. Importantly, the CD4/CD8 ratio is routinely recorded and readily available in clinical practice. Therefore, it could be easily used as a predictor of non-AIDS-related illnesses, including liver damage and others, provided that more data are acquired to substantiate its applicability.

Owing to an insufficient frequency of occurrences in each category, this study was not originally designed to explore the association between the CD4/CD8 ratio and specific non-AIDS-defining events, notably liver damage [9,11,12]. Furthermore, previous studies have evaluated the effectiveness of various ART regimens within drug classes on the recovery and elevation of the CD4/CD8 ratio. However, findings concerning the potential benefit of the CD4/CD8 ratio exhibit inconsistency. Variations between the ART drugs within each class might lead to differing impacts on the CD4/CD8 ratio, and the relationship between liver damage and these effects remains inadequately investigated [13].

The aim of this study was to determine whether the CD4/CD8 ratio, a readily obtainable metric in most clinical settings and frequently evaluated in patients infected with HIV exhibited an association with the occurrence of liver damage within a large cohort of patients infected with HIV, receiving ART. Additionally, we assessed a potential link between 3 distinct antiviral medications and the prevalence of liver damage, along with their impact on the CD4/CD8 ratio.

### Methods

#### Participants and Study Setting

This was a retrospective cohort study. We collected the baseline and follow-up data from the National Free Antiretroviral Treatment Program database, a surveillance system that consistently records data on people living with HIV receiving free ART along with long-term follow-up care in China. We included all treatment-naïve people living with HIV aged 18 years or older who initiated therapy between November 15, 2004, and December 31, 2020, and were residents of Chongzuo City, Guangxi Province, China. Within the National Free Antiretroviral Treatment Program, routine monitoring of immunological parameters, including CD4+ and CD8+ cell counts, as well as other pertinent laboratory parameters and clinical treatment data such as bacterial-fungal infections and liver and kidney function, was performed.

#### Statistical Analysis

Quantitative data were presented as the median (IQR), whereas categorical variables were given as the frequencies. To mitigate any potential bias stemming from treatment selection, the CD4/CD8 ratio recovered and unrecovered groups underwent a 1:2 propensity score matching (PSM) process. Similarly, the lopinavir-based, efavirenz-based, and NPV-based groups were subjected to a 1:1:1 PSM procedure. Following the methods reported previously [14], PSM analysis was used to match specific sociodemographic characteristics (such as sex, marital status, HIV transmission route, age at diagnosis, age at ART initiation, and BMI) between the groups with a recovered and unrecovered CD4/CD8 ratio. A caliper of 0.05 was used to ensure the alignment of characteristic factors between the 2 groups. Subsequently, the standardized mean differences (SMDs) and chi-square test were then used to assess the efficacy of the PSM, with SMDs exceeding 10% indicating substantial differences. Similarly, a caliper of 0.00001 was established for the ART regimen groups.

Subsequently, we used Cox proportional hazard regression models to estimate the relationship between CD4/CD8 ratio recovery and the occurrence of liver damage. To determine whether CD4/CD8 ratio recovery is linked to a reduced risk of liver damage when competing risk of mortality is accounted for, we also used Fine-Gray competing risk regression models [15,16]. Using these same methodologies, we assessed the association between the ART regimen and the incidence of liver damage with the intention of enhancing clinical decision-making. A set of factors were selected as covariates: CD4/CD8 ratio recovery, ART regimens, sex, marital status, age at HIV diagnosis, age at ART initiation, HIV transmission route, WHO (World Health Organization) HIV disease stage at
baseline, BMI at baseline, CD4⁺ lymphocyte count at baseline, CD8⁺ lymphocyte count at baseline, CD4/CD8 ratio at baseline, cytomegalovirus (CMV) infection, aspartate aminotransferase (AST) at baseline, alanine aminotransferase (ALT) at baseline, and total bilirubin (TBIL) at baseline.

To visually represent the hazard ratio before and after PSM, a forest plot was generated. Additionally, as a sensitivity analysis to assess the potential for residual confounding, we calculated end point E-values [17]. Statistical significance was assessed at α=.05, and all analyses were 2-tailed. The statistical analyses were performed with R (version 4.1.0; R foundation for Statistical Computing) and SPSS Statistics (version 26.0; IBM Corp). The “MatchIt” R package (Noah Greifer) was used for PSM, while the “survival,” “cmprsk,” and “splines” R packages for Fine-Gray competing risk regression.

Ethical Considerations

This study was approved by the Human Research Ethics Committee of Guangxi Medical University (2019-SB-102), and it was carried out in accordance with the Helsinki Declaration. In this study site, when the patients first attended the HIV prevention department of Center of Disease Control and the HIV care clinic of the hospital, they were informed that their anonymized data could be used for research. As the data collection and extraction were anonymized, individual informed consent was not needed for this study. An informed consent waiver was approved by the institutional review board of Guangxi Medical University. Participants were compensated 50 RMB (US $7.14) or a free blood routine test for the first visit and each follow-up.

We identified liver damage as the primary end point during the follow-up period. This was delineated by the combined assessment of 3 liver function parameters: AST, ALT, and TBIL. We established reference values of AST=40 unit liter, ALT=40 unit liter, and TBIL=20 mol/L as the upper limits of normality (ULN) based on the division of AIDS toxicity guidelines. All values surpassing these thresholds were classified as abnormal and categorized into different grades of liver enzyme elevation (LEE) or total bilirubin elevation (TBE) as defined by the guidelines. Specifically, grade 1-IV LEE was defined as elevations ranging from 1-2.5, 2.5-5, 5-10, to more than 10 times the ULN, while grade 1-IV TBE was determined by elevations of 1-1.5, 1.5-2.5, 2.5-5, or more than 5 times the ULN. Participants with no LEE or TBE or only grade I LEE or TBE were categorized into the normal liver function group, while those with grade II, III, or IV LEE or TBE were classified into the liver damage group.

In addition to 2 nucleoside reverse transcriptase inhibitors, cART (combination antiretroviral treatment) was defined as a combination regimen that included a nonnucleoside reverse transcriptase inhibitor, protease inhibitor, or integrase strand transfer inhibitor. The ART regimens were categorized based on the primary medication used within the initial cART regimen. Specifically, these categories encompassed lopinavir from the protease inhibitor class and nevirapine and efavirenz from the nonnucleoside reverse transcriptase inhibitor class.

The baseline encompassed various clinical indices before the initiation of ART. Follow-up persisted until the earlier occurrence of the following events: liver damage, death, or the final clinical visit recorded at the end of the study.

The inclusion criteria for patients in this study were as follows: (1) having normal liver function parameters at baseline (including normal serum ALT, AST, and TBIL), (2) having at least 2 records of CD4/CD8 ratio, and (3) receiving combination ART regimens. Patients who met any of the following criteria were excluded from the study: (1) absence of baseline CD4⁺ or CD8⁺ lymphocyte count data, (2) absence of records for any of the 3 liver function markers (AST, ALT, or TBIL), (3) lack of normal liver function at baseline, (4) presence of only 1 CD4/CD8 ratio record, and (5) were treated with regimens other than cART.

The recovery of the CD4/CD8 ratio was defined as presence of 2 consecutive values ≥1.0 (with an interval between 2 consecutive records ranging from 180 days to less than 2 years) during the follow-up period [18]. Individuals meeting this criterion were defined as the recovered group, whereas those not meeting the criterion were grouped as the unrecovered group.

Results

We identified a total of 2440 patients with HIV who met the inclusion and exclusion criteria. The patient flow for inclusion in the analysis is presented in Figure 1. Among these eligible patients, the cumulative person-years of follow-up amounted to 9963, with a median of 4 (IQR 1-7) person-years. Within the cohort, 1516/2440 (62.13%) individuals were male, while 1616/2440 (66.23%) were married or cohabiting. The majority of patients were diagnosed with HIV (1043/2440, 42.75%) or initiated ART (1074/2440, 44.02%) at the age of 50 years or more. Heterosexual transmission accounted for 93.40% (2279/2440) of HIV acquisition, and 33.40% (815/2440) had progressed to WHO HIV clinical stage I disease at baseline. Regarding baseline characteristics, 53.12% (1296/2440) of patients had a normal BMI ranging from 18.5 to 23.9 kg/m². Furthermore, 51.19% (1249/2440) of patients exhibited a CD4⁺ lymphocyte count below 200 cells/μL, while 50.08% (1222/2440) of them had a CD8⁺ lymphocyte count ≤800 cells/μL. Notably, 68.89% (1681/2440) of patients displayed a CD4⁺/CD8⁺ ratio below 0.30 at baseline. Additionally, 0.70% (17/2440) of patients showed anti-CMV Immunoglobulin G antibodies (overall 17/2440, 0.70%; when only patients with an available CMV serology were considered: 17/2440, 0.70%; Table 1).

As illustrated in Table S1 in Multimedia Appendix 1, 9.96% (243/2440) of patients achieved CD4/CD8 ratio recovery when the cutoff point for CD4/CD8 ratio recovery was set at 1. To determine the most informative cutoff point, 3 distinct thresholds were tested. Fine-Gray competing risk regression analysis indicated that the highest subdistribution hazard ratio (sHR) was observed (sHR 6.80, 95% CI 3.83-12.11) when the cutoff point was set at 1.0. This value notably exceeded the sHR for a cutoff of 0.5 (sHR 5.93, 95% CI 4.63-7.58) and a cutoff of
1.2 (sHR 4.90, 95% CI 2.51-12.11). Consequently, a CD4/CD8 ratio recovery cutoff threshold of 1.0 was selected for subsequent analysis.

Tables S2 and S3 in Multimedia Appendix 1 showed the results of PSM, revealing that 480 patients with recovered CD4/CD8 ratios and 242 patients with unrecovered CD4/CD8 ratios were included. The matched participants from the 3 ART regimen groups were merged, resulting in a total of 501 patients (167 in the lopinavir-based group, 167 in the efavirenz-based group, and 167 in the nevirapine-based group). After matching, these demographic characteristics no longer exhibited statistically significant disparities across groups (SMDs<10% and P>.05).

Figure 2 shows the cumulative incidence curve, illustrating the unadjusted incidence of liver damage in both the recovered and unrecovered CD4/CD8 ratio groups. The Fine-Gray test revealed a substantially higher cumulative incidence of liver damage in individuals with a non-recovered CD4/CD8 ratio compared to those with a recovered ratio over the 16-year follow-up period (P<.001). Upon conducting multivariable analysis, both Cox proportional hazard regressions and Fine-Gray competing risk regressions (Figure 3) indicated that a nonrecovered CD4/CD8 ratio was associated with a heightened risk of liver damage both before and after matching (aHR 7.899, 95% CI 4.391-14.209; P<.001; sHR 6.803, 95% CI 3.831-12.107; P<.001) and after PSM (aHR 6.939, 95% CI 3.409-14.123; P<.001; sHR 6.123, 95% CI 2.962-12.659; P<.001). The aHR and sHR values for the other covariates are presented in Table S4 in Multimedia Appendix 1. Additionally, the E-value analysis revealed that an unmeasured confounding factor would need to exhibit a minimum association (aHR) of 6.17 (95% CI 2.55-9.79) with liver damage, considering the measured covariates, to challenge the observed outcome.

The CD4/CD8 ratio recovery rate exhibited no significant differences among the 3 ART treatment groups (χ²= 0.04; P=.98; v=2). Especially, the CD4/CD8 ratio recovery rate was 10.1% for the nevirapine-based group, 10% for the efavirenz-based group, and 9.9% for the lopinavir-based group (Table S5 in Multimedia Appendix 1). In addition, there was a notable disparity in the time required for CD4/CD8 ratio recovery (defined as the interval between cART initiation and the attainment of CD4/CD8 ratio recovery, measured in months) across the 3 ART regimen groups (P=.01). The nevirapine-based group had the longest duration for recovery (113 months, IQR 95-131 months), substantially surpassing the efavirenz-based (71 months, IQR 49-88 months) and lopinavir-based (77 months, IQR 59-91 months) groups (Figure 4).

Through the use of stratified analysis based on different ART regimens, the relationship between CD4/CD8 ratio recovery and liver damage was assessed across 3 distinct subgroups characterized by competing mortality risks. In the nevirapine-based regimen, it was evident that the cumulative incidence of liver damage was significantly higher in the CD4/CD8 ratio unrecovered group compared to the recovered group, both before (Figure 5C) and after PSM (Figure 5F). However, within the group administered lopinavir and efavirenz regimens, no notable discrepancy in liver damage incidence was observed after PSM (Figures 5A, 5D, and 5F). Notably, patients subjected to the nevirapine-based regimen and exhibiting an unrecovered CD4/CD8 ratio displayed the highest liver damage incidence rate (7/100 person-years), a notably elevated figure in comparison to CD4/CD8 ratio recovered patients (Table S6 in Multimedia Appendix 1).

As shown in Figure 6, after being modified for a specified group of factors through forward-selection, both the Cox proportional hazard time-to-event regressions and Fine-Gray competing risk regressions (Figure 6) indicated that the efavirenz-based regimen was associated with a notably reduced incidence of liver damage than the nevirapine-based regimen. This distinction held true both before (aHR 0.675, 95% CI 0.544-0.836; P<.001; sHR 0.658, 95% CI 0.532-0.814; P<.001) and after PSM (aHR 0.564, 95% CI 0.322-0.987; P=.045; sHR 0.573, 95% CI 0.332-0.987; P=.045).

Figure 1. Cohort enrollment criteria. ALT: alanine aminotransferase; ART: antiretroviral treatment; AST: aspartate aminotransferase; HBV: hepatitis B virus; HCV: hepatitis C virus; TBIL: total bilirubin.
Table 1. Sociodemographic characteristics of patients with HIV receiving antiretroviral treatment (ART; N=2440).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Study population, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1516 (62.13)</td>
</tr>
<tr>
<td>Female</td>
<td>924 (37.87)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Married or cohabitation</td>
<td>1616 (66.23)</td>
</tr>
<tr>
<td>Divorced, separated, or widowed</td>
<td>493 (20.21)</td>
</tr>
<tr>
<td><strong>HIV transmission route</strong></td>
<td></td>
</tr>
<tr>
<td>Heterosexual contact</td>
<td>2279 (93.40)</td>
</tr>
<tr>
<td>Homosexual contact</td>
<td>37 (1.52)</td>
</tr>
<tr>
<td>Blood or plasma transfusion</td>
<td>124 (5.08)</td>
</tr>
<tr>
<td><strong>Age at diagnosis (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>442 (18.12)</td>
</tr>
<tr>
<td>30-49</td>
<td>955 (39.14)</td>
</tr>
<tr>
<td>≥50</td>
<td>1043 (42.75)</td>
</tr>
<tr>
<td><strong>Age at ART initiation (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>386 (15.82)</td>
</tr>
<tr>
<td>30-49</td>
<td>980 (40.16)</td>
</tr>
<tr>
<td>≥50</td>
<td>1074 (44.02)</td>
</tr>
<tr>
<td><strong>BMI at baseline, kg/m²</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>585 (23.98)</td>
</tr>
<tr>
<td>18.5-23.9</td>
<td>1296 (53.12)</td>
</tr>
<tr>
<td>24-27.9</td>
<td>169 (6.93)</td>
</tr>
<tr>
<td>≥28</td>
<td>27 (1.11)</td>
</tr>
<tr>
<td>Unknown</td>
<td>363 (14.88)</td>
</tr>
<tr>
<td><strong>WHO® HIV disease stage at baseline</strong></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>815 (33.40)</td>
</tr>
<tr>
<td>II</td>
<td>592 (24.26)</td>
</tr>
<tr>
<td>III</td>
<td>434 (17.79)</td>
</tr>
<tr>
<td>IV</td>
<td>492 (20.16)</td>
</tr>
<tr>
<td>Unknown</td>
<td>107 (4.39)</td>
</tr>
<tr>
<td><strong>CD4+ lymphocyte count at baseline, cells/μL</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;200</td>
<td>1249 (51.19)</td>
</tr>
<tr>
<td>200-349</td>
<td>813 (33.32)</td>
</tr>
<tr>
<td>≥350</td>
<td>378 (15.49)</td>
</tr>
<tr>
<td><strong>CD8+ lymphocyte count at baseline, cells/μL</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;800</td>
<td>1222 (50.08)</td>
</tr>
<tr>
<td>800-1499</td>
<td>913 (37.42)</td>
</tr>
<tr>
<td>≥1500</td>
<td>305 (12.50)</td>
</tr>
<tr>
<td><strong>CD4/CD8 ratio at baseline</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;0.3</td>
<td>1681 (68.89)</td>
</tr>
<tr>
<td>0.3-0.59</td>
<td>670 (27.46)</td>
</tr>
</tbody>
</table>
### Table

<table>
<thead>
<tr>
<th>Variable</th>
<th>Study population, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥0.6</td>
<td>89 (3.65)</td>
</tr>
<tr>
<td><strong>Cytomegalovirus infection</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2258 (92.54)</td>
</tr>
<tr>
<td>Yes</td>
<td>17 (0.70)</td>
</tr>
<tr>
<td>Unknown</td>
<td>165 (6.76)</td>
</tr>
</tbody>
</table>

*aWHO: World Health Organization.*

**Figure 2.** Unadjusted cumulative incidence of liver damage among adults infected with HIV receiving antiretroviral therapy, accounting for the competing risk of death. ART: antiretroviral treatment; LD: liver damage.

**Figure 3.** Forest plots of multivariable Cox proportional hazard regression and Fine-Gray competing risk regression analysis of the effect of CD4/CD8 ratio recovery on liver damage among patients with HIV receiving antiretroviral treatment (ART). Hazard ratio adjusted by CD4/CD8 ratio recovery, ART regimen, sex, marital status, age at HIV diagnosis, age at ART initiation, HIV transmission route, World Health Organization HIV disease stage, BMI at baseline, CD4+ lymphocyte count at baseline, CD8+ lymphocyte count at baseline, CD4/CD8 ratio at baseline, cytomegalovirus infection, aspartate aminotransferase at baseline, alanine aminotransferase at baseline, and total bilirubin at baseline. HR: hazard ratio; PSM: propensity score matching.

<table>
<thead>
<tr>
<th>Liver damage incidence rate (per 100 person-years)</th>
<th>Unadjusted CD4/CD8 ratio</th>
<th>P value</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrecovered</td>
<td>5.812</td>
<td>&lt;.001</td>
<td>7.899 (4.391, 14.209)</td>
</tr>
<tr>
<td>Recovered</td>
<td>0.697</td>
<td>–</td>
<td>1 (Reference)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PSM adjusted CD4/CD8 ratio</th>
<th>Unrecovered</th>
<th>P value</th>
<th>HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrecovered</td>
<td>4.503</td>
<td>&lt;.001</td>
<td>6.939 (3.409, 14.123)</td>
</tr>
<tr>
<td>Recovered</td>
<td>0.700</td>
<td>–</td>
<td>1 (Reference)</td>
</tr>
</tbody>
</table>

0.5  1  4  8  12  16 
Cox proportional hazard regression  Fine-Gray competing risk regression
Figure 4. Time for CD4/CD8 ratio recovery among different antiretroviral treatment regimen groups. NS: not statistically significant. ****P<.0001.

Figure 5. Cumulative incidence of liver damage (LD) for patients with HIV receiving antiretroviral treatment (ART), grouped by ART regimen. (A) The whole research population of patients who were treated with a lopinavir-based regimen. (B) The whole research population of patients who were treated with an efavirenz-based regimen. (C) The whole research population of patients who were treated with a nevirapine-based regimen. (D) The propensity score matching (PSM) of patients treated with a lopinavir-based regimen. (E) PSM patients treated with an efavirenz-based regimen. (F) PSM patients treated with a nevirapine-based regimen. The statistical significance was measured by the Fine-Gray test. LD: liver damage.
Discussion

Principal Results

This study represents the first comprehensive large cohort study conducted in a real-world setting with long-term follow-up to examine the association between CD4/CD8 ratio recovery and the risk of liver damage among patients infected with HIV receiving ART. Our findings underscored the notable incidence of liver damage and its strong association with the unrecovered CD4/CD8 ratio. Furthermore, we observed that participants treated with an efavirenz-based regimen exhibited a diminished incidence of liver damage and an unrecovered CD4/CD8 ratio in comparison to those administered with the nevirapine-based regimen. Given that this study centered on individuals with HIV with initially normal liver function, with the end point defined as the occurrence of liver damage, these results bear significant implications for comprehending the interrelation between the CD4/CD8 ratio and non–AIDS-defining events in patients with HIV. Thus, this study contributes substantially to the body of research in this field.

Comparison With Previous Work

In this study, liver damage was observed in 20% (491/2440) of the participants, a prevalence in line with earlier studies that reported figures ranging from 12% to 27% [19,20]. However, it is noteworthy that the highest incidence of liver damage reported in this study was 48%, occurring within the initial 3 months following the commencement of cART. This rate is both higher and faster than what was observed in a previous study (35% at 9 months) [19]. Nearly half of the patients in this cohort were older adults, aligning with the prevalent demographic trend among patients in Guangxi province [21,22]. Indeed, the elevated incidence of liver damage can be attributed to the advanced age of the patient cohort. The higher incidence underscores the substantial burden of liver disease within these individuals and the early onset of liver function impairment upon initiating cART, accentuating the significance of the initial cART regimen.

CD4/CD8 ratio recovery exhibited a robust correlation with a notable reduction in the incidence of liver damage, a relationship confirmed by both Cox regression and PSM analysis, methodologies designed to mitigate bias stemming from potential confounding variables. Notably, the observed risk reduction in liver damage remained consistent across various ratio thresholds, even when accounting for the presence of competing mortality risks. Despite variations in analytic approaches and threshold values, these findings align with results reported by Mussini et al [9], which demonstrated that a diminished CD4/CD8 ratio following ART initial independently contributed to an escalated risk of severe non–AIDS-defining events.

Furthermore, a study conducted on individuals coinfected with the hepatitis C virus and HIV demonstrated a close correlation between liver damage, assessed through transaminase levels, and the CD4/CD8 ratio [23]. These findings align with other research that has established an association between a low CD4/CD8 ratio and markers of liver health, such as liver stiffness and fibrosis [24]. Additionally, this ratio has been linked to unfavorable outcomes in smaller, often single-center cohorts [7,25,26]. However, contrasting results have been reported. For instance, a Canadian cohort study, after adjusting for variables including age and HIV RNA levels, found the CD4/CD8 ratio did not offer additional short-term prognostic value for clinical outcomes [27]. A diminished or inverted CD4/CD8 ratio reflects sustained immunological activation and is associated with immunodiscordance and the persistent inflammation that is characteristic of HIV infection [25,28,29]. Besides HIV-induced hepatitis, drug-induced hepatotoxicity constitutes a significant cause of liver damage. A post hoc analysis of previous studies revealed that an efavirenz-based ART regimen exhibited a more substantial improvement in the CD4/CD8 ratio compared to other frequently used ART
regimens [30,31]. In this study, we further explored the relationship between the extent and duration of CD4/CD8 ratio recovery and the incidence of liver damage in the 3 main ART regimens. Our findings indicated that there were no significantly differences in the rates of CD4/CD8 ratio recovery at a cutoff of 1.0 across the three ART regimen groups. However, it is worth highlighting that the efavirenz-based regimen was associated with a significant shorter period until CD4/CD8 recovery compared to the nevirapine-based regimen. Notably, users of efavirenz and lopinavir exhibited a greater enhancement in the CD4/CD8 ratio than nevirapine users, a result consistent with findings from a different research cohort [32].

In addition, this study demonstrated a consistent correlation between CD4/CD8 ratio recovery and reduced liver damage in nevirapine-based regimen groups, as validated by PSM analysis. This conclusion is further supported by the observation that patients receiving nevirapine-based therapy exhibited the longest duration to achieve CD4/CD8 ratio recovery in this study. The outcomes of this study hold clinical significance, particularly within resource-limited settings, and underscore the need for careful consideration of CD4/CD8 ratio recovery in this population to mitigate the risk of liver damage. This recommendation holds particular weight given the lifelong duration of ART regimens, where drug toxicity assumes paramount importance.

Limitations
This research has several limitations that warrant consideration. First, as a retrospective clinical study, the presence of unmeasured confounding variables is a potential concern. Second, the assessment of antiretroviral drug adherence, a factor that could affect the CD4/CD8 ratio recovery, was not feasible within the scope of this study. Third, this study lacked control over the use of medications other than ART regimens, some of which could be associated with hepatotoxicity. Last, in absence of viral suppression data in our cohort precludes us from definitively ascertaining occurrences of virological failure. Despite these limitations, the strength of our research lies in its substantial patient enrollment and robust multivariate analysis. Notably, the application of E-value analysis underscores that a potent confounding variable would be required to reverse the observed findings. An additional advantage is the use of competing-risk analysis, which provides unbiased estimations of cumulative events by accounting for death as a competing event. The stringent baseline liver function inclusion criteria further contribute to the study’s strengths; however, it should be acknowledged that selection bias could be present, given that a portion of patients were excluded due to these rigorous criteria.

Conclusions
This study had clinical and public health significance and offered additional evidence that the recovery of the CD4/CD8 ratio was associated with a lower risk of liver damage compared to ratio inversion. Notably, patients with HIV treated with efavirenz-based and lopinavir-based regimens exhibited a more pronounced increase in the CD4/CD8 ratio and a lower incidence of liver damage in contrast to those receiving nevirapine-based regimens. These findings highlight the importance of ongoing monitoring of the CD4/CD8 ratio and prompt intervention in the event of ratio inversion, particularly for individuals under nevirapine-based ART regimens.

Acknowledgments
We thank all the researchers involved in this study and the staff of the Chongzuo Center for Disease Control and Prevention. This work was supported by the the National Key R&D Program of China (2022YFC2305001), Guangxi Scientific and Technological Key Project (Guike 2022AC23005 and Guike 2022JJA141110), Thousands of Young and Middle-aged Key Teachers Training Program in Guangxi Colleges and Universities (To Bingyu Liang), and the Guangxi Medical University Training Program for Distinguished Young Scholars (To Junjun Jiang).

Authors' Contributions
BL, HL, and LY conceived and designed the study. BL, RS, and YL analyzed the data and wrote the manuscript. AN, JH, FQ, YO, JC, ZW, JC, YY, and JQ were responsible for data collection. HL, LY, and LB further edited the manuscript and gave final approval. HL, LY, and LB contributed equally to the manuscript. All authors have critically reviewed the paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary tables.

References


16. Van Der Pas S, Nelissen R, Fiocco M. Different competing risks models for different questions may give similar results in arthroplasty registers in the presence of few events. Acta Orthop 2018;89(2):145-151 [MEDLINE: 29388452]


Abbreviations

- aHR: adjusted hazard ratio
- ALT: alanine aminotransferase
- ART: antiretroviral treatment
- AST: aspartate aminotransferase
- cART: combination antiretroviral treatment
- CMV: cytomegalovirus
- LEE: liver enzyme elevation
- PSM: propensity score matching
- sHR: subdistribution hazard ratio
- SMD: standardized mean difference
- TBE: total bilirubin elevation
- TBIL: total bilirubin
- ULN: upper limits of normality
- WHO: World Health Organization
Please cite as:
CD4/CD8 Ratio Recovered as a Predictor of Decreased Liver Damage in Adults Infected With HIV: 16-Year Observational Cohort Study
JMIR Public Health Surveill 2024;10:e45818
URL: https://publichealth.jmir.org/2024/1/e45818
doi:10.2196/45818
PMID:37846087

©Bingyu Liang, Rujing Sun, Yanyan Liao, Aidan Nong, Jinfeng He, Fengxiang Qin, Yanyun Ou, Jianhua Che, Zhenxian Wu, Yuan Yang, Jiao Qin, Jie Cai, Lijuan Bao, Li Ye, Hao Liang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 09.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
AIDSVu Cities’ Progress Toward HIV Care Continuum Goals: Cross-Sectional Study

Nicole Hood, MPH; Nanette Benbow, MAS; Chandni Jaggi, MPH; Shamaya Whitby, MS; Patrick Sean Sullivan, DVM, PhD; AIDSVu Health Department Collaborators

Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States
Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
See Acknowledgements, Atlanta, GA, United States

Corresponding Author:
Nicole Hood, MPH
Department of Epidemiology
Rollins School of Public Health
Emory University
1518 Clifton Rd
Atlanta, GA, 30329
United States
Phone: 1 4047273956
Email: nicole.hood@emory.edu

Abstract

Background: Public health surveillance data are critical to understanding the current state of the HIV and AIDS epidemics. Surveillance data provide significant insight into patterns within and progress toward achieving targets for each of the steps in the HIV care continuum. Such targets include those outlined in the National HIV/AIDS Strategy (NHAS) goals. If these data are disseminated, they can be used to prioritize certain steps in the continuum, geographic locations, and groups of people.

Objective: We sought to develop and report indicators of progress toward the NHAS goals for US cities and to characterize progress toward those goals with categorical metrics.

Methods: Health departments used standardized SAS code to calculate care continuum indicators from their HIV surveillance data to ensure comparability across jurisdictions. We report 2018 descriptive statistics for continuum steps (timely diagnosis, linkage to medical care, receipt of medical care, and HIV viral load suppression) for 36 US cities and their progress toward 2020 NHAS goals as of 2018. Indicators are reported categorically as met or surpassed the goal, within 25% of attaining the goal, or further than 25% from achieving the goal.

Results: Cities were closest to meeting NHAS goals for timely diagnosis compared to the goals for linkage to care, receipt of care, and viral load suppression, with all cities (n=36, 100%) within 25% of meeting the goal for timely diagnosis. Only 8% (n=3) of cities were >25% from achieving the goal for receipt of care, but 69% (n=25) of cities were >25% from achieving the goal for viral suppression.

Conclusions: Display of progress with graphical indicators enables communication of progress to stakeholders. AIDSVu analyses of HIV surveillance data facilitate cities’ ability to benchmark their progress against that of other cities with similar characteristics. By identifying peer cities (e.g., cities with analogous populations or similar NHAS goal concerns), the public display of indicators can promote dialogue between cities with comparable challenges and opportunities.

Keywords: HIV; epidemiology; surveillance; HIV care continuum; cities; HIV public health; HIV prevention; diagnosis; HIV late diagnosis

Introduction

Public health surveillance data are foundational in the understanding of the current state of the HIV and AIDS epidemics [1]. Surveillance data provide significant insight into patterns within and progress toward achieving targets for each of the steps in the HIV care continuum—here, defined as timely diagnosis of HIV infection, linkage to HIV medical care, receipt
of HIV medical care, and HIV viral load suppression—for geographic areas and populations [2]. Dissemination of these data can be used to categorize certain steps in the continuum, geographic locations, and groups of people.

AIDSVu was created in 2009 as part of a cooperative private-public-academic partnership between Emory University’s Rollins School of Public Health, Gilead Sciences, Inc, and the Center for AIDS Research at Emory University [3]. AIDSVu aims to provide policy makers, health departments, and others who have a stake in health with an instrument to educate the public, monitor progress and trends, and judiciously advocate for resources and plan programs. Here, we report, for the first time, descriptive statistics for 36 US cities’ status for each of the HIV care continuum steps and their progress toward achieving 2020 National HIV/AIDS Strategy (NHAS) goals based on AIDSVu data [4]. We describe a standardized method to compare HIV care continuum data at the city, regional, and national levels.

Methods

Data Collection

We obtained the data used in this analysis through a data request in the spring of 2020 sent to AIDSVu-participating health departments for continuum indicators for people with HIV aged 13 years and older from the enhanced HIV/AIDS reporting system. The request was sent to health departments who had submitted data to AIDSVu in the past or who expressed interest in participating. A total of 36 cities provided stratified data without identifiers related to timely HIV diagnosis, linkage to HIV care, receipt of HIV care, and HIV viral suppression. Data provided on the AIDSVu website are regularly updated and city participation varies by year; these analyses use only the data as they were available for 2018.

In this study, we defined cities as single cities, single or multiple counties that included the named city, or metropolitan statistical areas (MSAs) that included the named city. Use of common SAS code to calculate indicators enabled comparability of indicators across jurisdictions. Data stratifications for display (ie, by age, sex, and race or ethnicity) were chosen by the jurisdictions. For stratification purposes, age was defined as age at initial diagnosis for timely diagnosis and linkage to care; age for the other HIV continuum measures was age as of the end of 2018.

Although AIDSVu HIV care continuum indicator definitions were guided by a process that involved the Centers for Disease Control and Prevention (CDC), the counts and percentages for each of the 4 HIV care continuum indicators were defined slightly differently from those of the CDC (see the AIDSVu website for more information) [3,5]. Timely HIV diagnoses were assessed among persons aged 13 years and older with new HIV diagnoses within the 5-year period from 2014 through 2018. An HIV diagnosis was considered timely if the person did not receive an AIDS diagnosis within 3 months of their initial HIV diagnosis. Timely diagnosis represented the inverse of the traditional late diagnosis continuum measure. Linkage to care was defined as having a CD4 lymphocyte or HIV viral load test within 1 month of diagnosis among those at least 13 years of age with a new HIV diagnosis in the 5-year period from 2014 to 2018. The numerator for the receipt of care definition included individuals who were 13 years of age or older and had been diagnosed with HIV by the end of 2017 and were alive throughout 2018 with ≥1 CD4 lymphocyte or HIV viral load test within 2018. The denominator included those living with HIV in 2018 (excluding those newly diagnosed in 2018). Those diagnosed with HIV by the end of 2017 and living with HIV throughout 2018 with an HIV viral load <200 copies/mL for their most recent HIV viral load test within 2018 were determined to be virally suppressed.

Data Analysis

Descriptive analyses were conducted for data as of 2018 from 36 cities using SAS (version 9.4, SAS Institute). Cities were categorized by US Census region (ie, Midwest, 2 cities; Northeast, 9 cities; South, 21 cities; and West, 4 cities). The national HIV case counts and percentages for timely diagnosis, receipt of care, and viral suppression were obtained from the CDC’s HIV Surveillance Report for 2018 [6]. We compared indicators reported by the cities with NHAS-designated HIV care continuum benchmarks proposed to be achieved by 2020 [4]. There is currently an updated set of goals through 2025 [7]; however, because our data are from 2018, we compared the data to 2020 goals. Given that there is not an NHAS indicator for timely diagnosis, we used the goal for the indicator called “Knowledge of HIV+ Status,” which was 90% (note: “Knowledge of HIV+ Status” is not tied to a particular time period of receipt of diagnosis) [4]. The NHAS linkage to care goal was to have 85% of persons with a new HIV diagnosis linked to HIV medical care within 1 month of diagnosis. This goal was applied to our linkage to care indicator. The NHAS retention in care goal of 85% was adapted for our receipt of care indicator, and the NHAS viral suppression goal of 80% was used for our viral suppression indicator.

The cities’ progress toward each of the HIV care continuum goals was assessed by determining the relative percentage difference from each goal. For example, the NHAS goal for linkage to care was 85% and the percentage of people in Atlanta with newly diagnosed HIV linked to an HIV care provider was 69% (n=5321). To calculate Atlanta’s progress, we subtracted 0.69 from 0.85, divided the resulting number by 0.69, and then multiplied it by 100 to say that Atlanta was within 25% of achieving the NHAS goal for linkage to care. We categorized city- and target-specific progress toward the goal as follows: met or surpassed the goal, within 25% of attaining the goal, or further than 25% from achieving the goal. We applied the same method to each of the priority populations for the individual cities’ progress tables. We defined priority populations as sex, race or ethnicity, and age groups. Finally, we categorized age as the following: 13-24, 25-44, 45-59, and 60 years of age and older. We calculated the percentage of cities that were at least within 25% of meeting each indicator goal within priority populations.

Ethical Considerations

As this study used publicly available, secondary, deidentified data (numbers were aggregated to the city level or beyond),
ethical evaluation via institutional review board review was not warranted.

Results

All Cities’ HIV Care Continuum Indicators

Within our 36 cities, 447,371 people were living with HIV as of 2018, with 82,827 people newly diagnosed with HIV during the 5-year period from 2014 to 2018. In 2018 alone, 15,767 people were newly diagnosed with HIV. Table 1 displays the numerators and percentages for HIV care continuum indicators at the national (except for linkage to care), regional, and city levels. More than half (n=22, 61%) of AIDSVu cities had a higher percentage for timely diagnosis than the national percentage for timely diagnosis. For receipt of care, more than half of the cities (n=24) had higher percentages than the percentage for the national average (n=661,816, 75.7%). Compared to the national percentage for viral suppression (n=565,195, 64.7%), 10 cities had higher percentages. City-to-city comparisons were made with the understanding that cities were broadly defined at the city, single county, multicounty, or MSA level.
Table 1. HIV care continuum indicator numerators and percentages by AIDSVu city with categorization by progress toward 2020 National HIV/AIDS Strategy goals, 36 US cities, as of 2018.

<table>
<thead>
<tr>
<th>City, n (%)</th>
<th>Timely diagnosis(^a)</th>
<th>Linkage to care(^a)</th>
<th>Receipt of care(^b)</th>
<th>Viral suppression(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NHAS(^c) goals, %</strong></td>
<td>90</td>
<td>85</td>
<td>85</td>
<td>80</td>
</tr>
<tr>
<td>National(^e), n (%)</td>
<td>153,331 (78.6)</td>
<td>N/A(^c)</td>
<td>661,816 (75.7)</td>
<td>565,195 (64.7)</td>
</tr>
<tr>
<td><strong>Region of participating cities, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>3894 (81.1)</td>
<td>3715 (77.4)</td>
<td>14,659 (72.8)</td>
<td>11,412 (56.7)</td>
</tr>
<tr>
<td>Northeast(^f)</td>
<td>12,144 (79.7)</td>
<td>1478 (75.7)</td>
<td>11,5136 (69.8)</td>
<td>98,377 (59.6)</td>
</tr>
<tr>
<td>South</td>
<td>44,040 (80.1)</td>
<td>33,625 (69.0)</td>
<td>166,653 (76.2)</td>
<td>132,020 (59.0)</td>
</tr>
<tr>
<td>West(^g)</td>
<td>6156 (79.0)</td>
<td>4706 (74.0)</td>
<td>2714 (70.2)</td>
<td>23,611 (61.2)</td>
</tr>
<tr>
<td><strong>City, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlanta(^h)</td>
<td>6206 (80.5)</td>
<td>5321 (69.0)</td>
<td>24,981 (73.6)</td>
<td>20,576 (59.5)</td>
</tr>
<tr>
<td>Austin(^h)</td>
<td>1228 (82.0)</td>
<td>946 (63.2)</td>
<td>5036 (85.4)</td>
<td>4330 (70.9)</td>
</tr>
<tr>
<td>Baltimore(^i)</td>
<td>2040 (77.9)</td>
<td>1986 (75.9)</td>
<td>13,358 (80.5)</td>
<td>11,326 (63.1)</td>
</tr>
<tr>
<td>Baton Rouge</td>
<td>788 (80.7)</td>
<td>711 (72.8)</td>
<td>3091 (85.7)</td>
<td>2624 (69.7)</td>
</tr>
<tr>
<td>Birmingham(^l)</td>
<td>655 (78.7)</td>
<td>576 (69.2)</td>
<td>3299 (85.7)</td>
<td>2757 (66.0)</td>
</tr>
<tr>
<td>Bridgeport(^h)</td>
<td>269 (73.3)</td>
<td>282 (76.8)</td>
<td>1897 (79.2)</td>
<td>1593 (63.5)</td>
</tr>
<tr>
<td>Charleston</td>
<td>344 (77.3)</td>
<td>329 (73.9)</td>
<td>1318 (70.3)</td>
<td>1138 (57.0)</td>
</tr>
<tr>
<td>Charlotte(^m)</td>
<td>1146 (83.5)</td>
<td>842 (61.4)</td>
<td>4734 (78.0)</td>
<td>4078 (61.8)</td>
</tr>
<tr>
<td>Chicago</td>
<td>3395 (81.2)</td>
<td>3227 (77.1)</td>
<td>12,297 (75.5)</td>
<td>9339 (54.3)</td>
</tr>
<tr>
<td>Columbia(^h)</td>
<td>585 (75.8)</td>
<td>598 (77.5)</td>
<td>1655 (66.9)</td>
<td>1384 (51.5)</td>
</tr>
<tr>
<td>Dallas(^h)</td>
<td>3355 (79.6)</td>
<td>2866 (68.0)</td>
<td>13,675 (80.6)</td>
<td>10,703 (59.5)</td>
</tr>
<tr>
<td>Denver(^l)</td>
<td>1119 (78.2)</td>
<td>—(^n)</td>
<td>6154 (74.0)</td>
<td>5463 (55.4)</td>
</tr>
<tr>
<td>Fort Lauderdale(^m)</td>
<td>2785 (81.5)</td>
<td>2544 (74.5)</td>
<td>15,279 (82.6)</td>
<td>13,285 (68.5)</td>
</tr>
<tr>
<td>Fort Worth</td>
<td>1086 (77.7)</td>
<td>893 (63.9)</td>
<td>4295 (80.3)</td>
<td>3587 (62.4)</td>
</tr>
<tr>
<td>Hampton Roads(^h)</td>
<td>1176 (80.4)</td>
<td>932 (63.7)</td>
<td>4607 (74.1)</td>
<td>4213 (61.8)</td>
</tr>
<tr>
<td>Hartford(^h)</td>
<td>332 (73.8)</td>
<td>344 (76.4)</td>
<td>2385 (83.0)</td>
<td>2093 (69.6)</td>
</tr>
<tr>
<td>Houston</td>
<td>5038 (80.0)</td>
<td>4028 (64.0)</td>
<td>18,750 (77.2)</td>
<td>15,471 (59.9)</td>
</tr>
<tr>
<td>Jacksonville(^h)</td>
<td>1266 (78.8)</td>
<td>1020 (63.5)</td>
<td>5611 (82.7)</td>
<td>4412 (62.7)</td>
</tr>
<tr>
<td>Las Vegas(^m)</td>
<td>1662 (78.1)</td>
<td>1715 (80.6)</td>
<td>5965 (72.1)</td>
<td>5149 (57.2)</td>
</tr>
<tr>
<td>Miami(^m)</td>
<td>5037 (82.7)</td>
<td>4144 (68.0)</td>
<td>18,128 (78.3)</td>
<td>15,418 (60.1)</td>
</tr>
<tr>
<td>Milwaukee(^m)</td>
<td>499 (81.0)</td>
<td>488 (79.2)</td>
<td>2362 (85.9)</td>
<td>2073 (70.8)</td>
</tr>
<tr>
<td>Newark</td>
<td>697 (80.3)</td>
<td>598 (68.9)</td>
<td>2566 (56.5)</td>
<td>1926 (62.3)</td>
</tr>
<tr>
<td>New Haven(^m)</td>
<td>319 (76.0)</td>
<td>340 (81.0)</td>
<td>2634 (87.5)</td>
<td>1847 (62.3)</td>
</tr>
<tr>
<td>New Orleans(^l)</td>
<td>1256 (79.9)</td>
<td>1103 (70.2)</td>
<td>5130 (81.0)</td>
<td>4421 (65.4)</td>
</tr>
<tr>
<td>New York City</td>
<td>8391 (79.9)</td>
<td>7972 (75.9)</td>
<td>70,945 (75.8)</td>
<td>61,702 (61.3)</td>
</tr>
<tr>
<td>Orlando(^l)</td>
<td>2397 (77.4)</td>
<td>1936 (62.6)</td>
<td>9763 (81.2)</td>
<td>8541 (68.1)</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>2039 (81.8)</td>
<td>1970 (79.0)</td>
<td>11,711 (69.6)</td>
<td>9832 (54.8)</td>
</tr>
<tr>
<td>City</td>
<td>Timely diagnosis</td>
<td>Linkage to care</td>
<td>Receipt of care</td>
<td>Viral suppression</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Phoenix</td>
<td>2127 (80.9)</td>
<td>1608 (61.2)</td>
<td>8051 (76.8)</td>
<td>6552 (57.4)</td>
</tr>
<tr>
<td>Providence</td>
<td>97 (68.8)</td>
<td>—</td>
<td>347 (67.1)</td>
<td>310 (58.1)</td>
</tr>
<tr>
<td>Raleigh</td>
<td>564 (79.8)</td>
<td>461 (65.2)</td>
<td>2685 (75.1)</td>
<td>2285 (60.2)</td>
</tr>
<tr>
<td>Richmond</td>
<td>808 (81.3)</td>
<td>658 (66.2)</td>
<td>3496 (76.5)</td>
<td>3066 (62.3)</td>
</tr>
<tr>
<td>San Antonio</td>
<td>1549 (82.3)</td>
<td>1041 (55.3)</td>
<td>5078 (79.8)</td>
<td>4287 (63.9)</td>
</tr>
<tr>
<td>Seattle</td>
<td>1248 (77.9)</td>
<td>1383 (86.3)</td>
<td>6934 (85.6)</td>
<td>6447 (77.6)</td>
</tr>
<tr>
<td>Tampa</td>
<td>2080 (75.9)</td>
<td>1724 (62.9)</td>
<td>10,442 (84.3)</td>
<td>9120 (71.5)</td>
</tr>
<tr>
<td>Washington DC</td>
<td>1527 (84.6)</td>
<td>1296 (71.8)</td>
<td>9293 (70.9)</td>
<td>7748 (56.1)</td>
</tr>
<tr>
<td>West Palm Beach</td>
<td>1124 (75.6)</td>
<td>950 (63.9)</td>
<td>5600 (77.9)</td>
<td>4865 (62.0)</td>
</tr>
</tbody>
</table>

aTimely diagnosis and linkage to care “n (%)” represents counts and percentages for 2014 through 2018.
bReceipt of care and viral suppression “n (%)” represents counts and percentages for 2018.
dThis section represents national estimates—not all 36 cities combined.
eN/A: not applicable.
fThe regional percentage for linkage to care does not include data from Providence.
gThe regional percentage for linkage to care does not include data from Denver.
hEach of these cities represents a metropolitan statistical area.
iWithin 25% of meeting NHAS goal.
j>25% of meeting NHAS goal.
kMet or surpassed NHAS goal.
lEach of these cities represents multiple counties.
mEach of these cities represents a single county.
n—: not available.

Cities’ Progress Toward the HIV Care Continuum Goals

Overall, cities were closest to meeting the 2020 goals for timely diagnosis and receipt of care and struggled the most with meeting the goal for viral suppression (Table 1). The 2020 NHAS goal specified that at least 85% of people with newly diagnosed HIV should be linked to care. At the end of 2018, most of the 34 cities that provided data for linkage to care (n=20, 59%) were within 25% of the goal; 1 city (Seattle) surpassed the goal. All but 3 of the 36 cities were at least within 25% from meeting the receipt of care goal. For viral suppression, most cities (n=25, 69%) were not within 25% of the goal, and no cities met or surpassed the goal. Table S1 in Multimedia Appendix 1 contains a color-coded version of Table 1, with red, yellow, and green colors indicating categories of progress toward each of the HIV care continuum indicators’ NHAS goals.

Cities’ Progress Toward the HIV Care Continuum Goals by Priority Populations

Table 2 displays the percentage of AIDSVu cities that were within 25% of meeting, had met, or had surpassed the NHAS goal for each HIV care continuum indicator within sex, race or ethnicity, and age groups. Overall, receipt of care was the indicator that had the highest percentage of cities that were at least within 25% of meeting the goal for select demographic groups, 97% (n=35) of cities met, surpassed, or were close to meeting the goal for females, Black people, and 13 to 24-year-olds. Viral suppression was the indicator for which most group-specific city proportions were not close to meeting the NHAS goal. Within the races and ethnicities group, the White population had the highest percentage of cities that were close to meeting, had met, or surpassed the goal for the viral suppression indicator. Only 1 demographic group—13 to 24-year-olds—had all 36 cities at least within 25% of meeting the goal for timely diagnosis. Compared to other race and ethnicity groups, the Black population had the highest percentage of cities that were close to meeting, had met, or surpassed the timely diagnosis and receipt of care goals (n=35, 97% for both). The greatest variation in percentages within a demographic category across all indicators was found for age groups. Compared to the younger age groups, older age groups had substantially lower percentages of cities whose residents had timely knowledge of their HIV+ status; however, the opposite trend was seen for the linkage to care and viral suppression indicators across age groups. Across the race and ethnicity categories, the greatest variation in percentages of cities close to meeting or having met or surpassed the NHAS goals was seen for viral suppression.
Table 2. Percentage of AIDSVu cities that were at least within 25% of achieving each National HIV/AIDS Strategy HIV care continuum indicator goal by sex, race or ethnicity, and age, as of 2018.

<table>
<thead>
<tr>
<th></th>
<th>Timely diagnosis</th>
<th>Linkage to care</th>
<th>Receipt of care</th>
<th>Viral suppression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, n (%)</td>
<td>36 (100)</td>
<td>21 (62)</td>
<td>33 (92)</td>
<td>11 (31)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>35 (97)</td>
<td>20 (59)</td>
<td>32 (89)</td>
<td>13 (36)</td>
</tr>
<tr>
<td>Female</td>
<td>33 (92)</td>
<td>23 (68)</td>
<td>35 (97)</td>
<td>13 (36)</td>
</tr>
<tr>
<td><strong>Race or ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>35 (97)</td>
<td>16 (47)</td>
<td>35 (97)</td>
<td>7 (19)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>33 (92)</td>
<td>24 (71)</td>
<td>27 (75)</td>
<td>11 (31)</td>
</tr>
<tr>
<td>White</td>
<td>34 (94)</td>
<td>27 (79)</td>
<td>31 (86)</td>
<td>23 (64)</td>
</tr>
<tr>
<td><strong>Age groups (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-24</td>
<td>36 (100)</td>
<td>15 (44)</td>
<td>35 (97)</td>
<td>11 (31)</td>
</tr>
<tr>
<td>25-44</td>
<td>35 (97)</td>
<td>22 (65)</td>
<td>34 (94)</td>
<td>5 (14)</td>
</tr>
<tr>
<td>45-59</td>
<td>7 (19)</td>
<td>31 (91)</td>
<td>32 (89)</td>
<td>20 (56)</td>
</tr>
<tr>
<td>60+</td>
<td>6 (17)</td>
<td>30 (88)</td>
<td>34 (94)</td>
<td>21 (58)</td>
</tr>
</tbody>
</table>

*aOnly 34 AIDSVu cities are represented in the percentages. Denver and Providence did not provide data for linkage to care.*

**Discussion**

**Principal Findings**

The HIV care continuum can be a useful public health tool to characterize the state of the HIV epidemic in defined areas. Much of the progress along the continuum is framed in the context of national goals, but all public health is local, and national goals will not be achieved unless they are consistently met in the cities with the largest HIV epidemics. Cities have a strong understanding of their community’s perceptions of HIV and the needs of those affected by and infected with HIV, which can be used to tailor and focus efforts. For example, in 2016 New York City clinical providers and community-based organizations created the PlaySure Network [8]. This network is dedicated to making HIV-related services such as testing, treatment, and pre-exposure prophylaxis available to all New Yorkers [8].

Here, we expand the geographic level of analysis of the HIV care continuum beyond the county-level data disseminated by the CDC [9]. A total of 27 of the 36 cities included here are located within Ending the HIV Epidemic jurisdictions [10]. The 36 cities in this analysis illustrate progress in meeting national goals overall and for particular race or ethnicity, sex, and age groups [11]. Understanding this level of patterns of inequities in access to and maintained engagement in HIV care can inform local public health action needed to improve the HIV care continuum.

Many cities publish their own surveillance reports, but different cities may use different methods to calculate key indicators. To date, a consistently derived set of indicators that surpasses the depth of the county level for at least some cities has not been available. Similarly, examining NHAS goals at the local level has a great impact. All cities were within 25% of meeting the timely diagnosis goal, but varied in how close they were to achieving the linkage to care and receipt of care goals, with 38% (n=13) of cities >25% from meeting the linkage to care goal and only 1 city that met or surpassed the linkage to care goal. Viral suppression posed the greatest challenge for cities, with only 31% (n=11) of cities falling within 25% of meeting the NHAS goal and none meeting or surpassing the goal. Among cities, there were substantial differences in meeting linkage to care and viral suppression goals among racial or ethnic groups. Finally, a higher percentage of cities were close to or met or surpassed the NHAS goals for younger populations (13 to 24 and 25 to 44-year-olds) compared to older ones (45 to 59-year-olds and those aged 60 or older) for timely diagnosis; in contrast, compared to all other age groups, the highest percentage of cities (58%, n=21) that were close to or met or surpassed the goal for viral suppression was for the oldest population (60 years of age or older). The former may reflect older people’s perception of a lower risk of HIV acquisition, while the latter may signal that younger people are less likely to stay engaged in care [12]. Additionally, younger people are unlikely to be diagnosed 8 years or more after their infection (the estimated duration of infection associated with late diagnosis) because their infection most likely happened between ages 15 and 25 years.

Among the 36 cities, there were none that achieved or surpassed the timely diagnosis NHAS goal for 2020. Thus, there is room for improvement in diagnosing people with HIV in a timely manner for cities overall. The CDC recommends that everyone aged 13 to 64 years be screened for HIV once in their lifetime, with those with certain risk factors tested annually [13]. Routine testing for those at high risk of acquiring HIV can make substantial inroads to increasing the number of timely diagnoses, but it will not allow cities to reach 100% of all those at risk of acquiring HIV. Anyone can contract HIV, so even for cities that are close to meeting the timely diagnosis goal, the last few who are not diagnosed in a timely manner are the hardest to reach.
The ability of providers and patients to assess risks for HIV may not be optimal. Some studies suggest that older people are more likely to be diagnosed with advanced HIV, and women—as well as older people—have a higher likelihood of being diagnosed with HIV in a hospital [12,14]. This finding may reflect low perceived risk among these demographic groups, leading to less testing as a part of routine care. Additionally, fear of stigma may also play a role in individuals not requesting or being offered testing at younger ages or outside of hospital settings. Offering HIV tests as part of routine primary care may therefore help increase the number of timely diagnoses overall.

Young people, who are often healthy, may not feel compelled to regularly engage with the health care system, so cities should focus on installing and making low-cost HIV testing services available at more accessible locations such as retail pharmacies and at home (eg, self-test kits) [15]. CDC recently announced a multiyear program to distribute HIV self-test kits directly to people who request them; this program will include direct marketing to populations with increased risks for HIV [16]. Cities will be able to incorporate referrals to the program into local prevention programming. The idea of distributing self-test kits by mail may be more acceptable after the national experience with government distribution of COVID-19 self-test kits [17].

Local, structural, and social factors can play a significant role in the ability of people with HIV to be linked to care. When there is no formal system to foster tight-knit communication between public health facilities, HIV testing sites, and HIV providers, it can be difficult to effectively link people newly diagnosed with HIV to care. Shortages of local providers who can treat people with HIV may lead to long wait times for appointments, potentially interfering with initiation of care [18-20]. It is also important to consider the local communities’ general attitudes toward HIV as they may discourage people from showing up at facilities that others know may be associated with treatment for HIV [19]. Implementation of the following has been found to promote successful linkage to care: patient navigators, mental health services, care coordination, staff training, and reduced time to provider appointments [18].

Our data highlight the importance of stratifying the HIV care continuum by demographic populations. Priority populations may differ across cities and existing prevention and care interventions may not be designed for the priority population of interest, thus requiring adaptations of the intervention. One study that constructed a dynamic HIV transmission model incorporated information regarding HIV micro epidemics within 6 differing US cities and evidence-based interventions to determine what cost-effective combination of approaches resulted in the most health benefit [21]. The researchers found that approaches needed to be tailored to each of the cities because they each had varying levels of high-risk, vulnerable populations for which some interventions worked better than others [21]. Using consistent methods of deriving HIV care continuum indicators can enable cities to discover commonalities, such as similar high-risk populations, that may promote discussions of best practices more easily.

**Limitations of the General Field**

In this report, we presented the proportion of timely diagnoses among people diagnosed with HIV in 36 cities from 2014 to 2018. However, late diagnosis (the inverse of timely diagnosis) has been found to be flawed and thus, late diagnosis and timely diagnosis indicators should be used with caution because they are affected by both HIV incidence and testing [22-24]. We recommend CDC dropping these indicators and developing a new one to monitor HIV diagnosis, for example, the probability of diagnosis within 1 year of infection [25]. The timely and late diagnosis definitions are also imperfect in that they may not accurately account for those who rapidly progress to AIDS despite having tested negative for HIV within the prior year or 2.

Intermittent access to care can be difficult to isolate from the receipt of care definition because the receipt of care definition relies on testing alone. Using the CDC definition for receipt of care means that people who seek care but do not receive either test for whatever reason will be excluded.

**Limitations of the Study**

Our study has several limitations. It only included data from people who were diagnosed. Thus, we do not provide insight into what populations of people are living with undiagnosed HIV. The lack of data on undiagnosed HIV means that our percentages for the 4 indicators underrepresent the true estimates. More work needs to be done to fully capture the entire population of people with HIV. Furthermore, to measure receipt of care and viral suppression, we used HIV case reporting data to estimate the denominator, that is, the number of people with HIV at the end of 2018. It has been shown that such a method can overestimate the number of people with HIV in a jurisdiction [26]. Therefore, we may have overestimated our denominator and underestimated receipt of care and viral suppression. However, because the same method was used across all 36 cities, the rankings of receipt of care and viral suppression by city and region were likely to be valid. It is recommended that HIV laboratory reporting data be used to estimate the number of people with HIV to counter the flaws with HIV case reporting data [26]. However, while this approach may lead to more accurate estimates in jurisdictions that have laws requiring full reporting of CD4 lymphocyte or viral load data, only 47 states and the District of Columbia had such laws as of December 2019 [6]. Among the 36 cities in our analysis, Newark and Philadelphia were in states that did not have such laws [6].

CITIES themselves decided whether they provided input for their city alone, the county that encompassed their city, multiple counties including the one with their city, or an MSA that included their city. Cities’ data reporting was not complete, including exclusions of cases missing a zip code and exclusions due to low case count at the zip code level and for specific cities (eg, linkage to care for Denver and Providence). Additionally, not all cities had the same capacity to deduplicate their data, contributing further to the potential inflation of denominators. Also, these data do not include categories for 2 groups of people who are known to be at increased risk for HIV acquisition—men who have sex with men and people who inject drugs. Furthermore, we did not have information on nativity.
These comparisons were made with the idea that our unit of analysis, a city, was broadly defined with cities themselves deciding how their jurisdiction was defined—whether that be at the city, single county, multicounty, or MSA level. Finally, we do not include data for all major cities across the United States, such as those in California which have their own reporting mechanisms. Therefore, we are not capturing the full spectrum of progress toward the HIV Care continuum goals in all major cities or the 4 regions across the United States.

**Public Health Implications**

National strategy goals cannot be achieved without first meeting these goals at the city level. Cities represent a fundamental unit for change, with cities knowing how to best reach and support the communities that dwell within their bounds. Cities have often been at the forefront of enacting fundamental policies—such as certain cities’ that make it their practice to provide pre-exposure prophylaxis at minimal cost—to prevent the transmission of HIV and improve the lives of those living with HIV. Similar to the SARS-CoV-2 vaccination for which about 67% of the US population had received at least 1 dose of the vaccine as of July 2021, with the remaining population being more difficult to vaccinate, it will also be difficult to reach the remaining people in the US population who have not received their diagnosis of HIV [27]. Advocating for primary care practices to offer HIV tests to general patient panels, rather than only high-risk patients, can facilitate timely diagnosis for all people at risk of acquiring HIV. Our data suggest that linkages to care present challenges to some cities. Cities can encourage their public health departments that may offer HIV testing to create strong ties and relationships with those that provide HIV treatment services so that there is minimal delay and a smooth transfer of patients to HIV providers. Most cities in our study seem to have some success in ensuring that people receive HIV care once they surpass the barrier of being linked to it in the first place. Our data suggest opportunities for improvement at the far end of the HIV care continuum, namely, viral suppression, which is the culmination of dedicated commitment and adequate and acceptable access to care. Those at highest risk for acquiring HIV may sometimes be some of the most vulnerable people in our cities who need the most resources to ensure that their local environment recognizes their needs and addresses them in a timely, steady, and supportive manner. With the nation’s renewed efforts to end the HIV epidemic, cities—especially high-prevalence ones—are receiving even more dedicated monies to combat this issue, making achieving national HIV care continuum goals more possible than ever before.

**Acknowledgments**

AIDSVu Health Department Collaborators are Qiang Xia, MD, MPH; Sarah Braunstein, PhD, MPH (Bureau of Hepatitis, HIV and STIs, New York City Department of Health and Mental Hygiene); Samira Khan, MPH, MSW (South Carolina Department of Health and Environmental Control); Lorene Maddox, MPH (Florida Department of Health); Lauren Ostrenga, MPH (Louisiana Department of Health); Caitlin Murano, MPH (New Jersey Department of Health); Vipul Shukla, MS, MPH; Abby Winkler, MPH (Wisconsin Department of Health Services); Tiffany Wilson, MPH (Alabama Department of Public Health). AIDSVu is supported by a grant from Gilead Sciences.

**Data Availability**

The data sets analyzed during this study are available from the state or local health departments on reasonable request.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1


[PDF File (Adobe PDF File), 99 KB - publichealth_v10i1e49381_app1.pdf ]

**References**


13. Screening for HIV. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/clinicians/screening/index.html [accessed 2023-12-19]


Abbreviations

CDC: Centers for Disease Control and Prevention
MSA: metropolitan statistical area
NHAS: National HIV/AIDS Strategy

Edited by A Mavragani; submitted 26.05.23; peer-reviewed by E Mokotoff, S Buskin; comments to author 13.09.23; revised version received 07.11.23; accepted 16.12.23; published 26.02.24.

Please cite as:
Hood N, Benbow N, Jaggi C, Whitby S, Sullivan PS, AIDSVu Health Department Collaborators
AIDSVu Cities’ Progress Toward HIV Care Continuum Goals: Cross-Sectional Study
JMIR Public Health Surveill 2024;10:e49381
URL: https://publichealth.jmir.org/2024/1/e49381
doi:10.2196/49381
PMID:38407961

©Nicole Hood, Nanette Benbow, Chandni Jaggi, Shamaya Whitby, Patrick Sean Sullivan, AIDSVu Health Department Collaborators. Originally published in JMIR Public Health and Surveillance (https://pubhealth.jmir.org), 26.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://pubhealth.jmir.org, as well as this copyright and license information must be included.
Characteristics of Hepatitis B Virus, Hepatitis C Virus, and Syphilis Coinfection in People With HIV/AIDS Contracted Through Different Sources: Retrospective Study

Rongrong Yang1,2, MD; Rui Yuan1, BA; Xien Gui1,2, BA; Hengning Ke1,2, MD; Ke Zhuang1, MD; Hui Hu1,2, MSN; Ling Li1, MSN; Ling Feng1,2, BSN; Xingxia Yu4*, MM; Yajun Yan1,2*, MM; Mingqi Luo1,2*, MD

1Zhongnan Hospital of Wuhan University, Wuhan, China
2Center for AIDS Research, Wuhan University, Wuhan, China
3Animal Biosafety Level-III Laboratory at the Center for Animal Experiment, State Key Laboratory of Virology, Wuhan University, Wuhan, China
4Department of Emergency, Renmin Hospital, Wuhan University, Wuhan, China
*these authors contributed equally

Corresponding Author:
Mingqi Luo, MD
Center for AIDS Research
Wuhan University
Donghu Road 169, Wuchang District
Hubei Province
Wuhan, 430071
China
Phone: 86 +02767812881
Email: mingqiluo@163.com

Abstract

Background: The burden of hepatitis B virus (HBV), hepatitis C virus (HCV), and syphilis coinfections remains disproportionately high among people living with HIV/AIDS. Hubei province is located in central China, where there are distinct regional characteristics of the distribution of people living with HIV/AIDS acquired via diverse transmission routes and the AIDS epidemic itself.

Objective: We aimed to estimate the magnitude of HBV, HCV, or syphilis coinfections among people living with HIV/AIDS with blood-borne transmission, which includes former paid blood donors, contaminated blood recipients, and intravenous drug users, as well as among people with sex-borne HIV transmission (including heterosexual people and men who have sex with men) and people with mother-to-child HIV transmission.

Methods: From January 2010 to December 2020, people living with HIV/AIDS were tested for hepatitis B surface antigen (HBsAg), HCV antibodies, and syphilis-specific antibodies. The positive patients were further tested for HBV markers, HBV DNA, and HCV RNA, and received a rapid plasma reagin circle card test. All people living with HIV/AIDS were first divided into transmission groups (blood, sex, and mother-to-child); then, people with blood-borne HIV transmission were divided into former paid blood donors, contaminated blood recipients, and intravenous drug users, while people with sex-borne HIV transmission were divided into heterosexual people and men who have sex with men.

Results: Among 6623 people living with HIV/AIDS, rates of chronic HCV infection were 80.3% (590/735) in former paid blood donors, 73.3% (247/337) in intravenous drug users, 57.1% (444/777) in contaminated blood recipients, 19.4% (21/108) in people with mother-to-child HIV transmission, 8.1% (240/2975) in heterosexual people, and 1.2% (21/1691) in men who have sex with men. Chronic HBV infection rates were similar among all people with blood-borne HIV transmission. However, compared to heterosexual people, the chronic HBV infection rate was greater in men who have sex with men (213/1691, 12.6% vs 308/2975, 10.4%; \( \chi^2 = 5.469; P = .02 \)), although HBV exposure was less common (827/1691, 48.9% vs 1662/2975, 55.9%; \( \chi^2 = 20.982; P < .001 \)). Interestingly, the combination of HBsAg and hepatitis B e antigen (HBeAg) was found in 11 patients with sex-borne HIV transmission, but in 0 people with blood-borne HIV transmission (11/196, 5.6% vs 0/521, 0%; \( \chi^2 = 29.695, P < .001 \)). In people with sex-borne HIV transmission, the proportions of patients with a syphilis titer \( \geq 1:16 \) and neurosyphilis were 8.6% (105/1227) and 7.8% (37/473), respectively, whereas these values were 0 in people with blood-borne HIV transmission.
Conclusions: In people living with HIV/AIDS, HCV transmission intensity was significantly associated with specific exposure modes of blood or sexual contact. The rate of chronic HBV infection among men who have sex with men was higher than in any other population. Attention should be paid to the high prevalence of neurosyphilis in people living with HIV/AIDS who contract HIV by sexual intercourse.

(JMIR Public Health Surveill 2024;10:e46750) doi:10.2196/46750

KEYWORDS
acquired immunodeficiency syndrome; AIDS; human immunodeficiency virus; HIV; hepatitis B virus; HBV; hepatitis C virus; HCV; syphilis

Introduction
According to a 2020 World Health Organization (WHO) report, 37.6 million people are infected with HIV worldwide [1]. Hepatitis B virus (HBV), hepatitis C virus (HCV), and syphilis infection are also all major public health problems worldwide. HBV, HCV, and syphilis are the 3 most common infections among people living with HIV/AIDS all over the world and share similar transmission routes [2,3]. The burden of HBV, HCV, and syphilis infection in people living with HIV/AIDS is even larger than in the general population. The ever-increasing burden of these infections has become a growing concern [4]. The prevalence of HBV, HCV, and syphilis coinfections in people living with HIV/AIDS is 8% to 12.2% [5-7], 33% to 80% [8,9], and 40.5% [10], respectively.

With the apparent efficiency of infectious illness prevention and control, as well as the recent introduction of direct-acting antiviral medicines, the WHO has proposed long-term worldwide prevention and control as the ultimate goal. A full assessment of HBV, HCV, and syphilis coinfection in people living with HIV/AIDS will aid in the development of a phased management plan that will contribute to the global eradication of hepatitis B disease by 2030 [11] and make HCV eradication a reality [12].

As the distribution of high-risk groups varies by country, so does the burden of HBV, HCV, and syphilis coinfection. Currently, the characteristics of the AIDS epidemic and the distribution of high-risk populations have changed. On the one hand, people who became infected with HIV as a result of nonroutine HIV screening of blood products still have HBV or HCV-related diseases; on the other hand, with the increasing transmission of AIDS among men who have sex with men, the pressure to control syphilis transmission differs from the blood transmission era of AIDS and heterosexual unprotected contact. Although HIV, HBV, HCV, and syphilis infections have been reported in a wide range of populations, including prisoners [13], blood donors [14], parturient women [15], and even inpatients [16], data on comparative infection rates among different populations in the same region remain scarce. A more extensive comparison of the prevalence and transmission intensity of coinfection with HBV, HCV, and syphilis among different risk groups will serve as the foundation for a comprehensive assessment of the disease burden in the HIV-positive community.

The Hubei provincial region includes populations of all high-risk HIV-infected groups and can be described as an appropriate place to examine the coinfection status of HBV, HCV, and syphilis in different populations of people living with HIV/AIDS in the same province. Hubei is located in central China. Poor rural farmers supplied plasma under unsanitary conditions in the 1990s, resulting in HIV infection among former paid blood donors and contaminated blood recipients [17-19]. Later, as a result of imported cases, an AIDS epidemic began to spread among intravenous drug users, heterosexual people, and men who have sex with men. As a result, the Hubei province AIDS epidemic has distinct geographical characteristics.

In this study, HIV-infected people were classified according to 3 classical sources of transmission: blood, sex, and mother-to-child transmission (MTCT); further subgroup analysis was performed in the blood-source and sex-source populations, with the goal of comparing differences in the disease prevention and control burden of HIV combined with HBV, HCV, and syphilis among various high-risk groups. This can not only maximize the allocation of comprehensive disease prevention and control expenditures in resource-rich locations, but it can also serve as a reference point for resource-limited places to achieve larger returns with a small investment, allowing for progress in reaching the final goal.

Methods

Study Population
From January 2010 to December 2020, a total of 6623 people living with HIV/AIDS were tested for hepatitis B surface antigen (HBsAg), HCV antibodies, and syphilis-specific antibodies at Zhongnan Hospital of Wuhan University. Patients who tested positive for HBsAg were also tested for HBV markers and HBV DNA. HCV RNA was investigated further in HCV antibody–positive patients, and the rapid plasma reagin circle card test was used in syphilis antibody–positive patients. First, all people living with HIV/AIDS were divided into 3 groups based on their HIV acquisition route: blood, intercourse, or MTCT. They were then redivided into former paid blood donors and contaminated blood recipients, and intravenous drug users in the blood-borne transmission group, and heterosexual people and men who have sex with men in the sex-borne transmission group. The inclusion criteria were (1) confirmed HIV infection and (2) a definite route of HIV acquisition. The exclusion criteria were (1) having 2 or more high-risk behaviors for HIV infection, (2) acute HBV infection, and (3) acute HCV infection.

Ethical Considerations
This study was a secondary analysis of a preexisting data set, and ethical review and approval were waived for this study by
the institutional review board of the work unit (Zhongnan Hospital of Wuhan University) of the first author (RY). Informed consent was obtained, and all the analyzed data were anonymous.

**Study Protocol**

Differences in HBV, HCV, and syphilis transmission intensity were analyzed in groups divided according to the 3 classical transmission routes of HIV acquisition: blood, sex, and MTCT. Then, in the blood-origin population, the participants were further divided into former paid blood donors, contaminated blood recipients, and intravenous drug users, whereas the sex-origin population was further divided into heterosexual people and men who have sex with men. Differences in HBV or HCV coinfection rates and characteristics of syphilis infection were compared in the different high-risk groups with the same HIV origin.

**Laboratory Testing**

Two rapid assays or joint detection modes were used to assess HIV status: enzyme immunoassay (EIA) and Western blotting. Anti-HCV, HBsAg, and *Treponema pallidum* antibody detection testing were done with a third-generation EIA (Shanghai Kehua Bio-Engineering). HBV markers, including HBsAg, hepatitis B surface antibody (anti-HBs), hepatitis B e antigen (HBeAg), hepatitis B e antibody (anti-HBe), and hepatitis B core antibody (anti-HBc), were examined with an enzyme-linked immunosorbent assay (ELISA). HCV RNA was tested using a polymerase chain reaction fluorescence detection kit (Shanghai Kehua Bio-Engineering) with a lower limit of detection of 20 copies/ml. The conventional rapid plasma reagin (RPR) circle card test, which uses cardiolipin antigen with a carbon particle to detect reagin, was used to determine the titer of syphilis. The experimental procedure and the analysis of results were performed according to the manufacturer’s instructions.

**Definitions**

Passive immunization is defined as the presence of anti-HBe and a vaccination history. Natural immunization is characterized as having a vaccine history and HBV markers in addition to anti-HBe, or not having a vaccination history but having HBV markers. Natural HBV exposure was defined as not having had a hepatitis B vaccine but having one or more HBV indicators or having received a hepatitis B vaccine and having one or more HBV markers except for anti-HBe.

As in a previous report [20], a reactive cerebrospinal fluid (CSF) venereal disease research laboratory test result, a CSF white blood cell count of >20 cells/l, or both were used to define neurosyphilis. Patients who had received penicillin previous to lumbar puncture or who had missing data, such as from a full skin examination or a complete neurological examination, were excluded from this study.

**Statistical Analysis**

Differences in characteristics across groups were compared using the Wilcoxon rank-sum test (2-group comparisons) and the Kruskal-Wallis rank test (≥3 group comparisons). The Pearson chi-square test was used to compare categorical variables across groups. A 2-tailed t test was applied to compare differences in measurement data. Statistical calculations were performed using SPSS (version 17.0; IBM Corp), with a P value <.05 considered significant.

**Results**

**Demographic Characteristics of People Living With HIV/AIDS**

This study investigated 6623 people living with HIV/AIDS from January 2010 to December 2020, of whom 71.27% (4720/6623) were male. Their average age was 37.6 (SD 12.3) years, and most (n=5838, 88.15%) were aged 18 to 55 years. Most (n=5805, 87.65%) of the individuals were receiving antiretroviral therapy (ART), and the average CD4+ T lymphocyte count was 226 (SD 86) cells/ul. The routes of HIV acquisition for all individuals are shown in Table 1. On the whole, the positive rates for HBsAg, anti-HCV, and antisyphilis antibodies were 10.92% (723/6623), 23.6% (1563/6623), and 21% (1391/6623), respectively.
Table 1. Demographic characteristics of the study population (N=6623).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>People living with HIV/AIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>4720 (71.27)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>1903 (28.73)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>≤18</td>
<td>138 (2.08)</td>
</tr>
<tr>
<td>19-35</td>
<td>2974 (44.9)</td>
</tr>
<tr>
<td>36-55</td>
<td>2864 (43.24)</td>
</tr>
<tr>
<td>≥56</td>
<td>647 (9.77)</td>
</tr>
<tr>
<td><strong>HIV transmission route, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Former paid plasma donation</td>
<td>735 (11.1)</td>
</tr>
<tr>
<td>Contaminated blood transfusion</td>
<td>777 (11.73)</td>
</tr>
<tr>
<td>Intravenous drug use</td>
<td>337 (5.09)</td>
</tr>
<tr>
<td>Mother-to-child transmission</td>
<td>108 (1.63)</td>
</tr>
<tr>
<td>Men who have sex with men contact transmission</td>
<td>1691 (25.53)</td>
</tr>
<tr>
<td>Heterosexual contact transmission</td>
<td>2975 (44.92)</td>
</tr>
<tr>
<td>CD4+ T lymphocyte count (cells/ul), mean (SD)</td>
<td>226 (86)</td>
</tr>
<tr>
<td>Antiretroviral therapy coverage, n (%)</td>
<td>5805 (87.65)</td>
</tr>
<tr>
<td>Hepatitis B surface antigen positive, n (%)</td>
<td>723 (10.92)</td>
</tr>
<tr>
<td>Anti–hepatitis C virus positive, n (%)</td>
<td>1563 (23.6)</td>
</tr>
<tr>
<td>Hepatitis C virus RNA detectable, n (%)</td>
<td>1112 (16.79)</td>
</tr>
<tr>
<td>Antisyphilis antibody positive, n (%)</td>
<td>1391 (21)</td>
</tr>
<tr>
<td>Positive in rapid plasma reagin circle card test, n (%)</td>
<td>520 (7.85)</td>
</tr>
</tbody>
</table>

Figure 1A and Figure 1B depict the study location and population geographic distribution, whereas Figure 1C, Figure 1D, and Figure 1E depict the overall positive rates for HBsAg, HCV, and syphilis, respectively.
**Figure 1.** Maps of the study area, sample sites, and prevalence of hepatitis B virus, hepatitis C virus (HCV), or syphilis coinfections among people living with HIV/AIDS. Figures 1A and 1B depict the study location and population geographic distribution. Figures 1C, 1D, and 1E depict the overall positive rates of hepatitis B surface antigen (HBsAg), HCV, and syphilis, respectively.

**Coinfection Rates of HBV and HCV in Patients Who Acquired HIV via the 3 Routes Of Transmission**

According to the route of HIV acquisition, the 6623 patients were divided into those with blood-borne transmission (n=1849), MTCT (n=108), and sex-borne transmission (n=4666). The coinfection rates of HBV were 10.6% (196/1849), 5.56% (6/108), and 11.17% (521/4666), respectively. There was no significant difference in HBV coinfection rate among the 3 groups ($\chi^2=3.680; P=.16$). However, the coinfection rates of HCV in people with blood-borne transmission, MTCT, and sex-borne transmission were 69.28% (1281/1849), 19.44% (21/108), and 5.59% (261/4666), respectively. These HCV infection rates were significantly different among the 3 groups ($\chi^2=2980.043; P<.001$). People with blood-borne HIV transmission had a higher prevalence of HCV coinfection than people with MTCT (1281/1849, 69.28% vs 21/108, 19.44%; $\chi^2=113.812; P<.001$) and people with sex-borne HIV transmission.
transmission (1281/1849, 69.28% vs 261/4666, 5.59%; \( \chi^2_{1}=2973.003; P<.001 \)); people with MTCT had a higher prevalence of HCV coinfection than people with sex-borne transmission (21/108, 19.44% vs 261/4666, 5.59%; \( \chi^2_{1}=36.434; P<.001 \)).

### Coinfection Rates of HBV and HCV in Different Populations With Same Origin of HIV

According to differences in risk factors, 1849 people living with HIV/AIDS who acquired HIV by blood were further divided into former paid blood donors (n=735), contaminated blood recipients (n=777), and intravenous drug users (n=337). The coinfection rates of HBV were 9.9% (73/735), 10.2% (79/777), and 13.1% (44/337), respectively. There was no significant difference in the HBV coinfection rate among individuals who acquired HIV through blood transmission (\( \chi^2_{1}=2.645; P=.27 \)). However, HCV infection rates in former paid blood donors, intravenous drug users, and contaminated blood recipients were 80.3% (590/735), 73.3% (247/337) and 57.1% (444/777), respectively, indicating a statistically significant difference between the three groups (\( \chi^2_{1}=98.060, P<.001 \)). Former paid blood donors had a higher prevalence of HBV coinfection than intravenous drug users (590/735, 80.3% vs 247/337, 73.3%; \( \chi^2_{1}=6.574; P=.01 \)) and contaminated blood recipients (590/735, 80.3% vs 444/777, 57.14%; \( \chi^2_{1}=93.462; P<.001 \)), and intravenous drug users had a higher prevalence of HCV coinfection than contaminated blood recipients (247/337, 73.3% vs 444/777, 57.1%; \( \chi^2_{1}=26.032; P<.001 \)). Similarly, 4666 people living with HIV/AIDS who acquired HIV by sexual contact were further divided into heterosexual people (n=2975) and men who have sex with men (n=1691). The coinfection rates of HBV in men who have sex with men were greater than in heterosexual people (213/1691, 12.6% vs 308/2975, 10.35%; \( \chi^2_{1}=5.469; P=.02 \)), but the reverse was the case for HCV coinfection (21/1691, 1.24% vs 240/2975, 8.07%; \( \chi^2_{1}=95.113; P<.001 \)).

### HBV Exposure and Chronic HBV Coinfection Among Individuals Who Acquired HIV Through Sexual Intercourse

The coverage of active immunization against HBV infection in men who have sex with men was greater than in heterosexual people (534/1691, 31.58% vs 574/2975, 19.29%; \( \chi^2_{1}=89.860; P<.001 \)), and the reverse was the case for passive immunity (614/1691, 36.31% vs 1354/2975, 45.51%; \( \chi^2_{1}=37.440; P<.001 \)). Compared with heterosexual people, the natural HBV exposure rate was lower in men who have sex with men (827/1691, 48.91% vs 1354/2975, 55.87%; \( \chi^2_{1}=20.982; P<.001 \)), but the chronic infection rate following natural exposure to HBV was greater in men who have sex with men (213/827, 25.76% vs 308/1662, 18.53%; \( \chi^2_{1}=17.412; P<.001 \)). These results are shown in Table 2.

#### Table 2. Hepatitis B virus (HBV) exposure and chronic HBV coinfection among individuals who acquired HIV through sexual intercourse.

<table>
<thead>
<tr>
<th></th>
<th>Heterosexual people (n=2975)</th>
<th>Men who have sex with men (n=1691)</th>
<th>Chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive immunization, n (%)</td>
<td>574 (19.29)</td>
<td>534 (31.58)</td>
<td>89.860 (1)</td>
<td>.001</td>
</tr>
<tr>
<td>Natural immunization, n (%)</td>
<td>1354 (45.51)</td>
<td>614 (36.31)</td>
<td>37.440 (1)</td>
<td>.001</td>
</tr>
<tr>
<td>Overall</td>
<td>254 (8.54)</td>
<td>125 (7.39)</td>
<td>1.897 (1)</td>
<td>.17</td>
</tr>
<tr>
<td>Anti-HBe/anti-HBc</td>
<td>515 (17.31)</td>
<td>363 (21.47)</td>
<td>12.189 (1)</td>
<td>.001</td>
</tr>
<tr>
<td>Anti-HBs/anti-HBc</td>
<td>106 (3.56)</td>
<td>46 (2.72)</td>
<td>2.430 (1)</td>
<td>.12</td>
</tr>
<tr>
<td>Anti-HBc</td>
<td>479 (16.1)</td>
<td>80 (4.73)</td>
<td>132.176 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Natural HBV exposure, n (%)</td>
<td>1662 (55.87)</td>
<td>827 (48.91)</td>
<td>20.982 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Proportion of chronic HBV infection to HBV natural exposure population, n/N (%)</td>
<td>308/1662 (18.53)</td>
<td>213/827 (25.76)</td>
<td>17.412 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chronic HBV infection, n (%)</td>
<td>308 (10.35)</td>
<td>213 (12.6)</td>
<td>5.469 (1)</td>
<td>.02</td>
</tr>
</tbody>
</table>

a Anti-HBe: hepatitis B e antibody.
b Anti-HBs: hepatitis B surface antibody.
c Anti-HBc: hepatitis B core antibody.

### Uncommon Combination of HBV Markers Among Chronic HBV Infections

In people with blood- or sex-borne HIV transmission, 196 and 521 people, respectively, were found to be coinfected with HBV in this study. Three uncommon combinations of HBV markers were detected in these chronic HBV infection patients: HBsAg and anti-HBe positivity; HBeAg and anti-HBs positivity; and only HBsAg and HBeAg positivity. Among those with blood-borne HIV transmission, 6 of 196 (3.1%) people with HIV/HBV coinfection had HBsAg and anti-HBe positivity, compared to 12 of 521 (2.3%) people with sex-borne HIV transmission. The rates of HBsAg and anti-HBe positivity showed no statistical difference between the people with blood-
or sex-borne HIV transmission (6/196, 3.1% vs 12/521, 2.3%; \( \chi^2 = 0.334; P = .56 \)). Similarly, 1 of 196 (0.5%) people with HIV/HBV coinfection among those with blood-borne HIV transmission had HBeAg and anti-HBs positivity, compared to 8 of 521 (1.5%) people with sex-borne HIV transmission. The rates of HBeAg and anti-HBs positivity showed no statistical difference between people with blood- or sex-borne HIV transmission (1/196, 0.5% vs 8/521, 1.5%; \( \chi^2 = 1.208; P = .27 \)).

It was interesting that when testing HBV markers, coinfection with only HBsAg/HBeAg positivity was found in 11 patients with HIV/HBV among people with sex-borne HIV transmission, while none of the people with blood-borne HIV transmission (11/196, 5.6% vs 0/521, 0%; \( \chi^2 = 29.695; P < .001 \)) had this coinfection. The clinical indicators related to HBV infection in these 11 patients are shown in Table 3.

### Table 3. Clinical indicators related to hepatitis B virus (HBV) infection among the 11 patients who were positive for only hepatitis B e antigen (HBeAg) and hepatitis B surface antigen (HBsAg).

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender, age (years)</th>
<th>HIV infection history (years)</th>
<th>CD4+ T lymphocyte count (cells/ul)</th>
<th>ART(^a)</th>
<th>ART regimen</th>
<th>ART duration</th>
<th>HBV DNA, IU/ml</th>
<th>Quantitative detection of HBsAg, IU/ml</th>
<th>Quantitative detection of HBeAg, s/co</th>
<th>ALT(^b) (U/L)</th>
<th>AST(^c) (U/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M, 33</td>
<td>10</td>
<td>23</td>
<td>N</td>
<td>N/A(^d)</td>
<td>N/A</td>
<td>4.36×10⁶</td>
<td>&gt;250</td>
<td>752.7</td>
<td>69</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>F, 26</td>
<td>5</td>
<td>280</td>
<td>Y</td>
<td>TDF(^e)/3TC(^f)/EFV(^g)</td>
<td>5 years</td>
<td>&lt;20</td>
<td>&gt;250</td>
<td>438.45</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>M, 39</td>
<td>_h</td>
<td>220</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>&lt;20</td>
<td>1.68</td>
<td>3.54</td>
<td>679</td>
<td>732</td>
</tr>
<tr>
<td>4</td>
<td>M, 39</td>
<td>—</td>
<td>36</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>3.36×10⁷</td>
<td>&gt;250</td>
<td>1840.69</td>
<td>17</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>M, 51</td>
<td>—</td>
<td>57</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>2.33×10⁴</td>
<td>11.09</td>
<td>1.02</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>M, 32</td>
<td>3</td>
<td>2</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>6.57×10⁵</td>
<td>353.1</td>
<td>1109.894</td>
<td>16</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>M, 27</td>
<td>—</td>
<td>38</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>6.08×10²</td>
<td>10.04</td>
<td>1.46</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>M, 34</td>
<td>12</td>
<td>319</td>
<td>Y</td>
<td>AZT(^i)/3TC/Lpv/r</td>
<td>2 years</td>
<td>7.74×10⁵</td>
<td>11121</td>
<td>618.552</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>9</td>
<td>M, 28</td>
<td>—</td>
<td>14</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
<td>6.04×10⁸</td>
<td>&gt;250</td>
<td>1678</td>
<td>17</td>
<td>38</td>
</tr>
<tr>
<td>10</td>
<td>M, 31</td>
<td>—</td>
<td>9</td>
<td>Y</td>
<td>TDF/3TC/EFV</td>
<td>1 month</td>
<td>3.07×10⁵</td>
<td>40751</td>
<td>239.68</td>
<td>23</td>
<td>49</td>
</tr>
<tr>
<td>11</td>
<td>F, 62</td>
<td>14</td>
<td>220</td>
<td>Y</td>
<td>TDF/3TC/Lpv/r</td>
<td>2 years</td>
<td>&lt;20</td>
<td>1837</td>
<td>1558.827</td>
<td>22</td>
<td>22</td>
</tr>
</tbody>
</table>

\(a\) ART: antiretroviral therapy.  
\(b\) ALT: alanine aminotransferase.  
\(c\) AST: aspartate aminotransferase.  
\(d\) N/A: not applicable.  
\(e\) TDF: tenofovir disoproxil fumarate.  
\(f\) 3TC: lamivudine.  
\(g\) EFV: efavirenz.  
\(h\) Not available (These patients did not receive HBsAg testing before but were diagnosed with chronic HBV during the implementation period of this study).  
\(i\) AZT: zidovudine.  
\(j\) Lpv/r: lopinavir/ritonavir.

Characteristics of Syphilis Infection Among HIV Blood-Borne And Sex-Borne Transmission Population

The positive rate of \( T pallidum \) antibodies in the sex-borne transmission group was greater than in the blood-borne transmission group (1227/4666, 26.3% vs 164/1849, 8.87%; \( \chi^2 = 239.499; P < .001 \)). Among those individuals who were positive for \( T pallidum \) antibodies, the proportion of patients who were negative for RPR in the sex-borne transmission group was lower than in the blood-borne transmission group (1127/1227, 61.45% vs 117/164, 71.34%; \( \chi^2 = 6.046; P = .01 \)). Similarly, among those who tested positive for RPR, the proportion of patients without symptoms was lower in the people with sex-borne transmission than in the people with blood-borne transmission (346/473, 73.15% vs 42/47, 89.36%; \( \chi^2 = 5.932; P = .02 \)). It was intriguing to note that high-titer syphilis infection and neurosyphilis only occurred in the people with sex-borne transmission. In people with sex-borne transmission, the proportion of patients with a syphilis titer \( \geq 1:16 \) and neurosyphilis were 8.56% (105/1227) and 7.82% (37/473), respectively, indicating a statistically significant difference between the sex- and blood-borne transmission groups (105/1227, 8.56% vs 0/164, 0%; \( \chi^2 = 15.180; P < .001 \) and 37/473, 7.82% vs 0/47, 0%; \( \chi^2 = 3.958; P = .047 \)). These results are shown in Table 4.
The frequency of HCV exposure or the amount of blood contact is higher in former paid blood donors, contaminated blood recipients, and intravenous drug users. The chronic HCV infection rate was found to be highest in former paid blood donors (80.27%), followed by contaminated blood recipients, and intravenous drug users (73.29%). The difference was related to the outbreak of HIV and HCV infection in the special historical period of this study. Therefore, the prevalence of HIV and HBV coinfection among adults in China was mainly acquired during the preschool period. Therefore, the prevalence of HIV and HBV coinfection in Hubei province shows obvious regional characteristics. This large-scale epidemiological survey used HBV markers (anti-HCV combined with HCV RNA and T. pallidum titer level) to investigate the prevalence of coinfections with HBV, HCV, or syphilis, as well as the different intensities of propagation in individuals with different risk behaviors but the same HIV transmission route.

The recognized transmission routes of HCV are blood transmission, sexual transmission, and MTCT. The 3 main routes of HCV infection demonstrated varying levels of transmission intensity, with blood transmission being the most intense, followed by MTCT and sexual transmission. In order to determine the characteristics of HCV transmission from a blood origin, the populations in this study were further divided into former paid blood donors, contaminated blood recipients, and intravenous drug users. We confirmed that HCV infection was significantly associated with the specific exposure mode of blood contact. The chronic HCV infection rate was found to be highest in former paid blood donors (80.27%), followed by intravenous drug users (73.29%), and contaminated blood recipients (57.14%). The chronic HCV infection rate in all these groups was greater than the prevalence of posttransfusion HCV infection among individuals with β-thalassemia [21] and blood donors in sub-Saharan Africa [22]. The difference was related to the outbreak of HIV and HCV infection in the special historical period of this study. The strong relationship between the frequency of HCV exposure or the amount of blood contact and the risk of HCV infection among former paid blood donors, contaminated blood recipients, and intravenous drug users was validated using population characteristics from a specific historical time. Although this viewpoint is conceptually evident and has long been acknowledged, obtaining specific and objective data is challenging. This study, however, provides evidence on exactly this problem.

Multiple sexual partners, mucosally traumatic intercourse, and bleeding during anal contact have been found to be common among men who have sex with men, which can raise the risk of HCV infection [23]. However, it was found that the HCV infection rate in men who have sex with men was lower than in people who acquired HIV through heterosexual contact (1.24% vs 8.07%). In this study, the male-to-male contact behavior that caused the spread of AIDS in this region was relatively late, and HCV coinfection in this community was relatively infrequent, so it did not result in the broad spread of HCV in men who have sex with men.

The prevalence of HBV infections is affected by multiple factors, including epidemic HBV in the birthplace, rates of HBV vaccination, and risk behaviors for HBV exposure. In the general population, chronic HBV infection was reported as 2.18% in China [24]. A previous study showed that people living with HIV/AIDS were 7 times more likely to have been infected with HBV [25]. However, the overall rate of chronic HBV infection in this study was 10.92%, which was much lower than the above inferred values. As is well known, China is a country with a moderate prevalence of HBV, and the majority of chronic HBV infections occur through vertical MTCT during the preschool period, which is different from European and American countries. Almost 95% of people infected with HBV can completely clear the infection when it is acquired during adulthood, and only 5% to 10% of people infected with HBV will develop a persistent infection. The regional characteristics of HBV prevalence may well account for the result that chronic HBV coinfection was not as high as HCV in this study. Interestingly, we found that men who have sex with men had a higher chronic HBV infection rate but a lower HBV exposure...
rate compared to people living with HIV/AIDS among the heterosexual population, implying that sexual behavior among men who have sex with men was closely related to the increase in the chronic HBV infection rate, providing another reason to strengthen disease safety education among men who have sex with men. Previous research found that men are 1.5 times more vulnerable to HBV than women, implying that gender is a factor influencing persistent HBV infection after HBV exposure [26]. In addition, another key aspect that may account for the greater HBV infection rate in men who have sex with men is exposure to HBV-contaminated blood during mucosally traumatic sex. Combining all these data shows that in places where vertical HBV transmission is a key epidemic component, both the intensity of transmission and the effect of HBV exposure on the prevalence of chronic HBV infection are substantially lower than for HCV.

In this study, among people who acquired HIV by either the blood route or by sexual contact, the combinations of HBsAg/anti-HBe and HBeAg/anti-HBs were detected, which could account for the transient coexisting phases of antigens and antibodies, the overlapping infection of different subtypes of HBV, or the deletion of genes in the S region or pre-S region of HBV. An interesting finding of this study was that an unusual pattern for HBV infection (the combination of only HBsAg/HBeAg antigens) was found only among people who acquired HIV by sexual contact, which might stimulate interest in further exploration of this phenomenon. In a previous study focused on HBV vaccination among people living with HIV/AIDS, it found that anti-HBe titers in a low CD4+ group were much lower than a high CD4+ group (low CD4+ group: 462 mIU/ml vs high CD4+ group: 8834 mIU/ml), which shows that an immunosuppressed state can affect the ability to produce protective anti-HBe [27]. Although immune deficiency is a possible cause, it still could not fully explain this unusual pattern. Further follow-up monitoring in these patients will be considered to explicitly reveal whether HBV-related antibodies appear after implementation of ART. This large-scale epidemiological study based on the comprehensive HBV marker test could provide a clue for further research into the mechanism of this unusual combination of HBV markers.

As a transfusion-transmissible infection, the prevalence of syphilis coinfection among people who acquired HIV from the blood route was 8.87% in this study, which is higher than in reports on the HIV-negative populations in Nigeria (4.2%) [28], Burkina Faso (3.96%) [29], Kenya (4.3%) [30], Pakistan (0.43%) [31], and India (0.43%) [32]. Nonuse of protection in sexual relations has already been identified as a risk factor. Except for differences in sociocultural practices, geographical location, and condom coverage, people living with HIV/AIDS may be more susceptible to T pallidum than HIV-negative people, which is a possible reason for the observed discrepancy in the magnitude of syphilis in the different studies. High T pallidum and neurosyphilis titers were found to be more common in people living with HIV acquired through sexual contact, implying that safe sex education is essential to reduce the risk of late syphilis; providing easily obtained information, education, and communication materials with pictures in HIV and sexually transmitted disease clinics, even in public places, is an important measure.

Overall, there were some highlights in this study. First, the transmission mode of AIDS in Hubei province has an obvious time node, which makes the regional characteristics of this study more prominent. Second, this study used syphilis antibody titers combined with related clinical symptoms and a complete marker test of HBV, in addition to HBsAg, to conduct a large-scale epidemiological investigation on the prevalence of syphilis and HBV infection, making the observational indicators of this study more comprehensive and superior. Third, another highlight is the integration into the study of subject grouping. For example, the people with blood-borne HIV transmission were further subdivided into intravenous drug users, contaminated blood recipients, and former paid blood donors. Similarly, the people with sex-borne HIV transmission were further subdivided into men who have sex with men and heterosexual people. These divisions deepened the understanding of the intensities of propagation in people with varying risk behaviors but the same HIV-borne transmission route.

However, we recognize that our study has some limitations. Given the cross-sectional nature of the study, we are unable to observe changes in HBV markers in those patients who had unusual patterns for HBV (the combination of only HBsAg/HBeAg antigens), but further follow-up observation will be conducted in the future. Also, due to the large sample sizes involved in epidemiological analyses, HCV RNA detection was only performed in patients positive for anti-HCV, which may miss some patients with detectable HCV RNA who are negative for anti-HCV. In addition, HBV DNA screening was only done for HBsAg-positive patients, not for all people living with HIV/AIDS, which could lead to underestimating the HBV infection rate due to the presence of occult HBV infections. Finally, the level of stratified heterogeneity is something that deserves attention and needs improvement. However, given the low frequency of false negative anti-HCV and occult HBV infections in this investigation and the large sample size, the aforesaid limitations should not have influenced the overall trend of the results.

**Conclusion**

HCV, HBV, and syphilis all have similar transmission routes; however, the varied routes do not carry the same weight in terms of disease prevalence. The risk intensity of HCV transmission by blood is substantially higher than that by sexual contact, and it is closely related to exposure frequency. In Hubei province, where vertical transmission of HBV is the primary route, the influence of adult HIV infection and high-risk behavior on HBV transmission is much lower than on HCV, and the range of chronic HBV infection prevalence in different high-risk groups is relatively limited. In people with sex-borne HIV transmission, high T pallidum and neurosyphilis titers are common. The data presented in this study provide a comprehensive picture of the prevalence of HBV, HCV, and syphilis in various populations with HIV risk behaviors, deepening our understanding that repeated exposure to unclean blood or sexual contact increases the risk of HCV and emphasizing the close relationship between neurosyphilis and repeated exposure.
As the distribution of HIV high-risk groups varies by country, so does the burden of HBV, HCV, and syphilis coinfection, and these precise comparative and analytical data serve as the foundation for future local policy creation. It is necessary to invest in preventive and control funds based on the regional distribution of people living with HIV/AIDS in order to optimize the returns.

Acknowledgments
We thank the patients, nurses, and physicians who provided care for the patients. We also thank the investigators at Zhongnan Hospital of Wuhan University. This study was supported by the National Natural Science Foundation of China (82003511) and Discipline and Platform Construction Project of the Center for AIDS Research of Wuhan University (PTPP2023002).

Data Availability
All data generated or used during this study are included in this published article.

Authors’ Contributions
RY, KZ, HH, LL, and LF collected the clinical data. XG and HK interpreted the results. RY and XY drafted the manuscript. YY and ML modified the final manuscript.

Conflicts of Interest
None declared.

References


Abbreviations

- **Anti-HBc**: hepatitis B core antibody
- **Anti-HBe**: hepatitis B e antibody
- **Anti-HBs**: hepatitis B surface antibody
- **ART**: antiretroviral therapy
- **CSF**: cerebrospinal fluid
- **EIA**: enzyme immunoassay
- **ELISA**: enzyme-linked immunosorbent assay
- **HBeAg**: hepatitis B e antigen
- **HBsAg**: hepatitis B surface antigen
- **HBV**: hepatitis B virus
- **HCV**: hepatitis C virus
- **MTCT**: mother-to-child transmission
- **RPR**: rapid plasma reagin circle card test
- **WHO**: World Health Organization

© Rongrong Yang, Rui Yuan, Xien Gui, Hengning Ke, Ke Zhuang, Hui Hu, Ling Li, Ling Feng, Xingxia Yu, Yajun Yan, Mingqi Luo. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 27.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
SARS-CoV-2 Infection, Hospitalization, and Associated Factors Among People Living With HIV in Southeastern China From December 2022 to February 2023: Cross-Sectional Survey

Wei Cheng¹, MPH; Yun Xu¹, MPH; Haibo Jiang², MPH; Jun Li³, MPH; Zhigang Hou⁴, MPH; Haibin Meng⁵, MPH; Wei Wang⁶, MPH; Chengliang Chai¹, MPH; Jianmin Jiang¹, PhD

¹Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
²Ningbo Center for Disease Control and Prevention, Ningbo, China
³Wenzhou Center for Disease Control and Prevention, Wenzhou, China
⁴Jiaxing Center for Disease Control and Prevention, Jiaxing, China
⁵Shaoxing Center for Disease Control and Prevention, Shaoxing, China
⁶Quzhou Center for Disease Control and Prevention, Quzhou, China

Abstract

Background: Limited studies have explored the impact of the Omicron variant on SARS-CoV-2 infection, hospitalization, and associated factors among people living with HIV, particularly in China. The adjustment of preventive policies since December 2022 in China presents an opportunity to evaluate the real-world factors influencing SARS-CoV-2 infection and related hospitalization among people living with HIV.

Objective: This study aimed to investigate SARS-CoV-2 infection, hospitalization rates, and associated factors among people living with HIV following the adjustment of preventive policies from December 2022 to February 2023 in southeastern China.

Methods: A cross-sectional telephone or web-based survey was conducted among people living with HIV in 5 cities in southeastern China from December 2022 to February 2023. Demographic information, SARS-CoV-2 infection and related hospitalization, and HIV-specific characteristics were collected from existing databases and special investigations. Multivariate logistic regression analyses were conducted to determine the associated factors for infection and hospitalization rates of SARS-CoV-2. Additionally, subgroup analyses were conducted for the association between vaccination and infection across different vaccination statuses and time since the last vaccination.

Results: Among people living with HIV with a COVID-19 testing history, the SARS-CoV-2 infection rate was 67.13% (95% CI 65.81%-68.13%), whereas the hospitalization rate was 0.71% (95% CI 0.46%-0.97%). Factors such as age, latest CD4 cell count, latest HIV viral load, and transmission route were found to be associated with SARS-CoV-2 infection, while age, cancer, latest CD4 cell count, and latest HIV viral load were associated with SARS-CoV-2 hospitalization. In terms of SARS-CoV-2 vaccination, compared to unvaccinated people living with HIV, there was a lower infection rate among those who had been vaccinated for <3 months in the booster vaccination group (adjusted odds ratio [aOR] 0.72, 95% CI 0.53-0.98; P=.04); and there was also a lower risk of hospitalization among people living with HIV who had received vaccination in the past 6-12 months (aOR 0.33, 95% CI 0.14-0.81; P=.02) and more than 12 months ago (aOR 0.22, 95% CI 0.07-0.72; P=.01).

Conclusions: After the ease of prevention and control measures in China, we observed a high SARS-CoV-2 infection rate but a low hospitalization rate. General risk factors, such as higher age and vaccination status, and HIV-related parameters, such as the latest CD4 cell count and HIV viral load, were associated with SARS-CoV-2 infection and hospitalization. A booster vaccination campaign for booster doses should be considered among people living with HIV in confronting possible COVID-19 epidemic emergencies in the near future.
Introduction

HIV/AIDS has remained a significant global public health threat despite being discovered over 40 years ago. Currently, there are 38 million people living with HIV worldwide, and in 2021 alone, 650,000 individuals died from HIV-related illnesses [1]. The situation regarding global HIV epidemic control has become more severe due to the emergence of the COVID-19 pandemic caused by SARS-CoV-2 [2,3]. It can be understood that the weakening of the immune system of patients with HIV/AIDS increases the possibility of severe outcomes from SARS-CoV-2 coinfection. However, there is limited data on the susceptibility and vaccine effectiveness of people living with HIV to SARS-CoV-2 infection and severe outcomes. As there is no cure for HIV and there is a chance of SARS-CoV-2 infection, coinfection continues to pose a problem [3-5].

Previous studies have suggested that older age and comorbidities such as hypertension, diabetes, and cardiovascular disease are risk factors for developing severe COVID-19 and are associated with high mortality rates [6]. However, the interaction of people living with HIV-specific factors, such as CD4 cell count, HIV viral load, and antiretroviral treatment (ART), with COVID-19 remains inclusive [3,6-8]. Moreover, most of these studies were conducted before the vaccination or during the epidemics of the Alpha and Delta variants. Limited studies have explored the impact of the Omicron variant, which exhibits increased transmissibility and immune escape abilities, on people living with HIV [9]. Furthermore, because of the strict prevention and control policies for COVID-19 in China, few studies concerning the SARS-CoV-2 infection and hospitalization among people living with HIV have been reported in the country.

Due to a unified, tightly coordinated COVID-19 vaccination campaign during the dynamic COVID-19 Zero period, partial and full vaccination coverage in China reached 91.5% and 89.3%, respectively [10,11]. In the context of this high vaccination coverage and the relatively weak virulence of the SARS-CoV-2 Omicron variant, the Chinese government relaxed COVID-19 prevention and control measures in December 2022. Notably, several studies reported a significant surge in cases between December 2022 and January 2023, with reported infection rates ranging from 40.9% to 92.3% in the general population [11-15]. However, there are currently no reports regarding SARS-CoV-2 infection and hospitalization among people living with HIV during this period. Moreover, although a few studies demonstrate the short-duration protection of the COVID-19 vaccine against infection in the general population [11,12,16,17], there was no study on the impact of the COVID-19 vaccine on infection and hospitalization of SARS-CoV-2 among people living with HIV in China. The substantial epidemic presents an opportunity to evaluate the real-world factors influencing the infection and hospitalization of SARS-CoV-2 among people living with HIV. Therefore, this study investigated the infection and hospitalization of SARS-CoV-2 and identified associated factors among people living with HIV from December 2022 to January 2023.

Methods

Study Population

Zhejiang Province is located in southeastern China and is known for its thriving economy and high population density. This province is home to over 65 million local residents spread across 11 metropolitan areas.

By the end of 2022, the total number of people living with HIV living in Zhejiang Province was 39,744. For our study, we selected 5 metropolitan areas in Zhejiang province as our study sites: Ningbo in the east, Quzhou in the west, Jiaxing in the north, Wenzhou in the south, and Shaoxing in the middle of the province (Figure 1). The total number of people living with HIV in the 5 cities was unbalanced, with Ningbo the highest (6096) and Quzhou the smallest (1346). As a result, the 5 selected cities accounted for 46% of all living cases of people living with HIV in the entire province.
Our target sample size was 10,000, which was estimated by experts. Center for Disease Control and Prevention (CDC) staff from the 5 cities downloaded the database from the HIV/AIDS case reporting information system, established by the Chinese CDC. Considering the work foundation of each city, we randomly selected 50% (5933/11,865) of people living with HIV from Wenzhou and Ningbo and 80% (5102/6378) from Quzhou, Jiaxing, and Shaoxing as potential participants (total number=11,035) to be investigated. The inclusion criteria for people living with HIV in the final analysis were as follows: (1) be aged ≥15 years; (2) currently live in the abovementioned 5 cities; and (3) agree to participate in the survey. Exclusion criteria mainly included people living with HIV who were not currently taking ART or who did not have results of CD4 cell count and HIV viral load.

Data Collection
To collect information regarding SARS-CoV-2 infection, admission status, and related symptoms, a questionnaire was designed by the Zhejiang Provincial CDC. The investigation was conducted by trained county CDC staff between January 16 and February 10, 2023. Uniform training was organized by the Zhejiang CDC and provided to investigators to ensure the quality of the investigation. The method of investigation was determined by the investigators. Ningbo staff conducted a web-based survey using Questionnaire Star (one of the most widely used commercial web-based survey platforms in China). Staff from 4 other cities conducted an investigation by telephone. During the investigation, we first inquired whether the participants had experienced any COVID-19–related symptoms since December 7, 2022. Subsequently, we asked if they had tested positive for the SARS-CoV-2 infection. If they responded with a “yes” or “no,” we further asked for the method of confirmation, specifically whether it was through an antigen test or nucleic acid test. For those who self-reported having SARS-CoV-2 infection, we asked for the confirmation date and whether they had been hospitalized after the infection. We further investigated the reasons for hospitalization for individuals admitted to the hospital.

Demographic information such as age, sex, and marital status was collected from the HIV/AIDS case reporting information system. Information related to ART, such as the initiation date, the latest CD4 cell count, viral load, and treatment regime, was obtained from the national HIV treatment subdatabase of Zhejiang province. Regarding the ART regime, we focused on the national free drugs that are widely used among people living with HIV in China, such as tenofovir (TDF), azidothymidine.
COVID-19 vaccine–related data, including vaccine type, vaccine dose, and last vaccination date, were obtained from the vaccination information system. In terms of vaccine brand and manufacturer, an inactivated vaccine was produced by Sinopharm CNBG Beijing, Sinopharm CNBG Wuhan and the Beijing-based Sinovac Biotech; a recombinant protein (Chinese hamster ovary [CHO] cell) vaccine was produced by Anhui Zhifei Longcom Biopharmaceutical Institute of Microbiology; and an adenovirus vector vaccine was used, referred to as Cansino Ad5-nCoV-S COVID-19 vaccine.

**Definition**

Due to the absence of COVID-19 test results for all participants at the time of investigation, we categorized the study participants into groups A, B, and C. Group A consisted of individuals with no history of COVID-19 testing but reported COVID-19–related symptoms. Among these participants, those who presented with symptoms such as fever or cough were classified as having SARS-CoV-2 infection, while those who did not exhibit such symptoms were categorized as noninfected. Group B comprised individuals who had undergone COVID-19 testing between December 7, 2022, and the investigation date. Infection status was determined based on the positive or negative results of nucleic acid or antigen tests. Group C encompassed all study participants, including groups A and B.

Hospitalizations that were unlikely to be attributed to COVID-19, such as cases involving non–COVID-19 infections or long-term hospitalizations, were excluded from the study.

According to the Technical Vaccination Recommendations for COVID-19 Vaccines in China and previous studies [16-18], vaccination status was categorized into 4 levels: (1) unvaccinated, which indicates no history of COVID-19 vaccination before the last SARS-CoV-2 exposure date; (2) partially vaccinated, which includes people who have received 1 dose of inactivated vaccine, received 2 doses of inactivated vaccine but received the second dose within 14 days before the last SARS-CoV-2 exposure date, received 2 doses of recombinant protein vaccine (3 doses are recommended for full primary vaccination), and received 1 dose of adenovirus vector vaccine, or 3 doses of recombinant protein vaccine but with the last dose within 14 days before the last SARS-CoV-2 exposure date; (3) full primary vaccination, which includes people who received either 2 doses of inactivated vaccine, 1 dose of adenovirus vector vaccine, or 3 doses of recombinant protein (CHO cell) vaccine with the last vaccination 14 days or more before the last SARS-CoV-2 exposure date and with no booster dose; 2 doses of inactivated vaccine with 1 booster dose of inactivated vaccine, adenovirus vector vaccine, or recombinant protein (CHO cell) vaccine within 7 days before the last SARS-CoV-2 exposure date; or 2 doses of adenovirus vector vaccine within 7 days before the last exposure date; (4) booster vaccinated, which includes people who received either 2 doses of Ad5-vector vaccine with the second dose 7 days or more before the last exposure date; or 2 doses of inactivated vaccine and 1 or 2 booster doses of inactivated vaccine, adenovirus vector vaccine, or recombinant protein vaccine 7 days or more before the last exposure date.

The time since the last vaccination was calculated as the interval between the date of the last vaccination dose and the date of the last SARS-CoV-2 exposure. The date of last SARS-CoV-2 exposure was defined as the date of the investigation if the participants had not self-reported SARS-CoV-2 infection; otherwise, it was the date of the positive result of self-tested SARS-CoV-2.

**Statistical Analysis**

The means (SDs) or medians (IQRs) were obtained for continuous variables. Categorical variables were analyzed by calculating characteristic percentages. To analyze the associated factors for SARS-CoV-2 infection, logistic regression analysis was performed among those with a history of SARS-CoV-2 test (group B). To identify risk factors for hospitalization, a multivariable logistic regression analysis was conducted among participants who had undergone SARS-CoV-2 testing and received a positive result. Univariate factors with a significance level of P < .10 and factors previously identified to be associated with COVID-19 infection were included in the full multivariable regression models. Additionally, to make the results more robust, we conducted a sensitivity analysis for the association between SARS-CoV-2 infection rate and vaccination among groups A and C. Data cleaning and organization were performed using Microsoft Excel 2016 software (Microsoft Corporation). Statistical analysis was conducted using R statistical software (version 4.1.2, R Project for Statistical Computing). A significance level of P < .05 was considered statistically significant in the final model.

**Ethics Statement**

The study was approved by the ethical review board of the Zhejiang Provincial CDC (IRB approval number: 2023-003-01). No risk was involved in participating in this study, and we protected the confidentiality of the participants. For those participants aged 18 years or younger, we sought parental consent and collected data from their parents.

**Results**

**Characteristics of the Study Participants**

In total, 9922 people living with HIV (group C) were included in the final analyses, with 3628 classified as group A and 6294 as group B. The response rate was 89.91% (9922/11,035) for the potential participants and 24.96% (9922/39,744) for all people living with HIV in the province. The median age of the participants was 46 years (IQR 22), of which 7996 (80.59%) were male. Among them, 6848 (69.02%) were registered residents of Zhejiang Province. The prevalence of chronic diseases among the participants was as follows: hypertension (930/9992, 9.37%), diabetes 395/9922, 3.98%), and cardiovascular diseases (103/9922, 1.04%). Heterosexual transmission accounted for a higher percentage (5825/9922, 58.71%) compared to homosexual transmission (4001/9922, 40.32%). The median latest CD4 count was 462 cells/µL (IQR 301), and the median ART duration was 5.38 years (IQR 4.65). Among the people living with HIV, 9274 (93.47%) had an HIV (AZT), lamivudine (3TC), efavirenz (EFV), and lopinavir and ritonavir (LPV/r).
viral load below 50 copies/mL. Regarding vaccination status, 1530 (15.42%) were unvaccinated, 455 (4.59%) had received their last vaccination less than 3 months ago, 237 (2.39%) had received their last vaccination between 3 and 5 months ago, 4707 (47.44%) had received their last vaccination between 6 and 11 months ago, and 2993 (30.17%) had received their last vaccination more than 12 months ago (Table 1). Among the entire study population, the proportion of individuals with partial vaccination, full primary vaccination, and booster vaccination was 2.59%, 15.32%, and 66.67%, respectively (Table 1).
Table 1. Characteristics of the study participants among people living with HIV in southeastern China from December 2022 to February 2023.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group A (n=3628)</th>
<th>Group B (n=6294)</th>
<th>Group C (n=9922)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), median (IQR)</strong></td>
<td>49 (22)</td>
<td>44 (23)</td>
<td>46 (22)</td>
</tr>
<tr>
<td><strong>Age categories (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-29</td>
<td>362 (9.98)</td>
<td>955 (15.17)</td>
<td>1317 (13.27)</td>
</tr>
<tr>
<td>30-39</td>
<td>720 (19.85)</td>
<td>1671 (26.55)</td>
<td>2391 (24.1)</td>
</tr>
<tr>
<td>40-49</td>
<td>807 (22.24)</td>
<td>1254 (19.92)</td>
<td>2061 (20.77)</td>
</tr>
<tr>
<td>50-59</td>
<td>923 (25.44)</td>
<td>1397 (22.2)</td>
<td>2320 (23.38)</td>
</tr>
<tr>
<td>≥60</td>
<td>816 (22.49)</td>
<td>1017 (16.16)</td>
<td>1833 (18.47)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2813 (77.54)</td>
<td>5183 (82.35)</td>
<td>7996 (80.59)</td>
</tr>
<tr>
<td>Female</td>
<td>815 (22.46)</td>
<td>1111 (17.65)</td>
<td>1926 (19.41)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illiteracy</td>
<td>208 (5.73)</td>
<td>227 (3.61)</td>
<td>435 (4.38)</td>
</tr>
<tr>
<td>Primary school</td>
<td>822 (22.66)</td>
<td>1065 (16.92)</td>
<td>1887 (19.02)</td>
</tr>
<tr>
<td>Middle school</td>
<td>1462 (40.3)</td>
<td>2195 (34.87)</td>
<td>3657 (36.86)</td>
</tr>
<tr>
<td>High school or technical secondary school</td>
<td>633 (17.45)</td>
<td>1356 (21.54)</td>
<td>1989 (20.05)</td>
</tr>
<tr>
<td>College degree or above</td>
<td>503 (13.86)</td>
<td>1451 (23.05)</td>
<td>1954 (19.69)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single or divorced</td>
<td>1701 (46.89)</td>
<td>3589 (57.02)</td>
<td>5290 (53.32)</td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td>1922 (52.98)</td>
<td>2695 (42.82)</td>
<td>4617 (46.53)</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (0.14)</td>
<td>10 (0.16)</td>
<td>15 (0.15)</td>
</tr>
<tr>
<td><strong>Household registration, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhejiang province</td>
<td>2494 (68.74)</td>
<td>4354 (69.18)</td>
<td>6848 (69.02)</td>
</tr>
<tr>
<td>Other provinces</td>
<td>1134 (31.26)</td>
<td>1940 (30.82)</td>
<td>3074 (30.98)</td>
</tr>
<tr>
<td><strong>Hypertension, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>363 (10.01)</td>
<td>567 (9.01)</td>
<td>930 (9.37)</td>
</tr>
<tr>
<td>No</td>
<td>3265 (89.99)</td>
<td>5727 (90.99)</td>
<td>8992 (90.63)</td>
</tr>
<tr>
<td><strong>Diabetes, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>133 (3.67)</td>
<td>262 (4.16)</td>
<td>395 (3.98)</td>
</tr>
<tr>
<td>No</td>
<td>3495 (96.33)</td>
<td>6032 (95.84)</td>
<td>9527 (96.02)</td>
</tr>
<tr>
<td><strong>Chronic obstructive pneumonia, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>12 (0.33)</td>
<td>19 (0.3)</td>
<td>31 (0.31)</td>
</tr>
<tr>
<td>No</td>
<td>3616 (99.67)</td>
<td>6275 (99.7)</td>
<td>9891 (99.69)</td>
</tr>
<tr>
<td><strong>Cardiovascular diseases, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>40 (1.1)</td>
<td>63 (1)</td>
<td>103 (1.04)</td>
</tr>
<tr>
<td>No</td>
<td>3588 (98.9)</td>
<td>6231 (99)</td>
<td>9819 (98.96)</td>
</tr>
<tr>
<td><strong>Chronic kidney disease, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20 (0.55)</td>
<td>27 (0.43)</td>
<td>47 (0.47)</td>
</tr>
<tr>
<td>No</td>
<td>3608 (99.45)</td>
<td>6267 (99.57)</td>
<td>9875 (99.53)</td>
</tr>
<tr>
<td><strong>Cancer, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18 (0.5)</td>
<td>38 (0.6)</td>
<td>56 (0.56)</td>
</tr>
<tr>
<td>No</td>
<td>3610 (99.5)</td>
<td>6256 (99.4)</td>
<td>9866 (99.44)</td>
</tr>
<tr>
<td><strong>Cerebrovascular disease, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Group A (n=3628)</td>
<td>Group B (n=6294)</td>
<td>Group C (n=9922)</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Yes</td>
<td>15 (0.41)</td>
<td>27 (0.43)</td>
<td>42 (0.42)</td>
</tr>
<tr>
<td>No</td>
<td>3613 (99.59)</td>
<td>6267 (99.57)</td>
<td>9880 (99.58)</td>
</tr>
<tr>
<td>Transmission route, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>2322 (64)</td>
<td>3503 (55.66)</td>
<td>5825 (58.71)</td>
</tr>
<tr>
<td>Homosexual</td>
<td>1268 (34.95)</td>
<td>2733 (43.42)</td>
<td>4001 (40.32)</td>
</tr>
<tr>
<td>Other or unknown</td>
<td>38 (1.05)</td>
<td>58 (0.92)</td>
<td>96 (0.97)</td>
</tr>
<tr>
<td>Latest CD4 count (cells/μL), median (IQR)</td>
<td>456 (303)</td>
<td>466 (299)</td>
<td>462 (301)</td>
</tr>
<tr>
<td>Latest CD4 count (cells/μL), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-199</td>
<td>337 (9.29)</td>
<td>514 (8.17)</td>
<td>851 (8.58)</td>
</tr>
<tr>
<td>200-499</td>
<td>1765 (48.65)</td>
<td>3020 (47.98)</td>
<td>4785 (48.23)</td>
</tr>
<tr>
<td>≥500</td>
<td>1526 (42.06)</td>
<td>2760 (43.85)</td>
<td>4286 (43.2)</td>
</tr>
<tr>
<td>Latest HIV load (copies/mL), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-49</td>
<td>3374 (93)</td>
<td>5900 (93.74)</td>
<td>9274 (93.47)</td>
</tr>
<tr>
<td>50-999</td>
<td>115 (3.17)</td>
<td>170 (2.7)</td>
<td>285 (2.87)</td>
</tr>
<tr>
<td>≥1000</td>
<td>139 (3.83)</td>
<td>224 (3.56)</td>
<td>363 (3.66)</td>
</tr>
<tr>
<td>Time since ART initiation (years), median (IQR)</td>
<td>5.35 (4.65)</td>
<td>5.35 (4.66)</td>
<td>4.65 (4.65)</td>
</tr>
<tr>
<td>TDF&lt;sup&gt;b&lt;/sup&gt; use, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1281 (35.31)</td>
<td>2537 (40.31)</td>
<td>3818 (38.48)</td>
</tr>
<tr>
<td>Yes</td>
<td>2347 (64.69)</td>
<td>3757 (59.69)</td>
<td>6104 (61.52)</td>
</tr>
<tr>
<td>3TC&lt;sup&gt;c&lt;/sup&gt; use, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>363 (10.01)</td>
<td>938 (14.9)</td>
<td>1301 (13.11)</td>
</tr>
<tr>
<td>Yes</td>
<td>3265 (89.99)</td>
<td>5356 (85.1)</td>
<td>8621 (86.89)</td>
</tr>
<tr>
<td>EFV&lt;sup&gt;d&lt;/sup&gt; use, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1072 (29.55)</td>
<td>2202 (34.99)</td>
<td>3274 (33)</td>
</tr>
<tr>
<td>Yes</td>
<td>2556 (70.45)</td>
<td>4092 (65.01)</td>
<td>6648 (67)</td>
</tr>
<tr>
<td>AZT&lt;sup&gt;e&lt;/sup&gt; use, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2804 (77.29)</td>
<td>4935 (78.41)</td>
<td>7739 (78)</td>
</tr>
<tr>
<td>Yes</td>
<td>824 (22.71)</td>
<td>1359 (21.59)</td>
<td>2183 (22)</td>
</tr>
<tr>
<td>LPV/r&lt;sup&gt;f&lt;/sup&gt; use, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3311 (91.26)</td>
<td>5812 (92.34)</td>
<td>9123 (91.95)</td>
</tr>
<tr>
<td>Yes</td>
<td>317 (8.74)</td>
<td>482 (7.66)</td>
<td>799 (8.05)</td>
</tr>
<tr>
<td>Time since last vaccination (months), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No vaccination</td>
<td>539 (14.86)</td>
<td>991 (15.75)</td>
<td>1530 (15.42)</td>
</tr>
<tr>
<td>0-2</td>
<td>170 (4.69)</td>
<td>285 (4.53)</td>
<td>455 (4.59)</td>
</tr>
<tr>
<td>3-5</td>
<td>96 (2.65)</td>
<td>141 (2.24)</td>
<td>237 (2.39)</td>
</tr>
<tr>
<td>6-11</td>
<td>1734 (47.79)</td>
<td>2973 (47.24)</td>
<td>4707 (47.44)</td>
</tr>
</tbody>
</table>
Among people living with HIV with a COVID-19 testing history, the SARS-CoV-2 infection rate was 66.97% (95% CI 65.81%-68.13%). The SARS-CoV-2 infection rate was significantly higher in the younger age group, those with single or divorced status, those with homosexual transmission routes, those with higher education levels, those with higher latest CD4 counts, those with lower latest HIV viral load, those not taking TDF, 3TC, and EFV, and those taking AZT, and those who had a long time since their last vaccination (Table 2). Multivariate logistic regression analysis revealed that age, education, transmission route, latest CD4 count, and latest HIV viral load were statistically associated with SARS-CoV-2 infection. Regarding vaccination, compared to those who had not received any vaccination, there was no statistically significant difference among those who received vaccination 4–6 months and 7-12 months since their last dose. However, there was a statistically significant association among those who had been vaccinated for 0-3 months (aOR 0.74, 95% CI 0.56-0.98) and more than 12 months (aOR 1.20, 95% CI 1.01-1.43; Table 2).
<table>
<thead>
<tr>
<th>Variable</th>
<th>Infection People living with HIV, n (%)</th>
<th>P value</th>
<th>Multivariable analysis aOR^a (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age categories (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-29</td>
<td>754 (78.95)</td>
<td>.001</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>30-39</td>
<td>1272 (76.12)</td>
<td>.02</td>
<td>0.78 (0.64-0.96)</td>
<td>.02</td>
</tr>
<tr>
<td>40-49</td>
<td>825 (65.79)</td>
<td>.001</td>
<td>0.51 (0.41-0.63)</td>
<td>.001</td>
</tr>
<tr>
<td>50-59</td>
<td>865 (61.92)</td>
<td>.001</td>
<td>0.48 (0.38-0.59)</td>
<td>.001</td>
</tr>
<tr>
<td>≥60</td>
<td>499 (49.07)</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3475 (67.05)</td>
<td>.001</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Female</td>
<td>740 (66.61)</td>
<td>.001</td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td>.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College degree or above</td>
<td>112 (49.34)</td>
<td>.01</td>
<td>0.66 (0.48-0.91)</td>
<td>.01</td>
</tr>
<tr>
<td>Illiteracy</td>
<td>571 (53.62)</td>
<td>.01</td>
<td>0.68 (0.55-0.84)</td>
<td>.01</td>
</tr>
<tr>
<td>Primary school</td>
<td>1434 (65.33)</td>
<td>.03</td>
<td>0.82 (0.70-0.98)</td>
<td>.03</td>
</tr>
<tr>
<td>Middle school</td>
<td>982 (72.42)</td>
<td>.49</td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td>High school or technical secondary school</td>
<td>1116 (76.91)</td>
<td>.49</td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td>.001</td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Single or divorced</td>
<td>2511 (69.96)</td>
<td>.001</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td>1696 (62.93)</td>
<td>.001</td>
<td>1.23 (1.08-1.40)</td>
<td>.001</td>
</tr>
<tr>
<td>Unknown</td>
<td>8 (80)</td>
<td>.001</td>
<td>2.33 (0.47-11.6)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Household registration</strong></td>
<td></td>
<td>.19</td>
<td></td>
<td>.19</td>
</tr>
<tr>
<td>Zhejiang province</td>
<td>2893 (66.44)</td>
<td>.19</td>
<td></td>
<td>.19</td>
</tr>
<tr>
<td>Other provinces</td>
<td>1322 (68.14)</td>
<td>.19</td>
<td></td>
<td>.19</td>
</tr>
<tr>
<td><strong>Hypertension</strong></td>
<td></td>
<td>.14</td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td>No</td>
<td>364 (64.20)</td>
<td>.14</td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td>Yes</td>
<td>3851 (67.24)</td>
<td>.14</td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td><strong>Diabetes</strong></td>
<td></td>
<td>.12</td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>No</td>
<td>164 (62.60)</td>
<td>.12</td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>Yes</td>
<td>4051 (67.16)</td>
<td>.12</td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td><strong>Chronic obstructive pneumonia</strong></td>
<td></td>
<td>.11</td>
<td></td>
<td>.11</td>
</tr>
<tr>
<td>No</td>
<td>16 (84.21)</td>
<td>.11</td>
<td></td>
<td>.11</td>
</tr>
<tr>
<td>Yes</td>
<td>4199 (66.92)</td>
<td>.11</td>
<td></td>
<td>.11</td>
</tr>
<tr>
<td><strong>Cardiovascular diseases</strong></td>
<td></td>
<td>.12</td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>No</td>
<td>48 (76.19)</td>
<td>.12</td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>Yes</td>
<td>4167 (66.88)</td>
<td>.12</td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td><strong>Chronic kidney disease</strong></td>
<td></td>
<td>.23</td>
<td></td>
<td>.23</td>
</tr>
<tr>
<td>Yes</td>
<td>21 (77.78)</td>
<td>.23</td>
<td></td>
<td>.23</td>
</tr>
<tr>
<td>No</td>
<td>4194 (66.92)</td>
<td>.23</td>
<td></td>
<td>.23</td>
</tr>
<tr>
<td><strong>Cancer</strong></td>
<td></td>
<td>.12</td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>Yes</td>
<td>30 (78.95)</td>
<td>.12</td>
<td></td>
<td>.12</td>
</tr>
<tr>
<td>Variable</td>
<td>People living with HIV, n (%)</td>
<td>$P$ value</td>
<td>Multivariable analysis</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------</td>
<td>-----------</td>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$P$ value</td>
<td>$\text{aOR}^a$ (95% CI)</td>
<td>$P$ value</td>
</tr>
<tr>
<td>No</td>
<td>4185 (66.90)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td></td>
<td>.97</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>18 (66.67)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>4197 (66.97)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Transmission</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>2142 (61.15)</td>
<td>Reference</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Homosexual</td>
<td>2033 (74.39)</td>
<td>1.30 (1.15-1.48)</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Other or unknown</td>
<td>40 (68.97)</td>
<td>1.30 (0.73-2.31)</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>Latest CD4 count (cells/μL)</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>0-199</td>
<td>282 (54.86)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>200-499</td>
<td>2034 (67.55)</td>
<td>1.35 (1.10-1.65)</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>≥500</td>
<td>1899 (68.80)</td>
<td>1.22 (0.99-1.50)</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Latest HIV load (copies/mL)</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>0-49</td>
<td>3984 (67.53)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>50-999</td>
<td>107 (62.94)</td>
<td>0.90 (0.65-1.26)</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>≥1000</td>
<td>124 (55.36)</td>
<td>0.68 (0.51-0.92)</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Time since ART$^c$ initiation (years)</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>0</td>
<td>156 (65.82)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>1</td>
<td>407 (61.39)</td>
<td>0.85 (0.61-1.19)</td>
<td>.35</td>
<td></td>
</tr>
<tr>
<td>2-4</td>
<td>1325 (65.63)</td>
<td>1.08 (0.79-1.48)</td>
<td>.61</td>
<td></td>
</tr>
<tr>
<td>≥5</td>
<td>2327 (68.95)</td>
<td>1.27 (0.93-1.73)</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>TDF$^d$ use</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>1825 (71.94)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>2390 (63.61)</td>
<td>0.99 (0.76-1.29)</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>3TC$^e$ use</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>718 (76.55)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>3497 (65.29)</td>
<td>0.60 (0.45-0.79)</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>EFV$^f$ use</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>1535 (69.71)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>2680 (65.49)</td>
<td>1.06 (0.91-1.23)</td>
<td>.45</td>
<td></td>
</tr>
<tr>
<td>AZT$^g$ use</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>3256 (65.98)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>959 (70.57)</td>
<td>1.23 (0.93-1.62)</td>
<td>.141</td>
<td></td>
</tr>
<tr>
<td>LPV/r$^h$ use</td>
<td></td>
<td>.32</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>3902 (67.14)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>313 (64.94)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Time since last vaccination (months)</td>
<td></td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No vaccination</td>
<td>633 (63.87)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>0-2</td>
<td>159 (55.79)</td>
<td>0.74 (0.56-0.98)</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>3-5</td>
<td>81 (57.45)</td>
<td>0.80 (0.55-1.17)</td>
<td>.25</td>
<td></td>
</tr>
</tbody>
</table>
Among the 4215 confirmed COVID-19 cases, 30 individuals (0.71%, 95% CI 0.46%-0.97%) required hospitalization. The highest hospitalized rates were observed among individuals aged ≥60 years, those with illiteracy, married individuals, individuals with hypertension, diabetes, chronic kidney disease, or cancer, individuals with the latest CD4 count below 200

The tests for COVID-19 in Zhejiang Province from December 2022 to February 2023

Among the 4215 confirmed COVID-19 cases, 30 individuals (0.71%, 95% CI 0.46%-0.97%) required hospitalization. The highest hospitalized rates were observed among individuals aged ≥60 years, those with illiteracy, married individuals, individuals with hypertension, diabetes, chronic kidney disease, or cancer, individuals with the latest CD4 count below 200.
cells/μL, an HIV load greater than 1000 copies/mL, and those who had not received any vaccination (Table 3).

In the multivariate analysis, several factors, including age, cancer, latest CD4 count, latest HIV viral load, and time since the last vaccination, remained significantly associated with COVID-19 hospitalization. Compared to people living with HIV aged <30 years, those aged >60 years showed an increased risk of hospital admission (aOR 7.44, 95% CI 1.23-45.17). Hospital admission was also associated with having cancer (aOR 12.19, 95% CI 2.98-49.93). Furthermore, compared to individuals who had not received any vaccination, people living with HIV who had received vaccination in the past 6-12 months (aOR 0.33, 95% CI 0.14-0.81) and those who had received vaccination more than 12 months ago (aOR 0.22, 95% CI 0.07-0.72) had a lower risk of hospitalization (Table 3).

Sensitivity analyses were conducted among groups A and C to evaluate the robustness of our results. The SARS-CoV-2 infection rate was 67.03% (95% CI 65.50%-68.56%) and 66.99% (95% CI 66.07%-67.92%) in groups A and C, respectively. Moreover, as shown in Figures 3 and 4, there was a statistically significant association between SARS-CoV-2 infection and vaccination for those who were vaccinated <3 months in group A (aOR 0.65, 95% CI 0.45-0.93; Figure 3) and group C (aOR 0.72, 95% CI 0.58-0.90; Figure 4). Subgroup analysis showed that the association between vaccination and SARS-CoV-2 infection occurred at the booster vaccination while not in the full primary group (Figures 3 and 4).
Table 3. SARS-CoV-2 hospitalization and associated factors analysis among people living with HIV who had positive tests of COVID-19 in Zhejiang province from December 2022 to February 2023.

<table>
<thead>
<tr>
<th>Variable</th>
<th>COVID-19 cases (n=4215), n (%)</th>
<th>Hospitalization</th>
<th>Multivariable analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>People living with HIV, n (%)</td>
<td>People living with HIV, n (%)</td>
<td>$P$ value</td>
</tr>
<tr>
<td>Age categories (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-29</td>
<td>754 (17.89)</td>
<td>2 (0.27)</td>
<td>.001</td>
</tr>
<tr>
<td>30-39</td>
<td>1272 (30.18)</td>
<td>1 (0.08)</td>
<td>.29</td>
</tr>
<tr>
<td>40-49</td>
<td>825 (19.57)</td>
<td>2 (0.24)</td>
<td>.45</td>
</tr>
<tr>
<td>50-59</td>
<td>865 (20.52)</td>
<td>6 (0.69)</td>
<td>.85</td>
</tr>
<tr>
<td>≥60</td>
<td>499 (11.84)</td>
<td>19 (3.81)</td>
<td>.03</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Male</td>
<td>3475 (82.44)</td>
<td>21 (0.6)</td>
<td>Reference</td>
</tr>
<tr>
<td>Female</td>
<td>740 (17.56)</td>
<td>9 (1.22)</td>
<td>1.39 (0.55-3.55)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>College degree or above</td>
<td>112 (2.66)</td>
<td>5 (4.46)</td>
<td>Reference</td>
</tr>
<tr>
<td>Illiteracy</td>
<td>571 (13.55)</td>
<td>8 (1.4)</td>
<td>2.96 (0.43-20.3)</td>
</tr>
<tr>
<td>Primary school</td>
<td>1434 (34.02)</td>
<td>9 (0.63)</td>
<td>0.83 (0.14-5.05)</td>
</tr>
<tr>
<td>Middle school</td>
<td>982 (23.3)</td>
<td>6 (0.61)</td>
<td>1.04 (0.19-5.80)</td>
</tr>
<tr>
<td>High school or technical</td>
<td>1116 (26.48)</td>
<td>2 (0.18)</td>
<td>1.72 (0.30-9.84)</td>
</tr>
<tr>
<td>secondary school</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Single or divorced</td>
<td>2519 (59.76)</td>
<td>7 (0.28)</td>
<td>Reference</td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td>1696 (40.24)</td>
<td>23 (1.36)</td>
<td>0.45 (0.17-1.20)</td>
</tr>
<tr>
<td>Household registration</td>
<td></td>
<td></td>
<td>.08</td>
</tr>
<tr>
<td>Zhejiang province</td>
<td>2893 (68.64)</td>
<td>25 (0.86)</td>
<td>Reference</td>
</tr>
<tr>
<td>Other provinces</td>
<td>1322 (31.36)</td>
<td>5 (0.38)</td>
<td>1.17 (0.40-3.48)</td>
</tr>
<tr>
<td>Hypertension</td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>No</td>
<td>364 (8.64)</td>
<td>6 (1.65)</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>3851 (91.36)</td>
<td>24 (0.62)</td>
<td>0.71 (0.26-1.95)</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>No</td>
<td>164 (3.89)</td>
<td>4 (2.44)</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>4051 (96.11)</td>
<td>26 (0.64)</td>
<td>1.10 (0.31-3.92)</td>
</tr>
<tr>
<td>Chronic obstructive pneumonia</td>
<td></td>
<td></td>
<td>&gt;.99</td>
</tr>
<tr>
<td>No</td>
<td>16 (0.38)</td>
<td>0 (0)</td>
<td>b</td>
</tr>
<tr>
<td>Yes</td>
<td>4199 (99.62)</td>
<td>30 (0.71)</td>
<td>—</td>
</tr>
<tr>
<td>Cardiovascular diseases</td>
<td></td>
<td></td>
<td>&gt;.99</td>
</tr>
<tr>
<td>No</td>
<td>48 (1.14)</td>
<td>0 (0)</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>4167 (98.86)</td>
<td>30 (0.72)</td>
<td>—</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>21 (0.5)</td>
<td>2 (9.52)</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>4194 (99.5)</td>
<td>28 (0.67)</td>
<td>1.10 (0.31-3.92)</td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>30 (0.71)</td>
<td>4 (13.33)</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>4185 (99.29)</td>
<td>26 (0.62)</td>
<td>12.19 (2.98-49.93)</td>
</tr>
<tr>
<td>Variable</td>
<td>COVID-19 cases (n=4215), n (%)</td>
<td>Hospitalization People living with HIV, n (%)</td>
<td>Multivariable analysis</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P value</td>
<td>aOR(^a) (95% CI)</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>18 (0.43)</td>
<td>0 (0)</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>4197 (99.57)</td>
<td>30 (0.71)</td>
<td>—</td>
</tr>
<tr>
<td>Transmission route</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>2142 (50.82)</td>
<td>25 (1.17)</td>
<td>Reference</td>
</tr>
<tr>
<td>Homosexual</td>
<td>2033 (48.23)</td>
<td>4 (0.2)</td>
<td>Reference</td>
</tr>
<tr>
<td>Other or unknown</td>
<td>40 (0.95)</td>
<td>1 (2.5)</td>
<td>Reference</td>
</tr>
<tr>
<td>Latest CD4 count (cells/μL)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-199</td>
<td>282 (6.69)</td>
<td>10 (3.55)</td>
<td>Reference</td>
</tr>
<tr>
<td>200-499</td>
<td>2034 (48.26)</td>
<td>8 (0.39)</td>
<td>Reference</td>
</tr>
<tr>
<td>≥500</td>
<td>1899 (45.05)</td>
<td>12 (0.63)</td>
<td>Reference</td>
</tr>
<tr>
<td>Latest HIV load (copies/mL)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-49</td>
<td>3984 (94.52)</td>
<td>25 (0.63)</td>
<td>Reference</td>
</tr>
<tr>
<td>50-999</td>
<td>107 (2.54)</td>
<td>2 (1.87)</td>
<td>Reference</td>
</tr>
<tr>
<td>≥1000</td>
<td>124 (2.94)</td>
<td>3 (2.42)</td>
<td>Reference</td>
</tr>
<tr>
<td>Time since ART(^c) initiation (cells/μL)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>156 (3.7)</td>
<td>1 (0.64)</td>
<td>—</td>
</tr>
<tr>
<td>1</td>
<td>407 (9.66)</td>
<td>4 (0.98)</td>
<td>—</td>
</tr>
<tr>
<td>2-4</td>
<td>1325 (31.44)</td>
<td>8 (0.6)</td>
<td>—</td>
</tr>
<tr>
<td>≥5</td>
<td>2327 (55.21)</td>
<td>17 (0.73)</td>
<td>—</td>
</tr>
<tr>
<td>TDF(^d) use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1825 (43.3)</td>
<td>14 (0.77)</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>2390 (56.7)</td>
<td>16 (0.67)</td>
<td>—</td>
</tr>
<tr>
<td>3TC(^e) use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>718 (17.03)</td>
<td>5 (0.7)</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>3497 (82.97)</td>
<td>25 (0.71)</td>
<td>—</td>
</tr>
<tr>
<td>EFV(^f) use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1535 (36.42)</td>
<td>13 (0.85)</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>2680 (63.58)</td>
<td>17 (0.63)</td>
<td>—</td>
</tr>
<tr>
<td>AZT(^g) use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3256 (77.25)</td>
<td>25 (0.77)</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>959 (22.75)</td>
<td>5 (0.52)</td>
<td>—</td>
</tr>
<tr>
<td>LPV/r(^h) use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3902 (92.57)</td>
<td>26 (0.67)</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>313 (7.43)</td>
<td>4 (1.28)</td>
<td>—</td>
</tr>
<tr>
<td>Time since last vaccination (months)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No vaccination</td>
<td>633 (15.02)</td>
<td>13 (2.05)</td>
<td>Reference</td>
</tr>
<tr>
<td>0-5</td>
<td>240 (5.69)</td>
<td>1 (0.42)</td>
<td>0.21 (0.03-1.73)</td>
</tr>
<tr>
<td>6-11</td>
<td>1990 (47.21)</td>
<td>12 (0.6)</td>
<td>0.33 (0.14-0.81)</td>
</tr>
<tr>
<td>Variable</td>
<td>COVID-19 cases (n=4215), n (%)</td>
<td>Hospitalization</td>
<td>Multivariable analysis</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td>People living with HIV, n (%)</td>
<td>P value</td>
<td>aOR(^a) (95% CI)</td>
</tr>
<tr>
<td>≥12</td>
<td>1352 (32.08)</td>
<td>4 (0.3)</td>
<td>0.22 (0.07-0.72)</td>
</tr>
</tbody>
</table>

\(a\) aOR: adjusted odds ratio.
\(b\) Not available.
\(c\) ART: antiretroviral treatment.
\(d\) TDF: tenofovir.
\(e\) 3TC: lamivudine.
\(f\) EFV: efavirenz.
\(g\) AZT: azidothymidine.
\(h\) LPV/r: lopinavir and ritonavir.

**Figure 3.** Association between vaccination and SARS-CoV-2 infection among group A in Zhejiang province from December 2022 to February 2023 (group A definition: those who presented with symptoms such as fever or cough were classified as having SARS-CoV-2 infection, while those who did not exhibit such symptoms were categorized as noninfected). aOR: adjusted odds ratio.
Figure 4. Association between vaccination and SARS-CoV-2 infection among group C in Zhejiang province from December 2022 to February 2023 (group C encompassed all study participants, including group A and group B. Group A definition: those who presented with symptoms such as fever or cough were classified as having SARS-CoV-2 infection, while those who did not exhibit such symptoms were categorized as noninfected. Group B comprised individuals who had undergone COVID-19 testing between December 7, 2022, and the investigation date. Infection status was determined based on the positive or negative results obtained from nucleic acid or antigen tests). aOR: adjusted odds ratio.

Discussion

Overview

The public health impact of the COVID-19 pandemic in the community continues. Previous studies have investigated the SARS-CoV-2 infection and hospitalization and their associated risk factors among people living with HIV [6,19,20]. However, most of these studies were conducted during the early stages of the pandemic or when the Delta variant was predominant. Furthermore, due to the limited number of infections resulting from implementing the Dynamic Zero policy, no studies regarding the Omicron variant have been reported in China. This study aimed to investigate SARS-CoV-2 infection and hospitalization among people living with HIV during the Omicron variant epidemic in China. The infection rate was 66.97% (4215/6294) among people living with HIV with a COVID-19 testing history, whereas the hospitalization rate was 0.71% (30/4215). Moreover, we found that general risk factors such as higher age, vaccination status, and HIV-related parameters such as latest CD4 count and HIV load were associated with SARS-CoV-2 infection and hospitalization, and these results will provide scientific information for future control and prevention of COVID-19 among people living with HIV.

The SARS-CoV-2 infection rate was 66.97% (4215/6294) in this study, which was lower than the rate reported by Fu et al [11] (82.4%), the rate estimated by modeling calculation conducted by Leung et al (92.3%) [14], and Bai et al (87.54%) [15], and higher than the rate investigated among children conducted by Su et al [12] (40.9%) in the general population after adjustment of the Zero-COVID policy. During the early pandemic, studies mostly demonstrated a lower or similar incidence of SARS-CoV-2 infection among people living with HIV when compared to those without HIV in the general population [4]. The lower infection rate among people living with HIV could be due to the following 2 reasons: first, many studies have indicated that COVID-19 has a great impact on people living with HIV, and thus they usually have greater social distancing [4,21]. Second, some researchers have argued that antiretroviral medication, specifically nucleoside or nucleotide reverse transcriptase inhibitors, may have contributed to a protective effect [8,22]. However, the lower rate can be attributed to other factors such as sample size, confounding biases, and geographical, social, and health care disparities, which should be concluded with caution and deserve further study.

We found that age, education, marital status, transmission route, latest CD4 cell count, and HIV viral load were statistically associated with SARS-CoV-2 infection. Younger adults had a higher risk of SARS-CoV-2 infection, potentially due to a higher force of infection in working age groups; this result was consistent with a study conducted among the general population in Shanghai [17] and people living with HIV in Lebanon [8]. In this study, being married or cohabiting was independently associated with infection, which could be related to intrafamily transmission. Furthermore, consistent with the study conducted by Kim and Jeong [23], we found that the higher the education
level, the higher the SARS-CoV-2 infection rate. Interestingly, people living with HIV with higher latest CD4 counts had a higher risk of infection, while those with higher HIV loads had a lower risk. This phenomenon may also contribute to illustrating the lower infection rate in people living with HIV in our study compared to studies in the general population. Nonetheless, the reasons for this phenomenon require further investigation. In addition, people living with HIV who contracted HIV from homosexual transmission had a higher risk of SARS-CoV-2 infection than those who contracted HIV from heterosexual transmission. This higher risk may be attributed to higher mobility and multiple sexual partners among men who have sex with men, which may increase their risk of infection [24-26].

According to a World Health Organization report, people living with HIV have a significantly higher risk of experiencing a severe or fatal COVID-19 infection [27]. Consistent with previous studies, we identified several factors associated with an increased risk of hospitalization among people living with HIV. These factors include older age, lower latest CD4 count, higher latest HIV viral load, and the presence of cancer [3,6,7,19,28]. Therefore, people living with HIV with these risk factors should be given priority care and protection.

In this study, we found 3TC had a protective effect against SARS-CoV-2 infection, and this protective effect was not shown in other commonly used drugs (TDF, EFV, AZT, and LPV/r). Previous studies have demonstrated that there was a molecular basis for the inhibition of the SARS-CoV-2 RNA-dependent RNA polymerase by these nucleotide analogs [29,30]. However, the effectiveness of antiretrovirals in preventing COVID-19 infection or severe illness among people living with HIV has yielded mixed results in different studies and trials [8,22,31]. Therefore, it is currently impossible to definitively conclude whether antiretrovirals can protect people from contracting COVID-19 or experiencing severe illness from the virus.

Data on the effectiveness of COVID-19 vaccines among people living with HIV in real-world studies remain limited, as they have not been adequately represented in initial vaccine effectiveness trials. In our study, we found that vaccination was associated with reduced SARS-CoV-2 infection within 3 months of the last vaccination. Our subgroup analyses further demonstrated that the protective effect only occurred in the booster vaccination participants. Our findings were consistent with the vaccine effectiveness of COVID-19 evaluated among the general population in China [11]. Furthermore, the significant effect of vaccination in reducing hospitalization in our study aligns with observations made in the general population and among people living with HIV in other studies [17,32]. Notably, the duration of protection against hospitalization varied and may be attributed to the small sample size of hospitalized cases. Since no studies specifically focused on vaccination efficacy against SARS-CoV-2 Omicron infection and hospitalization among the population of people living with HIV in the country, our study provided a scientific basis for the vaccination program among people living with HIV. Consequently, we highly recommend expanding vaccination coverage among people living with HIV when the epidemic reemerges since 15% (1530/9922) of people living with HIV in our study remained unvaccinated.

**Limitations**

This study had several limitations. Firstly, a causal relationship could not be established between the investigated factors and the infection or hospitalization of SARS-CoV-2 due to the cross-sectional nature of our study. Secondly, although we used random sampling at the beginning of participant selection, the inability to collect data from refusal would lead to selection bias. Our findings should be extrapolated with caution. Thirdly, diagnoses of conditions such as SARS-CoV-2 infection, hospitalization, and other chronic diseases such as diabetes, hypertension, and cardiovascular disease rely on self-reporting, which can be influenced by social desirability and reporting biases. Fourthly, other potential risk factors for infection, such as individual behaviors and close contact infections, which may impact infection, were not investigated. Finally, the number of hospitalizations was too small to conduct subgroup analyses and further estimations of the effectiveness of different vaccination strategies against hospitalization.

**Conclusions**

We conducted the first investigation of HIV/SARS-CoV-2 coinfection and hospitalization following the ease of the control policy in China. Our findings revealed a high prevalence of COVID-19 but a low hospitalization rate. Besides general characteristics such as age and vaccination status, we also found that people living with HIV–specific factors such as CD4 cell count and HIV viral load were associated with SARS-CoV-2 infection and hospitalization. These findings will add to the growing literature on SARS-CoV-2 and HIV coinfection and contribute to optimizing COVID-19 preventive strategies among people living with HIV. A booster vaccination campaign for booster doses should be considered among people living with HIV in confronting possible COVID-19 epidemic emergencies in the near future.

**Acknowledgments**

We would like to thank public health physicians and staff from the county and city level of the Centers for Disease Control and Prevention across Zhejiang province for their invaluable assistance with HIV/AIDS data collection.

This work was supported by the key research and development program of Zhejiang Province (2021C03200); the key program of the Health Commission of Zhejiang Province/Science Foundation of the National Health Commission (WKJ-ZJ-2221); the major program of the Zhejiang Municipal Natural Science Foundation (LD22H190001); the explorer program of Zhejiang Municipal Natural Science Foundation (LQ23H100001); and the Zhejiang Provincial Soft Science Research Projects (2024C35088).
The data analyzed in our study are available from the corresponding author on reasonable request.

JJ and CC conceived and designed the study. WC analyzed the data. YX, HJ, JL, ZH, HM, and WW contributed to reagents, materials, and field investigation. WC wrote the paper. JJ and CC reviewed the manuscript. All authors read and approved the final manuscript.

None declared.

References

1. HIV data and statistics. World Health Organization. 2023. URL: https://tinyurl.com/3n7ej46k [accessed 2023-03-20]

Data Availability

The data analyzed in our study are available from the corresponding author on reasonable request.

Authors’ Contributions

JJ and CC conceived and designed the study. WC analyzed the data. YX, HJ, JL, ZH, HM, and WW contributed to reagents, materials, and field investigation. WC wrote the paper. JJ and CC reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.


Abbreviations

3TC: lamivudine
aOR: adjusted odds ratio
ART: antiretroviral treatment
AZT: azidothymidine
CDC: Center for Disease Control and Prevention
CHO cell: Chinese hamster ovary cell
EFV: efavirenz
LPV/r: lopinavir and ritonavir
TDF: tenofovir
Edited by A Mavragani, T Sanchez; submitted 01.08.23; peer-reviewed by Z Wang, KC Chong, X Zhang; comments to author 08.12.23; revised version received 07.01.24; accepted 12.03.24; published 17.04.24.

Please cite as:
SARS-CoV-2 Infection, Hospitalization, and Associated Factors Among People Living With HIV in Southeastern China From December 2022 to February 2023: Cross-Sectional Survey
JMIR Public Health Surveill 2024;10:e51449
URL: https://publichealth.jmir.org/2024/1/e51449
doi:10.2196/51449
PMID:38630534

©Wei Cheng, Yun Xu, Haibo Jiang, Jun Li, Zhigang Hou, Haibin Meng, Wei Wang, Chengliang Chai, Jianmin Jiang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
A Machine Learning Model for Identifying Sexual Health Influencers to Promote the Secondary Distribution of HIV Self-Testing Among Gay, Bisexual, and Other Men Who Have Sex With Men in China: Quasi-Experimental Study

Yuxin Ni1,2,3*, MPH; Ying Lu2*, MA; Fengshi Jing4,5*, PhD; Qianyun Wang2,6*, MSW; Yewei Xie2,7, MS; Xi He8, BS; Dan Wu9, PhD; Rayner Kay Jin Tan2,10, PhD; Joseph D Tucker2,11,12, MD, PhD; Xumeng Yan2,13, MA; Jason J Ong11,14, PhD; Qingpeng Zhang15,16, PhD; Hongbo Jiang17, PhD; Wencan Dai18, MPH; Liqun Huang18, MPH; Wenhua Mei18, PhD; Yi Zhou18*, PhD; Weiming Tang1,2*, MD, PhD

1Dermatology Hospital of Southern Medical University, Guangzhou, China
2University of North Carolina at Chapel Hill Project-China, Guangzhou, China
3Department of Health Law, Policy, and Management, Boston University School of Public Health, Boston, MA, United States
4Institute for Healthcare Artificial Intelligence Application, Guangdong Second Provincial General Hospital, Guangzhou, China
5Faculty of Data Science, City University of Macau, Macao SAR, China
6Department of Social Welfare, University of California, Los Angeles, CA, United States
7Health Service and System Research Programme, Duke-NUS Medical School, Singapore, Singapore
8Zhuhai Xutong Voluntary Services Center, Zhuhai, China
9Department of Social Medicine and Health Education, School of Public Health, Nanjing Medical University, Nanjing, China
10Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
11Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom
12Central Clinical School, Monash University, Melbourne, Australia
13Muskeeters Foundation Institute of Data Science, The University of Hong Kong, Hong Kong, China (Hong Kong)
14Department of Pharmacology and Pharmacy, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
15Department of Epidemiology and Biostatistics, School of Public Health, Guangdong Pharmaceutical University, Guangzhou, China
16Zhuhai Center for Disease Control and Prevention, Zhuhai, China
17Department of Community Health Sciences, Fielding School of Public Health, University of California, Los Angeles, CA, United States
18these authors contributed equally

Corresponding Author:
Weiming Tang, MD, PhD
Dermatology Hospital of Southern Medical University
2 Lujing Rd, Yuexiu District
Guangzhou, 510091
China
Phone: 86 13302391532
Email: weiming_tang@med.unc.edu

Abstract

Background: Sexual health influencers (SHIs) are individuals actively sharing sexual health information with their peers, and they play an important role in promoting HIV care services, including the secondary distribution of HIV self-testing (SD-HIVST). Previous studies used a 6-item empirical leadership scale to identify SHIs. However, this approach may be biased as it does not consider individuals’ social networks.

Objective: This study used a quasi-experimental study design to evaluate how well a newly developed machine learning (ML) model identifies SHIs in promoting SD-HIVST compared to SHIs identified by a scale whose validity had been tested before.
Methods: We recruited participants from BlueD, the largest social networking app for gay men in China. Based on their responses to the baseline survey, the ML model and scale were used to identify SHIs, respectively. This study consisted of 2 rounds, differing in the upper limit of the number of HIVST kits and peer-referral links that SHIs could order and distribute (first round ≤5 and second round ≤10). Consented SHIs could order multiple HIV self-testing (HIVST) kits and generate personalized peer-referral links through a web-based platform managed by a partnered gay-friendly community-based organization. SHIs were encouraged to share additional kits and peer-referral links with their social contacts (defined as “alters”). SHIs would receive US $3 incentives when their corresponding alters uploaded valid photographic testing results to the same platform. Our primary outcomes included (1) the number of alters who conducted HIVST in each group and (2) the number of newly tested alters who conducted HIVST in each. We used negative binomial regression to examine group differences during the first round (February-June 2021), the second round (June-November 2021), and the combined first and second rounds, respectively.

Results: In January 2021, a total of 1828 men who have sex with men (MSM) completed the survey. Overall, 393 SHIs (scale=195 and ML model=198) agreed to participate in SD-HIVST. Among them, 229 SHIs (scale=116 and ML model=113) ordered HIVST on the web. Compared with the scale group, SHIs in the ML model group motivated more alters to conduct HIVST (mean difference [MD] 0.88, 95% CI 0.02-2.22; adjusted incidence risk ratio [aIRR] 1.77, 95% CI 1.07-2.95) when we combined the first and second rounds. Although the mean number of newly tested alters was slightly higher in the ML model group than in the scale group, the group difference was insignificant (MD 0.35, 95% CI –0.17 to –0.99; aIRR 1.49, 95% CI 0.74-3.02).

Conclusions: Among Chinese MSM, SHIs identified by the ML model can motivate more individuals to conduct HIVST than those identified by the scale. Future research can focus on how to adapt the ML model to encourage newly tested individuals to conduct HIVST.

Trial Registration: Chinese Clinical Trials Registry ChiCTR2000039632; https://www.chictr.org.cn/showprojEN.html?proj=63068

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-021-11817-2

(JMIR Public Health Surveill 2024;10:e50656) doi:10.2196/50656

KEYWORDS

artificial intelligence; HIV testing; key opinion leaders; machine learning; men who have sex with men; self-testing

Introduction

Globally, gay, bisexual, and other men who have sex with men (GBMSM) exhibit a significantly higher median HIV prevalence of 7.5%, in contrast to the global median of 0.7% [1]. In China, GBMSM are also disproportionately affected by HIV, accounting for 25.5% of new HIV infections in 2018 (vs 6.1% in 2008) [2]. Despite the efforts in HIV control, the HIV testing coverage among Chinese GBMSM is still low. In 2019, the national HIV prevalence of GBMSM was 6.3%, yet only 56% knew their HIV status in China [3]. Enhancing people’s awareness of their own HIV status could accelerate the initiation of treatment. People living with HIV who start treatment earlier have a higher likelihood of improving their survival and can also prevent HIV transmission by maintaining an undetectable viral load status (undetectable=untransmittable) [4].

Preventing HIV by increasing HIV testing rates is a key pillar in reaching the first “95-95-95” target, proposed by UNAIDS (the Joint United Nations Programme on HIV/AIDS), that 95% of all individuals living with HIV will be aware of their HIV status by 2025 [5]. One of the innovative approaches to expanding HIV testing coverage is HIV self-testing (HIVST). Specifically, HIVST is an effective decentralized testing tool that allows persons to test in a comfortable and private setting and mitigates concerns about inadvertent disclosure and confidentiality protection associated with facility-based testing [6]. In addition, HIVST may present an opportunity to address barriers that GBMSM may encounter when seeking care in health care facilities [7]. These barriers include but are not limited to, lack of access to health care, fear of HIV-related stigma, perceived discrimination, lack of confidence in health services, and limited knowledge and awareness about HIV [6-9].

HIVST can be conducted with or without assistance in various settings to reach diverse populations [10,11]. Directly assisted HIVST involves in-person guidance from trained providers or peers, while unassisted HIVST relies solely on manufacturer instructions [12]. A systematic review and meta-analysis indicated that self-testers achieve HIVST results with reliability and accuracy similar to trained health care workers [13]. In light of this, secondary distribution of HIV self-testing (SD-HIVST) was conducted within the unassisted HIVST implementation strategy.

In 2023, the World Health Organization (WHO) recommends expanding the use of HIVST services; one of the recommendations is using individuals’ sexual and social networks to promote HIVST and increase testing coverage [14]. One promising model is the social network–based SD-HIVST. As an innovative social network method, SD-HIVST is promising to help reach naïve testers and increase HIVST uptake among men who have never tested before [15-17]. The strategy of SD-HIVST involves individuals (defined as “index”) obtaining multiple HIVST kits and distributing them face-to-face or through peer referral (PR) links on the web to members within their social networks (defined as “alters”) [15,16]. There is increasing evidence demonstrating the effectiveness of SD-HIVST in improving HIV testing coverage and case identification worldwide [15-20]. In China, Wu et al [15] adapted the SD-HIVST model into a social media–based existing HIVST service provided by a GBMSM-friendly community-led
organization. Wu et al [15] also included components to improve test accuracy and result return rates. Building on the progress of Wu et al [15], Zhou et al [16] conducted a randomized trial with a modified model by providing monetary incentives to GBMSM index participants who motivated alters to upload their results to the web-based system. The study concluded that SD-HIVST with economic incentives could effectively promote SD-HIVST and encourage newly tested receivers to test. Consequently, this adapted digital secondary distribution strategy was culturally appropriate in the Chinese setting, where GBMSM influencers hold a premise in HIVST distribution to their social contacts with limited access to self-testing [15,16].

Sexual health influencers (SHIs) are individuals with a strong influence on disseminating information about HIV and sexually transmitted infections (STIs) within their social networks [21-23]. Studies have shown that those at the center of their networks may influence their peers’ adoption of healthier behaviors [21,22]. It was also reported that SHIs are important for improving HIV care-related services. For example, a study has shown that GBMSM with strong influence tend to participate in more sexual health campaigns and use HIV and syphilis testing [21]. The social comparison theory states that SHIs might serve as role models for well-being and self-improvement within the GBMSM community, offering an ideal to which other GBMSM can strive [24].

Artificial intelligence (AI), particularly machine learning (ML), shows promise in identifying key influencers and excels in classifying key populations, including pinpointing individuals at higher HIV risk [25-28]. The conventional empirical scales (ES) used a 6-item leadership questionnaire and were previously used to identify GBMSM SHIs and examine their demographic and behavioral characteristics [29]. A limitation of the ES is that there may be bias in the perception of one’s influence or leadership, and it did not consider the social network of these individuals. Thus, more objective and reliable measurements are required [21]. The ML model exhibits better flexibility, whereby it can incorporate more variables and measure SHIs more comprehensively.

To overcome the challenges of the ES to identify SHIs and to tailor this for SD-HIVST, we developed and validated an ML model (ie, an artificial algorithm) to help identify GBMSM SHIs [28]. This trial aimed to compare the effectiveness of SD-HIVST initiated by SHIs identified through the ML model in promoting alters for testing among Chinese GBMSM compared to the ES method. The authors of this paper hypothesized that SHIs identified using the ML model would be more effective in motivating more alters to conduct HIVST in comparison to SHIs identified by the scale.

**Methods**

**Ethical Considerations**

We registered this trial on the Chinese Clinical Trial Registry website (ChiCTR2000039632). We obtained the institutional review board approval through the Dermatology Hospital of Southern Medical University (2019020[R3]). We obtained different electronic consents from study participants before they started the screening survey, ordered HIVST kits, and uploaded the HIVST results. We provided each participant with 20 RMB (US $3) in compensation through WeChat Pay (China’s most popular social networking app with a mobile payment function) for their time spent completing a survey. We only used participants’ identifiable information for compensation and to eliminate any duplicate entries. We used deidentified data to conduct the analysis.

**Study Setting and Recruitment**

We used a quasi-experimental study design. This study was a collaboration between the Social Entrepreneurship to Spur Health team, the Zhuhai Center for Disease Control and Prevention, and the Zhuhai Xutong Voluntary Services Center, a GBMSM community-based organization (CBO).

The study participants were recruited from 5 provinces in southern China, including Guangdong, Guangxi, Fujian, Hunan, and Jiangxi, by putting recruiting banner ads on BlueD, China’s largest gay social networking app. Interested GBMSM who clicked the banner would answer the presurvey questions to complete the eligibility check. Eligible individuals were then sent an online baseline survey (Figure 1).
Eligibility Criteria

Participants met the following eligibility criteria: (1) were assigned male sex at birth, (2) were aged 18 years or older, (3) self-reported ever having sex with men, and (4) were willing to engage in the SD-HIVST program.

Identifying Sexual Health Influencers

To reduce self-report bias, we excluded respondents (survey outliers) who answered with more than 20 sexual partners in the previous 3 months. We randomly assigned eligible individuals to 1 of 2 groups in a 1:1 ratio. Identified SHIs were eligible to engage in the SD-HIVST program. Index participants were SHI who agreed to participate in the study.

A 6-item scale was used to identify SHIs among GBMSM allocated to the ES. This scale was adapted from the one (Cronbach α = 0.937) used in our previous study to assess GBMSM’s ability to spread information about sexual health [30,31]. The self-reported scale comprises 6 inquiries designed to evaluate each participant’s involvement in disseminating information about HIV and STIs, both through web-based and offline channels, over the past 3 months. These six questions assess (1) the frequency of their discussions with others regarding HIV and STI, (2) the extent to which they shared information about HIV and STI with others, (3) the number of individuals they informed about HIV and STI, (4) other people’s willingness to seek additional information about HIV and STI from them, (5) whether they or others within their social network were the primary sources of information about HIV and STI, and (6) how often they have sought advice about HIV and STIs. Participants received a total response score ranging from 0 to 24 based on their answers to these questions. Subsequently, SHIs were ranked according to their total scores. More details can be found in the protocol [32].

For the ML model group, we used our developed model to identify SHIs. The model was trained using data from our earlier SD-HIVST program [33]. We used an ensemble approach that involved 4 distinct ML models: logistic regression, random forest, decision tree, and support vector machine. By combining these 4 ML models, we developed our novel ensemble ML model. To ensure the models’ robustness, we applied a 5-fold cross-validation technique and evaluated the average metrics for each model [28]. We gathered data on the selected predictors from the baseline survey and fed it into our already-trained ML model, considering variables selected from social network characteristics, HIV testing, kit application willingness, GBMSM behavior, etc [16].

Conducting SD-HIVST

Identified SHIs were invited to conduct SD-HIVST on a web-based private platform, which we developed and piloted [15,16]. Index participants accessed the platform by following CBO’s public account on WeChat. CBO posted HIV knowledge and HIVST-related information (eg, kit manufacturer, things to be considered before ordering HIVST kits, and the ordering process) on their WeChat platform. Participants could click the link and read testing-related and HIV-related information. They could also send direct messages to CBO staff based on their needs. In addition, participants could also message research team staff if they had any research-related concerns.

HIVST kits, each including a delivery-tracking code, a result-tracking code, and a user guide, were sent by postal mail to the address provided by the applicant. Testers submitted their results by uploading images to the platform, accessed through a QR code on the package. CBO staff confirmed the results by reviewing the submitted photos. Then, CBO staff would message the verified result to each tester and provide counseling based on the tester’s needs. If a tester had reactive results, the CBO staff would provide information and link them to a confirmation test.

There were 2 rounds of HIVST ordering, and identified SHIs were invited at each round. The first round was from February to June 2021, and the second round was from June to November 2021.

In the first round, index participants could request a maximum of 5 HIVST kits and acquire a unique PR link, allowing them to invite up to 5 alters to order 1 HIVST kit each. In the second round, we lifted the upper limit to 10 HIVST kits and allowed index participants to invite up to 10 alters through the unique PR link. We asked for a refundable deposit of 40 RMB (US $7) per kit at each request.

Index participants would receive 20 RMB (US $3) as incentives for each validated testing result of alters. Furthermore, all unique testers earned 20 RMB (US $3) after verifying their testing results.

Outcomes

We identified unique alters as individuals who received tests from index participants and reported results using unique cell phone numbers. If alters were tested twice, we kept their most up-to-date results. We further classified alters who self-reported having never tested for HIV before this research as unique newly tested alters.

Primary outcomes included (1) the number of unique alters motivated by index participants in each group and (2) the number of unique newly tested alters motivated by index participants in each group. They were assessed when HIVST testers completed the return survey. Secondary outcomes included (1) the number of testers tested with HIV-reactive results, (2) the number of unique alters motivated by index participants categorized in subgroups, and (3) the economic cost comparison between the ES and the ML model.

Sample Size Calculation

We calculated a sample size of 400 SHIs with 80% power, a 2-sided α =.05, and a 20% loss to follow-up. According to our preliminary study data, we assumed the mean difference (MD) was 1 between the 2 groups in unique alters motivation and 0.3 in newly tested alters motivation [32]. We conducted sample size estimation using PASS software (version 16.0; NCSS, LLC).

Missing Data and Sensitivity Analysis

No index participants dropped out of the study during the entire study period. In data analysis, we only included baseline survey
data related to index participants’ sociodemographics without using follow-up survey data, and the baseline data was complete without any missing data. When uploading test results, some alters might choose their role as “index participants” to skip answering the alters’ survey through which we collected whether they were newly tested. Since CBO staff worked closely with index participants, they verified the identity of each uploaded result and made a note in the database for reference researchers.

We conducted sensitivity analyses under two assumptions: (1) considering all individuals who skipped the survey as newly tested, and (2) considering all individuals who skipped the survey as not newly tested.

Statistical Methods
We used the Shapiro-Wilk method to test the data distribution of baseline characteristics of index participants and presented the mean (SD) for the primary outcomes of each group. We calculated MD between groups and used bootstrapping to estimate corresponding 95% CIs. We used negative binomial regression for primary outcome analyses and adjusted covariates, including age group, income, education, residence permit, marital status, sexual orientation, and residence. We performed the same regressions on subgroup analysis, whereby we classified index participants based on their age, sexual orientation, sexual orientation disclosure, residence permit, education, and province of residence. We believe negative binomial regression is more suitable than other methods for the following reasons: (1) the variance of the count number (ie, outcomes) is larger than the mean value (Table 1), thus assumptions of Poisson regression are not met; (2) we conducted tests for the presence of excess 0s in the data. The results indicated that there were no excess 0s, suggesting that the data was not a suitable fit for 0-inflated negative binomial regression.

We considered $P<.05$ as statistically significant and performed data analysis in RStudio (version 1.2.5033; Posit).

We used a microcosting method to assess the economic cost (ie, the cost of all resources needed to implement the testing models) from the perspective of a health provider. We kept track of the resources used throughout the experiment and classified cost items. We collected cost data (Table S1 in Multimedia Appendix 1) and analyzed costs. All expenses are expressed in 2021 US dollars using OANDA currency conversions (US $1=6.51 RMB). We conducted the cost analysis in Excel 2019 (Microsoft Corporation).

Results
Overview
We received 2150 recruitment surveys from January 7 to January 28, 2021. After excluding 5% outliers from 1922 respondents, we identified 1828 unique participants with verified contact information. We randomly assigned 1828 participants to each group (ML model=914 and ES=914). We further identified 465 SHIs in the ES and 465 SHIs in the ML model group. After being invited to participate in SD-HIVST, 393 identified SHIs (ES=195 and ML model=198) further confirmed their interests, and 229 index participants (ES=116 and ML model=113) initiated SD-HIVST and were included in this study. SD-HIVST was implemented from February 2021 to November 2021 (Figure 2).
**Figure 2.** Flowchart of the quasi-experimental study: secondary distribution of HIV self-testing kits among gay, bisexual, and other men who have sex with men (GBMSM) in China, from January to November 2021. SHI: sexual health influencer.

### Sociodemographic Characteristics

In total, 229 index participants ordered SD-HIVST, with 59.5% (116/195) in the ES and 57.1% (113/198) in the ML model ordering HIVST kits on the web.

The social demographic characteristics were similar between the 2 groups, except for education status (Table 1). Overall, the mean age of index participants was 27.58 (SD 7.75) years, 76% (175/229) of them obtained at least a college degree, 87.3% (200/229) were single, 72.9% (167/229) self-identified as gay, 68.6% (157/229) disclosed their sexual orientation to people who are not gay, and 91.7% (210/229) ever tested for HIV.
Table 1. Baseline characteristics of index participants (n=229) in China, stratified by the empirical scale (ES) group versus the machine learning (ML) model group, from February to November 2021. A 2-tailed t test was applied to compare ages between the ES and ML models. Chi-square tests were applied to test differences in categorical characteristics between the ES and ML models.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>ES (n=116)a</th>
<th>ML model (n=113)b</th>
<th>Overall (N=229)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>27.43 (7.91)</td>
<td>27.73 (7.61)</td>
<td>27.58 (7.75)</td>
<td>.77</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>81 (69.8)</td>
<td>80 (70.8)</td>
<td>161 (70.3)</td>
<td></td>
</tr>
<tr>
<td>≥30</td>
<td>35 (30.2)</td>
<td>33 (29.2)</td>
<td>68 (29.7)</td>
<td></td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.047</td>
</tr>
<tr>
<td>High school or below</td>
<td>34 (29.3)</td>
<td>20 (17.7)</td>
<td>54 (23.6)</td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>73 (62.9)</td>
<td>88 (77.9)</td>
<td>161 (70.3)</td>
<td></td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>9 (7.8)</td>
<td>5 (4.4)</td>
<td>14 (6.1)</td>
<td></td>
</tr>
<tr>
<td>Annual income (US $), n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.68</td>
</tr>
<tr>
<td>&lt;2700</td>
<td>19 (16.4)</td>
<td>13 (11.5)</td>
<td>32 (14.0)</td>
<td></td>
</tr>
<tr>
<td>2700-5400</td>
<td>19 (16.4)</td>
<td>17 (15.0)</td>
<td>36 (15.7)</td>
<td></td>
</tr>
<tr>
<td>5401-9000</td>
<td>30 (25.9)</td>
<td>26 (23.0)</td>
<td>56 (24.5)</td>
<td></td>
</tr>
<tr>
<td>9001-14,400</td>
<td>22 (19.0)</td>
<td>27 (23.9)</td>
<td>49 (21.4)</td>
<td></td>
</tr>
<tr>
<td>&gt;14,400</td>
<td>26 (22.4)</td>
<td>30 (26.5)</td>
<td>56 (24.5)</td>
<td></td>
</tr>
<tr>
<td>Sexual orientation, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.83</td>
</tr>
<tr>
<td>Gay</td>
<td>84 (72.4)</td>
<td>83 (73.4)</td>
<td>167 (72.9)</td>
<td></td>
</tr>
<tr>
<td>Bisexual</td>
<td>26 (22.4)</td>
<td>26 (23.0)</td>
<td>52 (22.7)</td>
<td></td>
</tr>
<tr>
<td>Unsure</td>
<td>6 (5.2)</td>
<td>4 (3.5)</td>
<td>10 (4.4)</td>
<td></td>
</tr>
<tr>
<td>Sexual orientation disclosurec</td>
<td></td>
<td></td>
<td></td>
<td>.40</td>
</tr>
<tr>
<td>No</td>
<td>40 (34.5)</td>
<td>32 (28.3)</td>
<td>72 (31.4)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>76 (65.5)</td>
<td>81 (71.6)</td>
<td>157 (68.6)</td>
<td></td>
</tr>
<tr>
<td>Marital statusd, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.16</td>
</tr>
<tr>
<td>Engaged or married</td>
<td>12 (10.3)</td>
<td>8 (7.1)</td>
<td>20 (8.7)</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>97 (83.6)</td>
<td>103 (91.2)</td>
<td>200 (87.3)</td>
<td></td>
</tr>
<tr>
<td>Separated, divorced, or widowed</td>
<td>7 (6.0)</td>
<td>2 (1.8)</td>
<td>9 (3.9)</td>
<td></td>
</tr>
<tr>
<td>Residence, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.40</td>
</tr>
<tr>
<td>Guangdong Province</td>
<td>68 (58.6)</td>
<td>59 (52.2)</td>
<td>127 (55.5)</td>
<td></td>
</tr>
<tr>
<td>Other provinces</td>
<td>48 (41.4)</td>
<td>54 (47.8)</td>
<td>102 (44.5)</td>
<td></td>
</tr>
<tr>
<td>Residence permit, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>Rural</td>
<td>72 (62.1)</td>
<td>57 (50.4)</td>
<td>129 (56.3)</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>44 (37.9)</td>
<td>56 (49.6)</td>
<td>100 (43.7)</td>
<td></td>
</tr>
<tr>
<td>Condomless sex with other men in the past 3 monthse</td>
<td></td>
<td></td>
<td></td>
<td>.87</td>
</tr>
<tr>
<td>Yes</td>
<td>47 (40.5)</td>
<td>48 (42.5)</td>
<td>95 (41.5)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>69 (59.5)</td>
<td>65 (57.5)</td>
<td>134 (58.5)</td>
<td></td>
</tr>
<tr>
<td>Ever tested for HIV, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.17</td>
</tr>
<tr>
<td>Yes</td>
<td>103 (88.8)</td>
<td>107 (94.7)</td>
<td>210 (91.7)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>13 (11.2)</td>
<td>6 (5.3)</td>
<td>19 (8.3)</td>
<td></td>
</tr>
</tbody>
</table>

a A total of 59.5% (116/195) participants in the ES ordered HIV self-testing kits on the web.

b A total of 57.1% (113/198) participants in the ML model ordered HIV self-testing kits on the web.

c Index participants disclosed their sexual orientation to other people who were self-identified as not gay.
Marital status refers to heterosexual marriage.

In the past 3 months refers to October 2020 to January 2021.

**Application (First Round)**

Overall, 224 index participants ordered HIVST kits through the platform. In the ES, 113 index participants ordered 244 kits and referred to 21 PR links. We received 227 results, with a return rate of 85.7% (227/265), among which 72 results were from 72 unique alters.

In the ML model, 111 index participants ordered 267 kits and referred 41 PR links to alters. We received 261 results, with a return rate of 84.7% (261/308), among which 129 results were from 127 unique alters.

A total of 84.1% (95/113) index participants in the ES completed the 3-month follow-up survey, while 86.5% (96/111) index participants in the ML model completed the survey.

**Application (Second Round)**

In the second round, 85 index participants ordered kits through the platform. A total of 46 index participants in the ES ordered 127 kits and referred to 31 PR links. We received 133 results, with a return rate of 84.2% (133/158), among which 34 results were from 30 unique alters.

In the ML model, 39 index participants ordered 132 kits and referred 32 PR links to alters. We received 129 results, with a return rate of 78.6% (129/164), among which 79 results were from 72 unique alters.

A total of 87% (40/46) index participants in the ES and 87% (34/39) participants in the ML model completed the 3-month follow-up survey.

**Unique Alter Motivation Results**

Table 2 presents the primary outcome results for 229 index participants in the 6-month study period on the effect of motivating unique alters and newly tested alters. We estimated a total effect from 2 rounds, with separate effects for the first and second rounds.

Overall, index participants in the ML model motivated an average of 1.76 (199/113, SD 3.97) unique alters to conduct HIVST compared with 0.88 (102/116, SD 1.84) in the ES, with an MD of 0.88 (95% CI 0.02-2.22). Index participants in the ML model were more likely to motivate unique alters than those in the ES (adjusted incidence rate ratio [aIRR] 1.77, 95% CI 1.07-2.95).
Table 2. The effectiveness in motivating alters and newly tested alters to conduct HIV self-testing (HIVST), comparing index participants in the empirical scale (ES) group to those in the machine learning (ML) model group in China, February to November 2021. We kept the unique index participants when analyzing the outcomes of the first and second rounds. Therefore, the number of overall index participants is not equal to the sum of the absolute number of 2 rounds of index participants.

<table>
<thead>
<tr>
<th>Outcomes and number of index participants</th>
<th>Number of unique alters motivated by each index participant</th>
<th>Number of newly tested alters motivated by each index participant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First round and second round</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (n=116), n</td>
<td>102</td>
<td>44</td>
</tr>
<tr>
<td>ES (n=116), mean (SD)</td>
<td>0.88 (1.84)</td>
<td>0.38 (0.89)</td>
</tr>
<tr>
<td>ML model (n=113), n</td>
<td>199</td>
<td>82</td>
</tr>
<tr>
<td>ML model (n=113), mean (SD)</td>
<td>1.76 (3.97)</td>
<td>0.73 (2.32)</td>
</tr>
<tr>
<td>ES (reference) versus ML model, MD&lt;sup&gt;a&lt;/sup&gt; (95% CI)</td>
<td>0.88 (0.02 to 2.22)</td>
<td>0.35 (–0.17 to 0.99)</td>
</tr>
<tr>
<td>ES (reference) versus ML model, aIRR&lt;sup&gt;b&lt;/sup&gt; (95% CI)</td>
<td>1.77 (1.07 to 2.95)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.49 (0.74 to 3.02)</td>
</tr>
<tr>
<td><strong>First round (from February to June 2021)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (n=113), n</td>
<td>72</td>
<td>32</td>
</tr>
<tr>
<td>ES (n=113), mean (SD)</td>
<td>0.64 (1.39)</td>
<td>0.28 (0.70)</td>
</tr>
<tr>
<td>ML model (n=111), n</td>
<td>127</td>
<td>41</td>
</tr>
<tr>
<td>ML model (n=111), mean (SD)</td>
<td>1.14 (1.80)</td>
<td>0.37 (0.77)</td>
</tr>
<tr>
<td>ES (reference) versus ML model, MD (95% CI)</td>
<td>0.50 (–0.01 to 1.02)</td>
<td>0.09 (–0.17 to 0.34)</td>
</tr>
<tr>
<td>ES (reference) versus ML model, aIRR (95% CI)</td>
<td>1.85 (1.10 to 3.12)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.56 (0.75 to 3.23)</td>
</tr>
<tr>
<td><strong>Second round (from June to November 2021)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES (n=46), n</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>ES (n=46), mean (SD)</td>
<td>0.65 (1.48)</td>
<td>0.26 (0.74)</td>
</tr>
<tr>
<td>ML model (n=39)</td>
<td>72</td>
<td>41</td>
</tr>
<tr>
<td>ML model (n=39), mean (SD)</td>
<td>1.85 (3.98)</td>
<td>1.05 (3.03)</td>
</tr>
<tr>
<td>ES (reference) versus ML model, MD (95% CI)</td>
<td>1.20 (–0.25 to 3.20)</td>
<td>0.79 (–0.17 to 2.17)</td>
</tr>
<tr>
<td>ES (reference) versus ML model, aIRR (95% CI)</td>
<td>1.75 (0.83 to 3.69)</td>
<td>—.d</td>
</tr>
</tbody>
</table>

<sup>a</sup>MD: mean difference.

<sup>b</sup>aIRR: adjusted incidence rate ratio. A negative binomial model was used to calculate the incidence rate ratio (IRR). aIRR was adjusted for age (<30 and ≥30 years), income, education, marital status (referring to heterosexual marriage), sexual orientation, and residence.

<sup>c</sup>P<.05.

<sup>d</sup>The Hauck-Donner effect occurred, causing the estimated coefficients to be in extreme situations.

**Newly Tested Alters Motivation Results**
Throughout 2 rounds, index participants in the ML model averagely motivated 0.73 (82/113) unique newly tested alters, compared with 0.38 (44/116) in the ES. The performance of index participants in the ML model was slightly better than that in the ES (MD 0.35, 95% CI —0.17 to 0.99), but the difference between these 2 groups was not statistically significant. (Table 2.) We identified 4 alters who chose their role as index participants. Our sensitivity analysis results showed a similar pattern.

**Subgroup Analysis Results**
Table 3 reveals subgroup analysis results. Index participants in the ML model who were aged 30 years or younger, not gay, or residents of Guangdong were more likely to motivate unique alters to conduct HIVST than those who were in the ES.
Table 3. The effectiveness in motivating alters to conduct HIV self-testing (HIVST), comparing index participants in the empirical scale (ES) group to those in the machine learning (ML) model group, stratified by their baseline demographics in China, from February to November 2021.

<table>
<thead>
<tr>
<th>Variables</th>
<th>ES (reference) versus ML model, aIRR(^a) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>2.06 (1.09-3.91)(^b)</td>
</tr>
<tr>
<td>≥30</td>
<td>—(^c)</td>
</tr>
<tr>
<td><strong>Sexual orientation</strong></td>
<td></td>
</tr>
<tr>
<td>Gay</td>
<td>1.51 (0.85-2.69)</td>
</tr>
<tr>
<td>Others(^d)</td>
<td>4.16 (1.31-14.60)(^b)</td>
</tr>
<tr>
<td><strong>Sexual orientation disclosure(^e)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2.20 (0.82-6.06)</td>
</tr>
<tr>
<td>Yes</td>
<td>1.50 (0.80-2.83)</td>
</tr>
<tr>
<td><strong>Residence permit</strong></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>1.83 (0.98-3.44)</td>
</tr>
<tr>
<td>Urban</td>
<td>1.37 (0.65-2.95)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>High school or below</td>
<td>2.12 (0.94-4.87)</td>
</tr>
<tr>
<td>College or above</td>
<td>1.69 (0.92-3.09)</td>
</tr>
<tr>
<td><strong>Province</strong></td>
<td></td>
</tr>
<tr>
<td>Guangdong Province</td>
<td>1.81 (0.97-3.42)(^b)</td>
</tr>
<tr>
<td>Outside Guangdong</td>
<td>1.74 (0.76-3.97)</td>
</tr>
</tbody>
</table>

\(^a\)aIRR: adjusted incidence rate ratio. A negative binomial model was used to calculate the incidence rate ratio (IRR). aIRRs were adjusted for age, income, education, marital status, sexual orientation, and residence.

\(^b\)P<0.05.

\(^c\)The Hauck-Donner effect occurred, causing the estimated coefficients to be in an extreme situation.

\(^d\)Others include bisexual and unsure.

\(^e\)Index participants disclosed their sexual orientation to other people who were self-identified as not gay.

**HIVST Reactive Results**

Among 14 testers who were newly tested HIV reactive, 6 were index participants (3 in the ES and 3 in the ML model), and 8 were alters (4 in the ES and 4 in the ML model). CBO staff followed up with all testers with reactive results.

**Cost Analysis**

The total economic cost of implementing the SD-HIVST was US $6547.12 for the ES and US $6767.19 for the ML model. When comparing the ML model and ES approaches, the ML model is more effective in encouraging participants to get tested for HIV, but it is not cost-saving. The incremental cost to gain an additional tested participant is US $2.27, and for a newly tested participant, the incremental cost is US $5.79 (Table 4).
Table 4. Cost analysis evaluating the effectiveness in motivating alters and newly tested alters to conduct HIV self-testing (HIVST), comparing index participants in the empirical scale (ES) group to those in the machine learning (ML) model group, from February to November 2021.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>ES</th>
<th>ML model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost per alters (US $)</td>
<td>64.19</td>
<td>37.05</td>
</tr>
<tr>
<td>Incremental cost (US $)</td>
<td>Reference group</td>
<td>-27.13</td>
</tr>
<tr>
<td>Unique alters motivated by each index participant, n</td>
<td>102</td>
<td>199</td>
</tr>
<tr>
<td>ICER(^a)</td>
<td>N/A(^b)</td>
<td>Dominated(^c)</td>
</tr>
<tr>
<td>Total economic cost (US $)</td>
<td>6547.2</td>
<td>6767.19</td>
</tr>
<tr>
<td>Incremental cost (US $)</td>
<td>N/A</td>
<td>220.07</td>
</tr>
<tr>
<td>Newly tested alters motivated by each index participant being tested, n</td>
<td>44</td>
<td>82</td>
</tr>
<tr>
<td>ICER</td>
<td>N/A</td>
<td>5.79</td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.
\(^b\)ICER: incremental cost-effectiveness ratio.
\(^c\)Dominated means sexual health influencers in the ML model are more effective and cost-saving compared to their counterparts in the ES model.

**Discussion**

**Overview**

We conducted a quasi-experimental study among Chinese MSM and found that, overall, SHIs identified by the ML model can influence more individuals through their social network to conduct HIVST than SHIs identified by the conventional scale, highlighting the potential of artificial AI to enhance the efficiency of testing approaches. Building on the evidence that monetary incentives and peer referral can amplify the effect of SD-HIVST, this study extends the literature by comparing the effectiveness of the ES and a ML model in identifying SHIs for secondary distribution and reporting results on the total number of unique alters and newly tested alters motivated by the identified SHIs [15,16]. In addition, we acknowledged the important role that community engagement plays in promoting HIV testing among GBMSM. Identifying SHIs to implement community-led interventions could expand HIV testing services. As influential community members, SHIs build on their potential to support their peers in improving health awareness and behaviors, thereby engaging more individuals from their social networks and facilitating HIVST uptake.

**Principal Findings**

The ML model was more effective than the ES in engaging unique alters for HIVST during SD-HIVST. This is aligned with the results of a previous study using an interactive deep learning approach to identify HIV influencers; that is, not only ML models can detect HIV-related digital social influencers, but also ML models show higher accuracy in identification [34]. Despite studies showing that SHIs and noninfluencers can promote HIV testing within their social networks, the groups they can reach differ, particularly alters of SHIs who sought out more casual sexual partners, which increased their risk of HIV infection [22,23]. In light of this, it is promising to integrate them into the current SD-HIVST programs to facilitate HIV testing uptake and case identification further. Both the ML model and the ES revealed modest differences in the motivation of newly tested alters. The results of the ML model were statistically insignificant in promoting first-time testing compared to the ES; this may be due to the power of the study. However, in terms of implementation, SHIs identified by the ML model motivated a larger number of newly tested alters and promoted first-time testing coverage. Previous studies have demonstrated that first-time testing among GBMSM is associated with sociodemographic and other factors such as the number of sexual partners, condomless sex, and STI infection history [30,31]. Further refinement of the interventions may be warranted to reach more newly tested alters.

**Comparison to Previous Work**

Evidence-based HIV prevention strategies can be significantly enhanced through interdisciplinary collaboration between communities and academia [33]. At the community level, our research was rooted in strong partnerships between the academic and community sectors. We closely worked with the local gay-led CBO to evaluate the interventions designed for specific populations, ensuring a collaborative and informed approach to this study. A review paper demonstrated that AI, exemplified by Van Heerden et al [36] mobile app using natural language processing and Brown et al [37] partner notification model with unconditional logistic regression, shows promise in enhancing HIV counseling, testing, and partner notification strategies [33,36,37]. Our research extends the existing literature on the application of AI in HIV testing, specifically concentrating on identifying SHIs and expanding the effectiveness of the current SD-HIVST program.

Additionally, several studies have been done on the feasibility and acceptability of digital interventions, and further strategies to accelerate digital interventions for HIV prevention are needed [38,39]. ML as a data science approach has sparked significant research interest in developing digital health interventions [33,40]. The strength of our ML model lies in both accuracy and flexibility. In building the ML model, more specific variables were incorporated and weighted differently to define SHIs, such as the number of copies of HIVST wanted, the intention to deliver HIVST, and the willingness to obtain PR links [41]. Further, the absence of the data sets from the previous study for the ES or social network-based simulation optimization...
models renders them less accurate than the ML model in identifying SHIs during SD-HIVST [42].

**Strengths and Limitations**

This study also has several limitations. First, this study is quasi-experimental, which may be subject to selection bias. Only half of the identified SHIs agreed to participate, and only one-quarter initiated the HIV self-test kit application process. However, the social demographic characteristics were comparable between the 2 study arms. Second, the ML model did not outperform the ES in promoting newly tested alters. The main reason for this could be the selection of parameters when building the ML model that primarily motivates unique alters. Future research regarding motivating the newly tested alters requires additional consideration of the selection of variables and the weight given to them during modeling. Finally, compared to our previous studies, this study reported a relatively lower result return rate among the testers, mainly due to a much lower deposit, which reduces the “sunk cost” for the testers so that they are less inclined to return the results [15,16]. A higher deposit fee may be advisable for similar programs to ensure a higher rate of returns. However, our sensitivity analysis indicated that the impact of these nonreturns has a minimum impact on this study results.

**Future Directions**

As for implications, this study demonstrated that interventions led and delivered by CBO could facilitate participation. The local GBMSM-friendly CBO undertook the health service delivery in this study, including the construction and maintenance of the HIVST kit ordering platform, HIVST kit distribution, counseling, results interpretation, and linkage to care services. This was consistent with a systematic review that community engagement interventions are beneficial for marginalized populations in increasing health behaviors, health outcomes, participant self-efficacy, and perceived social support [43]. Particularly in COVID-19, our innovative digital health intervention can guarantee access to health services while saving costs. This ML model method of identifying SHIs may also apply to other regions and other interventions, as it has good extensibility and flexibility. Depending on the interventions (including but not limited to SD-HIVST), we can build the ML model based on different selected associated variables.

**Conclusion**

Identifying SHIs can enhance the effectiveness of social network–based intervention among GBMSM. We found that SHIs identified by the ML model can motivate more individuals to conduct HIVST than those identified by the ES in SD-HIVST programs among Chinese GBMSM. This study innovatively adopted data science techniques to develop the ML model for identifying SHIs and delivering interventions in collaboration with community-based organizations. This strategy can be adapted and implemented to improve HIV care services.

---

**Acknowledgments**

The authors would like to thank all the gay, bisexual, and other men who have sex with men who participated and Xutong community-based organization staff for their dedication to this study. We did not use ChatGPT or other generative AI tools.

This work was supported by the Key Technologies Research and Development Program (2022YFC2304900-4 to WT), the National Institute of Health (R34MH119963 to WT), CRDF Global (G-202104-67775 to WT), the Doctoral Workstation Foundation of Guangdong Second Provincial General Hospital (2023BSGZ014 to FJ), and the Guangzhou Science and Technology Project (SL2022A04J01130 to FJ).

The funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

**Data Availability**

The data sets generated or analyzed during this study are not publicly available due to identifiable information related to participants but are available from the corresponding author upon reasonable request.

**Authors’ Contributions**

YN, YL, FJ, QW, and YX drafted and finalized the paper with inputs from YZ and WT. DW, RKJT, JDT, XY, JJO, and QZ provided critical revisions of the paper. YZ, XH, SH, HJ, DW, and WT conceived the study. XH, WD, LH, and WM assisted with recruitment. All authors read and authorized the final version.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Table S1 and other supplementary material.

[DOCX File, 25 KB - publichealth_v10i1e50656_app1.docx ]

**References**

https://publichealth.jmir.org/2024/1/e50656

JMIR Public Health Surveill 2024 | vol. 10 | e50656 | p.982

(page number not for citation purposes)


Abbreviations

AI: Artificial intelligence  
aIRR: adjusted incidence risk ratio  
CBO: community-based organization  
ES: empirical scales  
GBMSM: gay, bisexual, and other men who have sex with men  
HIVST: HIV self-testing  
MD: mean difference  
ML: machine learning  
MSM: men who have sex with men  
PR: peer referral  
SD-HIVST: secondary distribution of HIV self-testing  
SHI: sexual health influencer  
STI: sexually transmitted infection  
UNAIDS: Joint United Nations Programme on HIV/AIDS  
WHO: World Health Organization

©Yuxin Ni, Ying Lu, Fengshi Jing, Qianyun Wang, Yewei Xie, Xi He, Dan Wu, Rayner Kay Jin Tan, Joseph D Tucker, Xumeng Yan, Jason J Ong, Qingpeng Zhang, Hongbo Jiang, Wencan Dai, Liqun Huang, Wenhua Mei, Yi Zhou, Weiming Tang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 24.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Attrition Rates in HIV Viral Load Monitoring and Factors Associated With Overdue Testing Among Children Within South Africa’s Antiretroviral Treatment Program: Retrospective Descriptive Analysis

Ahmad Haeri Mazanderani\textsuperscript{1,2,3}, MBChB, MMed, PhD; Lebohang Radebe\textsuperscript{1,3}, BSc (Hons), MSc; Gayle G Sherman\textsuperscript{1,2,3}, MBBCh, MMed, PhD

\textsuperscript{1}Centre for HIV & STIs, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, South Africa
\textsuperscript{2}Department of Paediatrics & Child Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
\textsuperscript{3}Paediatric HIV Diagnostics Division, Wits Health Consortium, University of the Witwatersrand, Johannesburg, South Africa

Corresponding Author:
Ahmad Haeri Mazanderani, MBChB, MMed, PhD
Centre for HIV & STIs
National Institute for Communicable Diseases
National Health Laboratory Service
1 Modderfontein Road
Sandringham
Johannesburg, 2031
South Africa
Phone: 27 826428609
Email: ahmadh@nicd.ac.za

Abstract

Background: Numerous studies in South Africa have reported low HIV viral load (VL) suppression and high attrition rates within the pediatric HIV treatment program.

Objective: Using routine laboratory data, we evaluated HIV VL monitoring, including mobility and overdue VL (OVL) testing, within 5 priority districts in South Africa.

Methods: We performed a retrospective descriptive analysis of National Health Laboratory Service (NHLS) data for children and adolescents aged 1-15 years having undergone HIV VL testing between May 1, 2019, and April 30, 2020, from 152 facilities within the City of Johannesburg, City of Tshwane, eThekwini, uMgungundlovu, and Zululand. HIV VL test–level data were deduplicated to patient-level data using the NHLS CDW (Corporate Data Warehouse) probabilistic record-linking algorithm and then further manually deduplicated. An OVL was defined as no subsequent VL determined within 18 months of the last test. Variables associated with the last VL test, including age, sex, VL findings, district type, and facility type, are described. A multivariate logistic regression analysis was performed to identify variables associated with an OVL test.

Results: Among 21,338 children and adolescents aged 1-15 years who had an HIV VL test, 72.70% (n=15,512) had a follow-up VL test within 18 months. Furthermore, 13.33% (n=2194) of them were followed up at a different facility, of whom 3.79% (n=624) were in a different district and 1.71% (n=281) were in a different province. Among patients with a VL of $\geq 1000$ RNA copies/mL of plasma, the median time to subsequent testing was 6 (IQR 4-10) months. The younger the age of the patient, the greater the proportion with an OVL, ranging from a peak of 52% among 1-year-olds to a trough of 21% among 14-year-olds. On multivariate analysis, 2 consecutive HIV VL findings of $\geq 1000$ RNA copies/mL of plasma were associated with an increased adjusted odds ratio (AOR) of having an OVL (AOR 2.07, 95% CI 1.71-2.51). Conversely, patients examined at a hospital (AOR 0.86, 95% CI 0.77-0.96), those with $\geq 2$ previous tests (AOR 0.78, 95% CI 0.70-0.86), those examined in a rural district (AOR 0.63, 95% CI 0.54-0.73), and older age groups of 5-9 years (AOR 0.56, 95% CI 0.47-0.65) and 10-14 years (AOR 0.51, 95% CI 0.44-0.59) compared to 1-4 years were associated with a significantly decreased odds of having an OVL test.

Conclusions: Considerable attrition occurs within South Africa’s pediatric HIV treatment program, with over one-fourth of children having an OVL test 18 months subsequent to their previous test. In particular, younger children and those with virological...
failure were found to be at increased risk of having an OVL test. Improved HIV VL monitoring is essential for improving outcomes within South Africa’s pediatric antiretroviral treatment program.

**Introduction**

South Africa has an estimated 8 million people living with HIV [1]. Although the national Prevention of Mother to Child Transmission program, launched in earnest in 2004, has reduced the vertical transmission rate to approximately 3%, there remain around 280,000 children and adolescents nationwide aged <15 years living with HIV. Furthermore, pediatric and adolescent populations living with HIV have far worse clinical outcomes than older age groups [2]. Modeling studies from South Africa estimate that while the proportion of the total population living with HIV who are diagnosed, those accessing antiretroviral therapy (ART), and those with viral suppression (<1000 RNA copies/mL of plasma [hereinafter “copies/mL”]) is 94%, 69%, and 63%, respectively, for pediatric populations, these proportions are considerably lower at 82%, 57%, and 39%, respectively [1]. Partly in response to the disparity between pediatric and adult treatment outcomes, the UNAIDS (Joint United Nations Programme on HIV/AIDS) launched the Global Alliance to End AIDS in Children by 2030 program, with South Africa officially having launched its chapter in 2023 [3].

Poor pediatric linkage to care, high attrition rates, and high mortality rates have been described from a number of cohort studies across the country [4-6]. Such studies support the identification of variables associated with attrition—this is considered an important step toward improving care. Virological nonsuppression has been identified as one such key risk factor for subsequent program loss from both improving care and rural cohorts [7]. However, as variables associated with attrition may vary among regions and are likely to be influenced by a myriad of socioeconomic and political factors, additional mechanisms are urgently needed to not only identify children at risk of disengaging with care but also support the challenging process of re-engagement. Effective use of routine laboratory data offers opportunities toward this end.

The National Health Laboratory Service (NHLS) is the only clinical laboratory service provider within South Africa’s public health sector. As such, the NHLS Data Warehouses can potentially be leveraged for cost-effective disease surveillance and near–real-time reporting, with the advantage of not being restricted to facility-based monitoring as is the case with the national electronic patient monitoring system, TIER.Net. The ability of laboratory data to support the ready disaggregation of population-level data to provide a more nuanced understanding of program outcomes has been demonstrated for the pediatric HIV program, with younger children found to have markedly lower viral suppression rates than older children, and male patients being less likely to be suppressed than female patients across all age groups [8]. Furthermore, reporting of patient-identifying consolidated exceptions, both at district and facility levels, as indicated in the Results for Action reports of the National Institute for Communicable Diseases (NICD), a division of the NHLS, enables the timely identification of children in need of urgent follow-up based on HIV diagnostic and monitoring findings [9]. However, the potential for routine laboratory data for evaluating pediatric attrition in near real time, thereby supporting retention in care, has not previously been described.

In the following analysis, we aimed to evaluate HIV viral load (VL) monitoring within the pediatric HIV program by reporting on patient mobility and overdue VL (OVL) testing within 5 high-burden health districts in South Africa.

**Methods**

**Study Design**

This study is a retrospective descriptive analysis of routine laboratory data from South Africa’s public health sector, spanning a 4-year period. Data for children and adolescents aged <15 years living with HIV, who had undergone HIV VL testing between May 1, 2019, and April 30, 2020, from 152 facilities within 5 districts, were extracted from the NICD Data Warehouse. Test-level data were deduplicated to patient-level data by 2 sequential methods. The first one involved assigning a unique patient identifier (UPI) to all results by using a probabilistic record-linking algorithm, referred to in the text as the NHLS CDW (Corporate Date Warehouse) unique identifier (hereinafter referred to as the “CDW algorithm”), which links HIV tests performed nationally within the NHLS. This algorithm has been validated within the adult HIV program, with a reported sensitivity of 84% and incorrectly matching 0.3% of results [10]. Second, records were manually deduplicated by 4 data experts, according to a set of agreed upon rules listed in Multimedia Appendix 1 and referred to in the text as the manual deduplication UPI. Manual deduplication linked HIV tests for each patient within the 152 facilities only. Prior and subsequent tests were linked using both the CDW algorithm and manual deduplication identifiers. Prior tests from October 1, 2017, were linked to the cohort to describe the number of previous tests and number of previous high HIV VL (≥1000 copies/mL) results. Subsequent tests for an 18-month period up to October 31, 2021, were linked to the cohort to describe follow-up testing. An OVL test was defined as no subsequent VL test within 18 months of the last VL test, irrespective of the last HIV VL result. An 18-month period was informed by national ART guidelines, which recommend that all patients undergo HIV VL monitoring at least once every 12 months. While other similar analyses
have used a 24-month period [11,12], an 18-month period (ie, considering a 6-month grace period) was deemed sufficient and supported the inclusion of data up until April 2020 at the time of analysis. Due to known limitations with linking infant tests using demographic data [13], analysis of OVL tests was restricted to children and adolescents aged 1-15 years.

Setting and Laboratory Testing Methods

HIV viral load routine laboratory data were analyzed from 5 high-HIV-burden districts in South Africa considered by the National Department of Health as having the largest pediatric treatment gaps in the country, as determined from modeling data: City of Johannesburg and City of Tshwane in Gauteng Province, and eThekwini, uMgungundlovu, and Zululand in KwaZulu-Natal Province [14]. Data were restricted to patients who had undergone HIV VL testing at one of 152 designated facilities that have additional pediatric HIV support from District Support Partner organizations, representing 30% of all facilities and approximately two-thirds of all HIV VL testing performed for children and adolescents aged <15 years within these districts.

HIV VL monitoring is routinely recommended 6 months after ART initiation, with repeat testing performed at 12 months and 12-monthly thereafter if viral suppression is detected. The cutoff that defines suppression was revised in the 2019 national guidelines from <1000 copies/mL to ≤50 copies/mL, with patients with an suppressed VL requiring repeat testing at 3 months after a thorough assessment and management of the cause of an elevated VL [15].

Within the NHLS, HIV VL testing is performed using plasma samples at centralized laboratories using either the Cobas HIV-1 Test (Roche Diagnostics) or the Abbott RealTime HIV-1 assay (Abbott Molecular, Inc).

Statistical Analysis

Variables associated with the last HIV VL finding per patient, tested between May 1, 2019, and April 30, 2020, are reported; these include age group (<1, 1-4, 5-9, or 10-14 years), sex (male, female, or unknown), VL test result (≤50, 50-999, or ≥1000 copies/mL), district type (metropolitan districts classified as urban versus nonmetropolitan districts classified as rural), and facility type (hospital versus primary health care facility). The number of patients according to each deduplication step is described by comparing the number and percentage decrease of patients reported through manual deduplication of UPIs and deduplication performed using the CDW algorithm per variable. The proportion of people with an OVL was evaluated for both the CDW algorithm, these corresponded to 24,398 patients, while the manually deduplicated UPIs corresponded to 22,978 patients, representing a deficit of 5.82% (n=1420). This deficit was associated with considerable variation among variables except for sex and facility type (Figure 1). Facilities in the urban districts had a greater deficit in the number of participants (n=1144) than those in rural districts (n=276); however, this represented a smaller deficit in the proportion of people (5.7% vs 6.3%, respectively; Figure 1A). In terms of age, the 1–4-year age group had the highest percentage of deduplication at 6.4% (n=217), while the ≤1-year age group had the lowest proportion of deduplication at 4.0% (n=68; Figure 1B). There was considerable variation in deduplication among HIV VL result categories: those with a VL of ≤50 copies/mL decreasing by 2.7% (n=376), those with a VL of 50-999 copies/mL decreasing by 7.6% (n=390), and those with a VL of ≥1000 copies/mL decreasing by 11.8% (n=654; Figure 1C). In terms of sex and facility type, the variation was minimal, with a 5.9% (n=733) reduction among females and a 5.8% (n=647) reduction among males (Figure 1D), and a 6.0% (n=806) reduction for clinics compared with a 5.6% (n=614) reduction for hospitals (Figure 1E). In terms of the number of linked previous tests, there was a marked decrease in number of patients with a single HIV VL test and those with no prior high (≥1000 copies/mL) HIV VL finding (Figures 1F and 1G).

Differences were also observed in the number and proportion of patients with an OVL test after manual deduplication when compared with the CDW algorithm. According to the CDW algorithm, 22,690 children and adolescents aged 1-15 years had undergone an HIV VL test between May 2019 and April 2020 within the 152 designated facilities, of whom 9685 (42.68%) had an OVL test. After manual deduplication, 21,338 children and adolescents were identified, of whom 5826 (27.3%) had an OVL test, representing a reduction of 39.85% (n=3859) of patients considered to have an OVL test.

Ethical Considerations

The National Institute for Communicable Diseases has ethics approval for communicable disease surveillance and analysis of routine laboratory data by the Human Research Ethics Committee of the University of the Witwatersrand (M160667; M210752), with the requirement for informed consent having been waived. All analysis with patient-identifiable data was performed on a secure password-protected server located on the NHLS campus.

Results

Data Deduplication

A total of 29,822 HIV VL tests, registered among children and adolescents aged <15 years, were carried out between May 2019 and April 2020 at 152 facilities. According to the CDW algorithm, these corresponded to 24,398 patients, while the manually deduplicated UPIs corresponded to 22,978 patients, representing a deficit of 5.82% (n=1420). This deficit was associated with considerable variation among variables except for sex and facility type (Figure 1). Facilities in the urban districts had a greater deficit in the number of participants (n=1144) than those in rural districts (n=276); however, this represented a smaller deficit in the proportion of people (5.7% vs 6.3%, respectively; Figure 1A). In terms of age, the 1–4-year age group had the highest percentage of deduplication at 6.4% (n=217), while the ≤1-year age group had the lowest proportion of deduplication at 4.0% (n=68; Figure 1B). There was considerable variation in deduplication among HIV VL result categories: those with a VL of ≤50 copies/mL decreasing by 2.7% (n=376), those with a VL of 50-999 copies/mL decreasing by 7.6% (n=390), and those with a VL of ≥1000 copies/mL decreasing by 11.8% (n=654; Figure 1C). In terms of sex and facility type, the variation was minimal, with a 5.9% (n=733) reduction among females and a 5.8% (n=647) reduction among males (Figure 1D), and a 6.0% (n=806) reduction for clinics compared with a 5.6% (n=614) reduction for hospitals (Figure 1E). In terms of the number of linked previous tests, there was a marked decrease in number of patients with a single HIV VL test and those with no prior high (≥1000 copies/mL) HIV VL finding (Figures 1F and 1G).

Differences were also observed in the number and proportion of patients with an OVL test after manual deduplication when compared with the CDW algorithm. According to the CDW algorithm, 22,690 children and adolescents aged 1-15 years had undergone an HIV VL test between May 2019 and April 2020 within the 152 designated facilities, of whom 9685 (42.68%) had an OVL test. After manual deduplication, 21,338 children and adolescents were identified, of whom 5826 (27.3%) had an OVL test, representing a reduction of 39.85% (n=3859) of patients considered to have an OVL test.
Figure 1. Number and percentage deduplication of test-level to patient-level data using 2 unique patient identifiers (UPIs) among children and adolescents aged <15 years with an HIV viral load (VL) test between May 2019 and April 2020 from the City of Johannesburg, City of Tshwane, eThekwini, uMgungundlovu, and Zululand by (A) district, (B) age group, (C) VL test result, (D) sex, (E) facility type, (F) number of previous linked tests, and (G) number of previous linked high VL (>999 RNA copies/mL of blood) findings. CDW: Corporate Data Warehouse; cps/ml: RNA copies/mL of blood.

Follow-Up HIV VL Testing and Movement Across Facilities

Among the manually deduplicated data set of 21,338 children and adolescents aged 1-15 years who had undergone an HIV VL test between May 2019 and April 2020, a total of 54,160 tests were registered at 771 different facilities. Considering the results of the first test registered with each patient, 11,889 (55.7%) had a VL of <50 copies/mL, 4416 (20.7%) had a VL of 50-999 copies/mL, and 5033 (23.6%) had a VL of ≥1000 copies/mL. In terms of location, 17,341 (81.3%) tests were first registered in an urban district and 3997 (18.7%) were registered in a rural district. At the time of the first test, there were 3303 (15.5%) patients aged 1-4 years, 6126 (28.7%) patients aged 5-9 years, and 11,909 (55.8%) patients aged 10-14 years.

Of the 21,338 children and adolescents, 77.16% (n=16,465) of them had undergone a follow-up VL test, of whom 72.70% (n=15,512) had undergone a follow-up HIV VL test within 18 months and 4.47% (n=953) had a follow-up test after 18 months. Overall, 14,363 (87.2%) patients were followed up at the same facility, 1963 (11.9%) were followed up at 2 facilities, 120 (0.7%) were followed up at 3 facilities, 10 (0.01%) were followed up at 4 facilities, and 8 (0.01%) were followed up at 5 facilities. In total, 2102 (12.8%) patients were followed up at a different facility, of whom 604 (3.7%) were in a different district and 270 (1.6%) were in a different province. The median time to subsequent testing varied by VL result (Figure 2). For those with a VL of <50 copies/mL (n=9944), the median time of subsequent testing was 12.1 (IQR 11.1-13.1) months. For those with a VL finding of 50-999 copies/mL (n=3505), the median time to subsequent testing was 11.2 (IQR 7.2-12.8) months, and for those with a VL finding of ≥1000 copies/mL (n=3016), the median time to subsequent testing was 6.5 (IQR 4.2-10.3) months. For those with a single VL finding of ≥1000 copies/mL (n=1056), the median follow-up time was 7.1 (IQR 4.7-11.4) months, whereas for those with a repeat VL of ≥1000 copies/mL
(n=1960), the median follow-up time was 6.1 (IQR 4.0-9.6) months. The change in the national guidelines in 2019, whereby the threshold for viral suppression was revised from <1000 to copies/mL to <50 copies/mL, was associated with a biphasic pattern of time to follow-up among patients with a finding between 50 and 999 copies/mL.

Urban districts had a greater number and proportion of people (n=5077, 29.17%) with an OVL test than rural districts (n=749, 19.05%; Figure 3A). In terms of age, the 1-4-year age group had the highest proportion of OVL tests at 42.98% (n=1372), while the 10-14-year age group had the lowest at 23.83% (Figure 3B). However, given the number of patients, this age category also had the highest number of people with an OVL test (n=2882). In terms of HIV VL test results, 25.57% (n=3209) of patients with a VL of <50 copies/mL and 25.83% (n=1150) of those with a VL of 50-999 copies/mL had an OVL test, while those with a VL of ≥1000 copies/mL accounted for the greatest proportion of OVL tests, at 33.83% (n=1467; Figure 3C). In terms of sex, a slightly higher proportion of females (n=3052, 28.45%) had an OVL test than males (n=2614, 26.06%; Figure 3D). Furthermore, a slightly higher proportion of people who underwent VL testing at a clinic had an OVL (n=3419, 28.51%) than those who underwent testing at a hospital (n=2407, 25.76%; Figure 3E).

In terms of the number of linked previous tests, the largest proportion of people with an OVL test was observed among patients with no linked prior tests (n=3126, 64.39%; Figure 3F). Not only did having at least one prior linked VL test considerably reduce the proportion of people with an OVL test, but also this proportion generally decreased as the number of previous linked tests increased. This ranged from a peak of 21.96% (n=1217) for those with one prior VL test to a trough of 6.64% (n=16) for those with 6 prior tests. This pattern was replicated in the number of previous high VL findings (>999 copies/mL; Figure 3G).

**Figure 2.** Time to follow-up HIV viral load (VL) testing (in months) by VL findings, with medians reported per result category based on manual deduplication of unique patient identifiers among children and adolescents aged 1-15 years having undergone an HIV VL test between May 2019 and April 2020 from the City of Johannesburg, City of Tshwane, eThekwini, uMgungundlovu, and Zululand. cps/ml: RNA copies/mL of blood.
Figure 3. Number and percentage deduplication of test-level to patient-level data after manual deduplication for overdue viral load (VL) tests among children and adolescents aged 1-15 years having undergone an HIV viral load test between May 2019 and April 2020 from the City of Johannesburg, City of Tshwane, uMgungundlovu, and Zululand by (A) district, (B) age, (C) VL findings, (D) sex, (E) facility type, (F) number of previous linked tests, and (G) number of previous linked high HIV VL (>999 RNA copies/mL of blood) findings. cps/ml: RNA copies/mL of blood.

OVL Tests by Age and Outcome
As the age of the patient increased, the proportion of patients having an OVL test decreased, ranging from a peak of 52% among 1-year-olds to a trough of 21% among 14-year-olds (Figure 4A). This trend was observed across HIV VL result categories, with a higher proportion of OVL tests among patients with a VL of ≥1000 copies/mL (Figure 4B). Because of differences in population size among the age groups, the absolute number of patients with an OVL test varied, with high numbers of patients in the older age range having an OVL test (Figure 4A).
Figure 4. Number and percentage of people per age group (in years) with an overdue HIV viral load (VL) test among children and adolescents aged 1-15 years having undergone an HIV VL test between May 2019 and April 2020 from the City of Johannesburg, City of Tshwane, eThekwini, uMgungundlovu, and Zululand (A) overall and (B) by HIV VL findings. OVL: overdue viral load.

Variables Associated With an OVL Test

Multivariate logistic regression analysis was performed to determine variables associated with an OVL test. Only patients with 2 or more HIV VL tests were included in the model to control for limitations associated with deduplication of the test-level data (details of patients excluded from the model are provided in Multimedia Appendix 1). In total, 16,648 patients were included in the model, of whom 1554 (9.33%) had an OVL test and 15,094 (90.67%) had a follow-up VL test within 18 months of their last VL test. The 7 categorical variables included in the model, owing to their associated univariate Pearson chi-square test $P$ value of <.15, were age, current VL finding, any single high VL ($\geq 1000$ copies/mL excluding current results), 2 consecutively determined high VLs, number of previous tests, district type, and facility type (Table 1).

The Hosmer-Lemeshow goodness-of-fit $P$ value for the model was .90. On multivariate analysis, 2 consecutive (prior and current) HIV VL test results of $\geq 1000$ copies/mL were associated with a significantly increased odds of having an OVL test. Conversely, the odds of having an OVL test were significantly low among patients examined at a hospital (rather than a clinic), those with $\geq 2$ previous tests (compared to those with <2 tests), those with a history of a prior high HIV VL finding (excluding the current findings), those examined in a rural district, and those with an older age. A current VL finding of $\geq 1000$ copies/mL was associated with increased odds but was not significant in the final model (Figure 5 and Table 2).
Table 1. Variables used in multivariate logistic regression analysis to evaluate overdue HIV viral load testing among children and adolescents aged 1-15 years having undergone an HIV viral load test between May 2019 and April 2020 from the City of Johannesburg, City of Tshwane, eThekwini, uMgungundlovu, and Zululand.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall, n (%)</th>
<th>Overdue HIV viral load test, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No (n=15,094)</td>
<td>Yes (n=1554)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4</td>
<td>2058 (12.36)</td>
<td>314 (15.26)</td>
</tr>
<tr>
<td>5-9</td>
<td>4771 (28.66)</td>
<td>428 (8.97)</td>
</tr>
<tr>
<td>10-14</td>
<td>9007 (58.98)</td>
<td>812 (8.27)</td>
</tr>
<tr>
<td>Facility type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinic</td>
<td>9251 (55.57)</td>
<td>923 (9.98)</td>
</tr>
<tr>
<td>Hospital</td>
<td>7397 (44.43)</td>
<td>631 (8.53)</td>
</tr>
<tr>
<td>Facility district</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>12,035 (80.37)</td>
<td>1345 (10.05)</td>
</tr>
<tr>
<td>Rural</td>
<td>3059 (19.63)</td>
<td>209 (6.40)</td>
</tr>
<tr>
<td>Current HIV viral load test outcome (RNA copies/mL of plasma)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1000</td>
<td>12,308 (81.17)</td>
<td>1205 (8.92)</td>
</tr>
<tr>
<td>≥1000</td>
<td>2786 (18.83)</td>
<td>349 (11.13)</td>
</tr>
<tr>
<td>Any high HIV viral load test result</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False</td>
<td>10,436 (68.96)</td>
<td>1045 (9.10)</td>
</tr>
<tr>
<td>True</td>
<td>4658 (31.04)</td>
<td>509 (8.85)</td>
</tr>
<tr>
<td>Number of previous HIV viral load tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>5886 (89.03)</td>
<td>725 (10.97)</td>
</tr>
<tr>
<td>≥2</td>
<td>9208 (91.74)</td>
<td>829 (8.26)</td>
</tr>
<tr>
<td>Consecutive high HIV viral load test results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False</td>
<td>13,773 (91.27)</td>
<td>1318 (8.73)</td>
</tr>
<tr>
<td>True</td>
<td>1321 (84.84)</td>
<td>236 (15.16)</td>
</tr>
</tbody>
</table>
Figure 5. Multivariate logistic regression odds ratios to evaluate variables associated with overdue HIV viral load (VL) testing among children and adolescents aged 1-15 years having undergone an HIV VL test between May 2019 and April 2020 from the City of Johannesburg, City of Tshwane, eThekwini, uMgungundlovu, and Zululand. cps/ml: RNA copies/mL of blood.
Table 2. Results of the univariate Pearson chi-square Test and multivariate logistic regression analysis of variables associated with an overdue HIV viral load test among children and adolescents aged 1-15 years who underwent an HIV viral load test between May 2019 and April 2020 from the City of Johannesburg, City of Tshwane, eThekwini, uMgungundlovu, and Zululand.

<table>
<thead>
<tr>
<th>Variables</th>
<th>P value (univariate Pearson chi-square test)</th>
<th>Outcomes of multivariate logistic regression analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P value</td>
<td>Adjusted odds ratio</td>
</tr>
<tr>
<td>Age (years; reference: 1-4 years)</td>
<td>&lt;.001</td>
<td>—</td>
</tr>
<tr>
<td>5-9</td>
<td>&lt;.001</td>
<td>0.47-0.65</td>
</tr>
<tr>
<td>10-14</td>
<td>&lt;.001</td>
<td>0.44-0.59</td>
</tr>
<tr>
<td>Facility type (reference: clinic)</td>
<td>.002</td>
<td>0.86</td>
</tr>
<tr>
<td>Testing at a hospital</td>
<td>&lt;.001</td>
<td>0.63</td>
</tr>
<tr>
<td>Facility district (reference: urban)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>&lt;.001</td>
<td>0.63</td>
</tr>
<tr>
<td>Current HIV viral load test result (RNA copies/mL of plasma; reference: &lt;1000 RNA copies/mL of plasma)</td>
<td>&lt;.001</td>
<td>1.17</td>
</tr>
<tr>
<td>≥1000</td>
<td>.13</td>
<td>0.72</td>
</tr>
<tr>
<td>Any high HIV viral load test result (reference: false)</td>
<td>.002</td>
<td>0.78</td>
</tr>
<tr>
<td>Number of previous HIV viral load tests (reference: &lt;2)</td>
<td>&lt;.001</td>
<td>2.07</td>
</tr>
<tr>
<td>≥2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>

aNot applicable.

Discussion

Principal Results

In this analysis, which included over 20,000 children and adolescents living with HIV who are aged 1-15 years from 5 districts in South Africa, we found that over one-fourth of individuals had no repeat HIV VL test within 18 months from their previous test, suggesting considerable attrition within the pediatric population. Furthermore, among patients who underwent follow-up testing, delays in repeat testing were found among those with an elevated VL of >1000 copies/mL, similar to delays described for adult cohorts in South Africa [16]. As viraemia is associated with both patient morbidity and the emergence of drug-resistant HIV strains, these findings highlight the need to improve VL monitoring and quality of care among children and adolescents within the ART program.

Individual variables associated with an OVL test included patients of a younger age, having more recently initiated ART (as determined by <2 prior HIV VL tests), having a repeatedly high HIV VL of >1000 copies/mL, being tested in an urban district rather than a rural district, and being tested at a clinic rather than a hospital. Factors accounting for higher rates of OVL in urban districts and clinics require further investigation but may be related to shifting patient populations (due to migration to or use of multiple clinics in urban areas) and access to multidisciplinary teams, respectively. Importantly, a consecutive high HIV VL finding of >1000 copies/mL was found to be a key risk factor for having an OVL test, whereas a single high VL result was protective, suggesting that patients initially identified with viraemia are successfully returned to care but if they remain unsuppressed (ie, have virological failure) they are at a high risk of subsequently disengaging with care.

Among those children and adolescent patients who had undergone additional VL testing, a fair amount of mobility was observed. In total, 13.33% of patients were followed up at a different facility, 3.79% were followed up in a different district, and 1.71% were followed up in a different province. As the time to follow-up for this analysis was restricted to only 18-30 months per patient, it is likely a far higher proportion of mobility can be expected over the course of a child’s treatment history. This highlights limitations with South Africa’s TIER.Net HIV treatment database, which uses facility-level monitoring and hence necessarily underestimates retention of HIV care [17-19]. Furthermore, the extent to which children who have initiated ART in South Africa cross national borders and resume HIV care outside of their countries has yet to be described, as well as the number of deaths among this cohort (on account of restricted access to the national death register).

One of the notable findings of this analysis was the impact that additional manual deduplication of an already automated deduplicated data set had on programmatic outcomes. As South Africa has yet to effectively implement a national UPI, automated record linking algorithms are applied to laboratory data in an attempt to deduplicate test-level data sets to represent patient-level data. However, our findings suggest that without additional manual deduplication, analysis of public health laboratory data using the NHLS CDW algorithm would have
overestimated the proportion of patients aged 1-15 years with an OVL by at least 40% (even though the data set was reduced by only an additional 6% after manual deduplication). This highlights the need to incorporate a national health UPI within the health system, including the laboratory information system and NHLS Data Warehouses, to effectively monitor the ART program.

Limitations
A number of important limitations need to be considered regarding these findings. The NHLS CDW data set was manually deduplicated at the facility level (for the 152 facilities evaluated in this analysis) without the opportunity to compare findings against a gold-standard data set. Although we report that 13% of patients with a repeat HIV VL test were followed up at a different facility, this is likely an underestimation on account of the reduced ability to accurately merge the test results of patients who were examined at multiple facilities. This may have been a contributing factor to the reason why the urban regions were associated with a higher rate of OVL testing, if patients examined in the urban areas are more likely to be followed up at multiple facilities within or even outside of the district. Additionally, patients may be retained in care but not have repeat HIV VL tests in accordance with the guidelines, for instance, on account of challenges with accessing pediatric phlebotomy services [20]; hence, the OVL testing rate may overestimate program attrition. This is likely to have been further exacerbated during the course of the COVID-19 pandemic, which impacted maternal and child access to HIV services [21]. As the multivariate analysis evaluating variables associated with OVL testing was restricted to patients with ≥2 VL tests (to ensure goodness-of-fit of the model by increasing the likelihood that if additional testing was performed, it would be linked via the deduplication methodology), this may have also influenced the findings by not including patients who were only followed-up for a single HIV VL test and hence disengaged with care soon after ART initiation. Lastly, as our analysis was restricted to 30% of facilities within 5 of the 52 national districts, the results may not be nationally representative, especially considering that the districts included in this study are known to have a high pediatric HIV disease burden.

Conclusions
In summary, linking population clinical data with laboratory data in a legal manner, which protects individuals’ rights, has the potential to improve patient outcomes for multiple health programs. A functional UPI for health care would enable improved longitudinal surveillance of children and adolescents living with HIV by using routine laboratory data for accurate, timely identification of patients who need to be returned to care, including infants. In the interim, 1-4-year-olds and those with a sustained VL of ≥1000 copies/mL should be managed as especially high risk for loss to follow-up.

Acknowledgments
The authors gratefully acknowledge the National Health Laboratory Services for access to laboratory data and Dr Trevor Graham Bell, principal health data analyst at the National Institute for Communicable Diseases, for assistance with data extraction and processing. The authors also wish to thank Mr Dumsani Mlotshwa, Dr Tanya Murray, and Dr Faith Moyo for their assistance with manual deduplication of facility patient lists. This work was partly funded by the ELMA Foundation. The funders played no role in the design of the study, data collection and analysis, or in the production of this manuscript.

Data Availability
The data sets analyzed in this study are not publicly available because they contain information that could compromise participant privacy, including patient first and last names, gender, locations, and tests conducted. Please contact the South African National Health Laboratory Service [22] for data requests.

Authors’ Contributions
GGS and AHM conceptualized the study, and LR analyzed the data. All authors contributed to the interpretation of the data. AHM played a major role in drafting the manuscript, and all the authors read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Manual deduplication rules.

References


13. Haeri Mazanderani et alJMIR PUBLIC HEALTH AND SURVEILLANCE


Examining HIV Testing Coverage and Factors Influencing First-Time Testing Among Men Who Have Sex With Men in Zhejiang Province, China: Cross-Sectional Study Based on a Large Internet Survey

Lin He¹, MS; Tingting Jiang¹, MS; Wanjun Chen¹, MS; Shaoqiang Jiang², BS; Jinlei Zheng¹, MPH; Weiyong Chen¹, BS; Hui Wang¹, PhD; Qiaoqin Ma¹, BS; Chengliang Chai¹, MPH

¹Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
²Coastal Service Center, Hangzhou, China

Corresponding Author:
Chengliang Chai, MPH
Zhejiang Provincial Center for Disease Control and Prevention
3399 Bin Sheng Road, Binjiang District
Hangzhou, 310051
China
Phone: 86 57187115193
Fax: 86 57187115194
Email: chlchai@cdc.zj.cn

Abstract

Background: Men who have sex with men (MSM) constitute a significant population of patients infected with HIV. In recent years, several efforts have been made to promote HIV testing among MSM in China.

Objective: This study aimed to assess HIV testing coverage and factors associated with first-time HIV testing among MSM to provide a scientific basis for achieving the goal of diagnosing 95% of patients infected with HIV by 2030.

Methods: This cross-sectional study was conducted between July 2023 and December 2023. MSM were recruited from the “Sunshine Test,” an internet platform that uses location-based services to offer free HIV testing services to MSM by visiting the WeChat official account in Zhejiang Province, China. Participants were required to complete a questionnaire on their demographic characteristics, sexual behaviors, substance use, and HIV testing history. A logistic regression model was used to analyze first-time HIV testing and its associated factors.

Results: A total of 7629 MSM participated in the study, with 87.1% (6647) having undergone HIV testing before and 12.9% (982) undergoing HIV testing for the first time. Multivariate logistic regression analysis revealed that first-time HIV testing was associated with younger age (adjusted odds ratio [aOR] 2.55, 95% CI 1.91-3.42), lower education (aOR 1.39, 95% CI 1.03-1.88), student status (aOR 1.35, 95% CI 1.04-1.75), low income (aOR 1.55, 95% CI 1.16-2.08), insertive anal sex role (aOR 1.28, 95% CI 1.05-1.56), bisexuality (aOR 1.69, 95% CI 1.40-2.03), fewer sex partners (aOR 1.44, 95% CI 1.13-1.83), use of rush poppers (aOR 2.06, 95% CI 1.70-2.49), unknown HIV status of sex partners (aOR 1.40, 95% CI 1.17-1.69), lack of awareness of HIV pre-exposure prophylaxis (aOR 1.39, 95% CI 1.03-1.88), and offline HIV testing uptake (aOR 2.08, 95% CI 1.80-2.41).

Conclusions: A notable 12.9% (982/7629) of MSM had never undergone HIV testing before this large internet survey. We recommend enhancing HIV intervention and testing through internet-based platforms and gay apps to promote testing among MSM and achieve the target of diagnosing 95% of patients infected with HIV by 2030.

(Keywords: HIV; testing; men who have sex with men; MSM; internet; pre-exposure prophylaxis; China; mobile phone)

(JMIR Public Health Surveill 2024;10:e56906) doi:10.2196/56906
Introduction

Men who have sex with men (MSM) constitute a significant demographic of patients infected with HIV in China. Due to widespread unprotected sexual behaviors, engagement with multiple sexual partners, and the use of recreational agents, HIV prevalence among MSM remains notably high [1]. The overall national prevalence of HIV among MSM in China from 2001 to 2018 was estimated to be 5.7% [2]. In China, the annual number of newly diagnosed HIV infections through homosexual transmission has increased from 2.5% in 2006 to 25.6% in 2022 [3]. The Joint United Nations Program on HIV/AIDS (UNAIDS) proposed a 95-95-95 target by 2030 [4]. The overall aim of the target was to achieve diagnosis of 95% of patients infected with HIV, 95% of diagnosed patients treated, and 95% of treated patients having achieved viral suppression to reduce HIV transmission, and ultimately end the HIV epidemic worldwide.

In China, with the implementation of the “Four Frees and One Care” policy and lifelong follow-up care, achieving the second and third 95% targets is feasible. In 2022, a total of 92.8% of patients diagnosed with HIV received antiretroviral therapy, and 97% of treated patients achieved viral suppression (defined as HIV viral load <1000 copies/mL) in China [3]. The national estimate is that the proportion of patients infected with HIV was diagnosed at only 84% in 2022, and achieving a 95% diagnosis target for patients infected with HIV may be a huge challenge in China by 2030.

The “HIV testing as prevention” strategy stands as an effective approach to mitigate HIV transmission by encouraging early diagnosis of patients infected with the disease and linking them to antiretroviral therapy, thus reducing disease transmission [5,6]. There are 3 main HIV testing strategies among MSM in China. The first strategy is scaling up testing, expanding HIV testing coverage, and increasing testing frequency to prevent a substantial number of new HIV infections among MSM [7,8]. The second strategy is regular testing, which is a key prevention strategy for identifying and treating HIV infections among MSM [9,10]. The third strategy is HIV self-testing [11], which recommends self-testing as an innovative strategy and an additional testing approach to attain the UNAIDS targets to end HIV by 2030. With the implementation of the HIV testing strategy, annual HIV testing among MSM has reached nearly 600,000 in China [3], and an increasing number of MSM have received HIV interventions and are undergoing testing. To achieve the 95% diagnosis target for patients infected with HIV, promoting HIV testing among MSM is very important.

Previous studies found that younger individuals and students had lower HIV testing rates [9,12,13]. A higher HIV testing rate was observed in the developed country compared with the developing country. A meta-analysis study from the United States stated that only 15% of the MSM reported having first-time HIV testing [9], while the first-time HIV testing rate was 50.5% in Brazil [14] and 30.6% in Malaysia [15]. The first-time HIV testing rate in China was similar to other developing countries. A study showed that 30.1% and 29.4% of MSM underwent first-time HIV testing in Tianjin and Zhuhai, China, respectively [16]. A study on a large Chinese gay social media platform showed that 19.5% of MSM had never been tested for HIV [17]. An internet-distributed HIV self-testing study revealed that 16.6% of MSM had never been tested for HIV [18]. Some factors that have been associated with first-time HIV testing include Han ethnicity, HIV status, and income [19]. With the increasing popularity of the internet and smartphone usage in China, the internet penetration rate was expected to reach 76.4% in 2023 [20]. MSM who sought potential sexual partners from traditional places such as bars, parks, and baths, gradually transferred to internet-based dating apps, such as Blued app (BlueCity Group; a large gay Chinese social media platform). Consequently, the prevalence of gay app use among MSM has rapidly increased from 12.5% in 2015 to 52.6% in 2017 [21], suggesting that most MSM rely on gay apps to find sexual partners [22], internet-based HIV interventions and testing [23], as well as gay apps [24], are considered simple ways of promoting HIV testing among MSM in China and increasing awareness of HIV status [25].

Zhejiang Province is located in eastern China, with an economically developed and prosperous internet economy. There are many companies that do e-commerce internet-based platforms, such as the Alibaba Group which was founded in Zhejiang Province and has had an HIV prevalence of 8% among MSM since 2008 [1]. Also, more than 40% of the annually newly diagnosed patients infected with HIV were MSM. Based on the developed internet economy, Zhejiang Province relies on the internet to conduct HIV intervention and testing among MSM, including internet-based consulting and mailing HIV testing kits for self-testing and using internet-based technology could increase the HIV self-testing rate among MSM [26]. Approximately 62,000 HIV testing cases occurred among MSM in Zhejiang Province in 2022, accounting for 10.6% of the national proportion [3], with over 70% of HIV testing in the province relying on the internet. Internet surveys offer broader coverage of the MSM population, rendering them more representative [27]. Identifying patients infected with HIV from the internet is key to diagnosing 95% of the infections by 2030. Therefore, this study aimed to assess HIV testing coverage based on a large internet survey and determine the factors associated with first-time HIV testing uptake among MSM in Zhejiang Province, China.

Methods

Sampling

A cross-sectional study based on a large internet survey was conducted between July 2023 and December 2023, targeting the MSM population.

Study Participants

MSM participants were recruited based on inclusion criteria, which are (1) a minimum age of 16 years, (2) reported having sex with men in the past year, and (3) living in Zhejiang Province.

Participant Recruitment and Data Collection

Sunshine Coast Public Welfare, a social organization, uses the internet to provide HIV interventions and testing services for the MSM population. The organization comprises 13 full-time social workers, 22 part-time social workers, and over 400

https://publichealth.jmir.org/2024/1/e56906
registered volunteers. Sunshine Coastal Public Welfare has established a digital HIV prevention service using location-based services. MSM populations in Zhejiang Province can apply for free HIV testing services by visiting the WeChat (Tencent Holdings Ltd, a very popular communication software in China) official account, “Sunshine Test.” Two options are available for MSM to access HIV testing services. First, they can opt to have the testing reagent mailed to them through courier and perform self-testing after receiving the reagent. Alternatively, they can choose offline services, where volunteers in their neighborhood provide HIV testing services, or they can visit nearby voluntary counseling and testing sites for MSM. During the study period, MSM had repeated HIV testing (more than 2 times); only the first testing record was retained in the study.

**Questionnaires**

MSM who are applying for HIV testing services through “Sunshine Test” (a website by visiting WeChat in China) are required to complete a routine surveillance questionnaire. The questionnaire consists of 20 questions focusing on demographic characteristics such as age, marital status, education level, and sexual behavior characteristics such as sexual roles, number of sexual partners, whether to use rush poppers, sexual history, pre-exposure prophylaxis (PrEP), post-exposure prophylaxis (PEP), and HIV testing history. All questionnaires were completed on the “Sunshine Test.” MSM can apply for testing services after completing the questionnaire and being screened by a social worker or volunteer.

**Statistical Analysis**

For descriptive analyses, categorical variables are presented as frequencies and proportions, while continuous variables are presented as medians and IQRs or means and SDs. The significance of the difference in the general demographic characteristics was tested by a chi-square test. Factors from the univariate analysis with \( P<0.10 \) or those previously shown to be associated with the differences in sociodemographic characteristics among first-time HIV testing were included in multivariate logistic regression models, and adjusted odd ratios (aORs) were calculated along with corresponding 95% CIs. The statistical significance level was \( P\leq0.05 \) and \( \beta=1 \). Statistical analyses were performed using SPSS (version 19.0; IBM Corp.).

**Ethical Considerations**

This study was approved by the Zhejiang Provincial Center for Disease Control and Prevention (2022-011-01). All participants provided written informed consent before the completion of the survey. Participants received free HIV testing and health counseling, and all study procedures were conducted in accordance with the approved guidelines and regulations.

**Results**

A total of 7629 MSM participated in the study (Figure 1), with an average age of 29.0 (SD 8.5) years. Among them, 78.6% (5997/7629) were single, 50.7% (3869/7629) had a bachelor’s degree or higher, 16.4% (1254/7629) were students, and 49.1% (2985/7629) had a monthly income of less than US $700. The majority of 78.2% (5964/7629) had only male sexual partners, while 33.3% (2543/7629) had more than 2 sex partners in the past 3 months. Additionally, 31.2% (2380/7629) reported using rush poppers during sexual activity and 45.2% (3451/7629) were aware of their sex partner’s HIV status. The awareness rates of PrEP and PEP were 76.2% (5816/7629) and 78.1% (5960/7629), respectively. Furthermore, 54.8% (4183/7629) had HIV testing through mail reagent self-testing.

Of the 7629 participants, 12.9% (982/7629) underwent HIV testing for the first time. Factors associated with a high proportion of first-time HIV testing included being younger than 20 years old (30.9%, 133/430), having high school and below education (16.2%, 299/1851), being a student (21.6%, 271/1254), having a low monthly income (21.4%, 282/1320), identifying as bisexual (18.2%, 303/1665), having a small number of sex partners in the past 3 months (17.5%, 224/1277), frequently using rush poppers (18.5%, 198/1068), being unaware of their sex partners’ HIV status (18.6%, 275/1477), lacking awareness of HIV PrEP (18.9%, 342/1813), undergoing offline HIV testing (17.8%, 614/3446), and being HIV positive (24.4%, 20/82). The chi-square test showed that all factors except marital status, sex role, and history of sexually transmitted infections were associated with a high proportion of first-time HIV testing (Table 1).

Multivariate logistic regression analysis revealed that first-time HIV testing was associated with being younger than 20 years old (aOR 2.55, 95% CI 1.91-3.42), having attained high school education or lower (aOR 1.39, 95% CI 1.03-1.88), being a student (aOR 1.35, 95% CI 1.04-1.75), having a monthly income of less than US $330 (aOR 1.55, 95% CI 1.16-2.08), insertive anal sex role (aOR 1.28, 95% CI 1.05-1.56), identifying as bisexual (aOR 1.69, 95% CI 1.40-2.03), having a low number of sex partners in the past 3 months (aOR 1.44, 95% CI 1.13-1.83), frequently using rush poppers in the past 3 months (aOR 2.06, 95% CI 1.70-2.49), being unaware of their sex partners’ HIV status (aOR 1.40, 95% CI 1.17-1.69), lacking awareness of HIV PrEP (aOR 1.39, 95% CI 1.03-1.88), and undergoing offline HIV testing (aOR 2.08, 95% CI 1.80-2.41; Table 2).
Figure 1. A cross-sectional study based on a large internet survey by visiting the official account of WeChat (a very popular communication software in China), “Sunshine Test.” A total of 7629 MSM participated in the study between July and December 2023 in Zhejiang Province, China. MSM: men who have sex with men.

The “Sunshine Test” recruited 9568 MSM from July to December 2023

Excluded (n=1939)
• HIV testing ≥2 times

First testing record of 7629 MSM was retained

45.2% (3446) through offline testing
54.8% (4183) through mail reagent self-testing

Total 7629 in the analysis
12.9% (982) underwent first-time HIV testing
Table 1. Social demographic and behavioral characteristics among men who have sex with men.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values (N=7692), n (%)</th>
<th>First-time HIV testing (n=982), n (%)</th>
<th>chi-square (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td>138.7 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;20</td>
<td>430 (5.6)</td>
<td>133 (30.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>4526 (59.3)</td>
<td>568 (12.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥30</td>
<td>2673 (35)</td>
<td>281 (10.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td>3.6 (2)</td>
<td>.16</td>
</tr>
<tr>
<td>Single</td>
<td>5997 (78.6)</td>
<td>792 (13.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1388 (18.2)</td>
<td>166 (12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>244 (3.2)</td>
<td>24 (9.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td>33.5 (3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High school and below</td>
<td>1851 (24.3)</td>
<td>299 (16.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>1909 (25)</td>
<td>262 (13.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>3145 (41.2)</td>
<td>335 (10.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>724 (9.5)</td>
<td>86 (11.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td>119.3 (3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Student</td>
<td>1254 (16.4)</td>
<td>271 (21.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Company employee</td>
<td>2843 (37.3)</td>
<td>265 (9.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freelance</td>
<td>1022 (13.4)</td>
<td>117 (11.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>2510 (32.9)</td>
<td>329 (13.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly income (US $)</td>
<td></td>
<td></td>
<td>142.1 (3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;350</td>
<td>1320 (17.3)</td>
<td>282 (21.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>350-599</td>
<td>1665 (21.8)</td>
<td>256 (15.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>700-1399</td>
<td>3259 (42.7)</td>
<td>328 (10.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥1400</td>
<td>1385 (18.2)</td>
<td>116 (8.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex roles</td>
<td></td>
<td></td>
<td>5.4 (2)</td>
<td>.07</td>
</tr>
<tr>
<td>Receptive anal sex</td>
<td>1750 (22.9)</td>
<td>197 (11.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insertive anal sex</td>
<td>2652 (34.8)</td>
<td>349 (13.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>3227 (42.3)</td>
<td>436 (13.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex of partner</td>
<td></td>
<td></td>
<td>53.9 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>5964 (78.2)</td>
<td>679 (11.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male and female</td>
<td>1665 (21.8)</td>
<td>303 (18.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sexual partners in the past 3 months</td>
<td></td>
<td></td>
<td>68.6 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>0</td>
<td>1277 (16.7)</td>
<td>224 (17.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3809 (49.9)</td>
<td>536 (14.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥2</td>
<td>2543 (33.3)</td>
<td>222 (8.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever used rush poppers during sex behavior</td>
<td></td>
<td></td>
<td>56.6 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never use</td>
<td>5249 (68.8)</td>
<td>677 (12.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occasional use</td>
<td>1312 (17.2)</td>
<td>107 (8.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Often use</td>
<td>1068 (14)</td>
<td>198 (18.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness of partner’s HIV status</td>
<td></td>
<td></td>
<td>59.4 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>3451 (45.2)</td>
<td>427 (12.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part</td>
<td>2701 (35.4)</td>
<td>280 (10.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1477 (19.4)</td>
<td>275 (18.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>Values (N=7692), n (%)</td>
<td>First-time HIV testing (n=982), n (%)</td>
<td>chi-square (df)</td>
<td>P value</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------</td>
<td>--------------------------------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>History of STI(^b)</td>
<td></td>
<td></td>
<td>&lt;0.01 (1)</td>
<td>.98</td>
</tr>
<tr>
<td>No</td>
<td>7428 (97.4)</td>
<td>956 (12.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>201 (2.6)</td>
<td>26 (12.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness of HIV PrEP(^c)</td>
<td></td>
<td></td>
<td>76.1 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>1813 (23.8)</td>
<td>342 (18.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5816 (76.2)</td>
<td>640 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge of HIV PEP(^d)</td>
<td></td>
<td></td>
<td>45.1 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>1669 (21.9)</td>
<td>296 (17.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5960 (78.1)</td>
<td>686 (11.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV testing pathway</td>
<td></td>
<td></td>
<td>137.1 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Offline testing</td>
<td>3446 (45.2)</td>
<td>614 (17.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mail reagent self-testing</td>
<td>4183 (54.8)</td>
<td>368 (8.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV status</td>
<td></td>
<td></td>
<td>9.8 (1)</td>
<td>&lt;.002</td>
</tr>
<tr>
<td>Positive</td>
<td>82 (1.1)</td>
<td>20 (24.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>7547 (98.9)</td>
<td>962 (12.7)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Median 27 (IQR 23-33).
\(^b\)STI: sexually transmitted infections.
\(^c\)PrEP: pre-exposure prophylaxis.
\(^d\)PEP: post-exposure prophylaxis.
## Table 2. Factors associated with first-time HIV testing among men who have sex with men.

<table>
<thead>
<tr>
<th>Variables</th>
<th>First-time HIV testing (n=982), n (%)</th>
<th>OR² (95% CI) P value</th>
<th>aOR³ (95% CI) P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years; reference: ≥30)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>133 (30.9)</td>
<td>3.81 (3.00-4.84)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>20-29</td>
<td>568 (12.5)</td>
<td>1.22 (1.05-1.42)</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Education (reference: master’s degree or above)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school and below</td>
<td>299 (16.2)</td>
<td>1.43 (1.11-1.85)</td>
<td>&lt;.006</td>
</tr>
<tr>
<td>College</td>
<td>262 (13.7)</td>
<td>1.18 (0.91-1.53)</td>
<td>.21</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>335 (10.7)</td>
<td>0.88 (0.69-1.14)</td>
<td>.34</td>
</tr>
<tr>
<td><strong>Occupation (reference: others)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>271 (21.6)</td>
<td>1.83 (1.53-2.18)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Company employee</td>
<td>265 (9.3)</td>
<td>0.68 (0.57-0.81)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Freelance</td>
<td>117 (11.4)</td>
<td>0.86 (0.69-1.07)</td>
<td>.18</td>
</tr>
<tr>
<td><strong>Monthly income (US $; reference: ≥US $1400)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;$350</td>
<td>282 (21.4)</td>
<td>2.97 (2.36-3.74)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>$350-$599</td>
<td>256 (15.4)</td>
<td>1.99 (1.58-2.51)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>$700-$1399</td>
<td>328 (10.1)</td>
<td>1.22 (0.98-1.53)</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Sex roles (reference: receptive anal sex)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insertive anal sex</td>
<td>349 (13.2)</td>
<td>1.20 (1.00-1.44)</td>
<td>.06</td>
</tr>
<tr>
<td>Both</td>
<td>436 (13.5)</td>
<td>1.23 (1.03-1.47)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Sex of partner (reference: male and female)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male and female</td>
<td>303 (18.2)</td>
<td>1.73 (1.49-2.00)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Number of sex partners in the past 3 months (reference: ≥2)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>224 (17.5)</td>
<td>2.22 (1.82-2.71)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1</td>
<td>536 (14.1)</td>
<td>1.71 (1.45-2.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Ever used rush poppers (reference: never use)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occasional use</td>
<td>107 (8.2)</td>
<td>0.60 (0.48-0.74)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Often use</td>
<td>198 (18.5)</td>
<td>1.54 (1.29-1.83)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Awareness of partner’s HIV status (reference: yes)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part</td>
<td>280 (10.4)</td>
<td>0.82 (0.70-0.96)</td>
<td>.01</td>
</tr>
<tr>
<td>No</td>
<td>275 (18.6)</td>
<td>1.62 (1.37-1.91)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Awareness of HIV PrEPc (reference: no)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>342 (18.9)</td>
<td>1.88 (1.63-2.17)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>HIV testing pathway (reference: offline testing)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offline testing</td>
<td>614 (17.8)</td>
<td>2.25 (1.96-2.58)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Or: odds ratio.  
AOR: adjusted odds ratio.  

### Discussion

**Principal Findings**

This study described HIV testing coverage based on a large internet survey and the factors associated with first-time HIV testing uptake among MSM in Zhejiang Province, China. Based on data from 7629 participants, we found that 87.1% (6647/7629) of MSM had undergone HIV testing, and 12.9% (982/7629) had never been tested before the survey. This 87.1% HIV testing coverage marks a significant increase from the 55.9% reported a decade ago (2013-2014) in Zhejiang Province [28]. Comparative analysis revealed that the HIV testing rate of 87.1% aligned with the 85% reported in a US meta-analysis...
study [9], surpassing rates found in Tianjin [16] and Zhuhai (70.6%) [19], as well as an internet-based study based on gay apps (80.5%) [17]. These findings underscore the need for continued efforts to expand HIV testing and increase testing coverage among MSM to achieve the target of diagnosing 95% of patients infected with HIV by 2030.

Consistent with previous studies, we found that younger individuals and students had lower HIV testing rates [9,12,13]. In China, nearly 3000 patients infected with HIV are diagnosed annually, more than 80% of whom are MSM [3]. An internet-based study found that younger MSM were probably deterred from seeking testing due to apprehension about health care or local office settings [13]. HIV molecular transmission clusters found students and nonstudents with HIV infections in the same transmission clusters [29]. However, the nonstudent MSM had an HIV prevalence of nearly 8%. Accordingly, a lack of awareness of the HIV epidemic in students among MSM may result in a low HIV testing rate. The internet and smartphones are essential tools for the student population, and a previous study showed that HIV intervention and testing through internet-based platforms or gay apps could increase HIV testing among younger MSM [24,30,31]. We suggest that strengthening awareness of the HIV epidemic among students and enhancing HIV testing education through internet-based platforms or gay apps will increase the HIV testing proportion of MSM students.

We found that participants with fewer sexual partners had low HIV testing rates. A study in China showed that individuals with 2 or more sexual partners were more likely to undergo HIV testing [32]. MSM with multiple sexual partners usually exhibit high-risk behaviors that prompt them to undergo HIV testing. Similarly, MSM who often used rush poppers had a higher HIV testing rate. Rush poppers are widely popular among MSM and have become the main substance type used during sexual activities [33]. Rush poppers prolong sexual activity and increase the rate of unprotected sex, thereby increasing their risk of HIV infection. Studies have shown that people who use rush poppers have more unprotected, casual sex, and increased HIV testing rates [9,34].

Moreover, we found that being unaware of a partner’s HIV status was associated with lower HIV testing rates. Serostatus disclosure strategies can reduce the risk of HIV infection and promote regular HIV testing which has been gradually promoted among MSM in China in recent years [35]. The serostatus disclosure strategy refers to knowing each other’s HIV status before any sexual activity. The strategy encourages using HIV status to decide whether to have sex or not. If both partners have the same HIV infection status, do not use condoms; and if the HIV infection status is inconsistent, use condoms [36]. Our study also showed that those who knew their sexual partners’ HIV status had a higher HIV testing rate. Therefore, it is recommended to continue implementing this strategy among MSM to increase HIV testing rates.

Participants who were aware of PrEP had higher rates of HIV testing. HIV PrEP and PEP are biological interventions that prevent HIV transmission, providing effective prevention opportunities for MSM at a high risk of HIV exposure, and can effectively reduce HIV infection and transmission [37]. In China, PEP guidelines were released by the National Center for Communicable Diseases Control in 2020. PEP has been widely promoted among MSM after 3 years of promotion. However, PrEP has only been piloted in some cities, and there was no national guideline for PrEP [3]. Therefore, in the results of this study, only those who had an awareness of PEP had a higher rate of HIV testing. Several studies have shown that unprotected sex and sexually transmitted infections continue to increase among MSM who have used PrEP drugs [38]. Previous studies have also shown that the PrEP project effectively promoted HIV testing among MSM [39], and greater odds of PrEP awareness were associated with HIV testing [40].

This study had some limitations. First, the study recruited MSM who used web-based apps for HIV testing, 88.5% (6753/7629) of whom were younger than 40 years. It did not capture a few MSM who came from traditional places such as bars and baths, especially those who were older than 50 years. Previous studies have shown that this population has a higher HIV infection rate and a lower HIV testing rate [32]. Therefore, the results of this study apply only to MSM who rely on the internet for HIV testing. Second, participants who opted for HIV self-testing through mailed reagents may introduce bias, as the feedback on true HIV-positive results could be limited. Finally, the use of routine surveillance questionnaires may not capture all factors associated with HIV testing as identified in previous studies.

Conclusions
A significant finding from the large internet survey was that 12.9% (982/7629) of MSM had never undergone HIV testing before the study. We suggest enhancing HIV intervention and testing through internet-based platforms and gay apps to promote HIV testing among MSM. Future studies could explore strategies to increase HIV testing rates among underrepresented demographics, such as older MSM and those who do not rely on web-based testing services, ultimately helping to meet the target of diagnosing 95% of patients infected with HIV by 2030.

Acknowledgments
We express our gratitude to the participants and Sunshine Coast Public Welfare for their contributions to the study. The study was funded by a grant from the Health Science and Technology Project in Zhejiang Province Social (2023KY632 and 2021RC048).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.
References


4. UNAIDS. Understanding fast-track: accelerating action to end the AIDS epidemic by 2030. URL: https://www.unaids.org/ [accessed 2015-07-07]


Abbreviations

- **aOR**: adjusted odds ratio
- **MSM**: men who have sex with men
- **PEP**: post-exposure prophylaxis
- **PrEP**: pre-exposure prophylaxis
- **UNAIDS**: Joint United Nations Program on HIV/AIDS

©Lin He, Tingting Jiang, Wanjun Chen, Shaoqiang Jiang, Jinlei Zheng, Weiyong Chen, Hui Wang, Qiaoqin Ma, Chengliang Chai. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 14.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Factors Associated With Surveillance Testing in Individuals With COVID-19 Symptoms During the Last Leg of the Pandemic: Multivariable Regression Analysis

Timothy Dotson¹, MS; Brad Price¹,², PhD; Brian Witrick¹,³, PhD; Sherri Davis¹, MA; Emily Kemper¹, MS; Stacey Whanger⁴, MPH; Sally Hodder¹,⁵, PhD; Brian Hendricks¹,⁶, PhD

Corresponding Author:
Timothy Dotson, MS

Abstract

Background: Rural underserved areas facing health disparities have unequal access to health resources. By the third and fourth waves of SARS-CoV-2 infections in the United States, COVID-19 testing had reduced, with more reliance on home testing, and those seeking testing were mostly symptomatic.

Objective: This study identifies factors associated with COVID-19 testing among individuals who were symptomatic versus asymptomatic seen at a Rapid Acceleration of Diagnostics for Underserved Populations phase 2 (RADx-UP2) testing site in West Virginia.

Methods: Demographic, clinical, and behavioral factors were collected via survey from tested individuals. Logistic regression was used to identify factors associated with the presence of individuals who were symptomatic seen at testing sites. Global tests for spatial autocorrelation were conducted to examine clustering in the proportion of symptomatic to total individuals tested by zip code. Bivariate maps were created to display geographic distributions between higher proportions of tested individuals who were symptomatic and social determinants of health.

Results: Among predictors, the presence of a physical (adjusted odds ratio [aOR] 1.85, 95% CI 1.3-2.65) or mental (aOR 1.53, 95% CI 0.96-2.48) comorbid condition, challenges related to a place to stay/live (aOR 307.13, 95% CI 1.46-10,6372), no community socioeconomic distress (aOR 0.99, 95% CI 0.98-1.00), no challenges in getting needed medicine (aOR 0.01, 95% CI 0.00-0.82) or transportation (aOR 0.23, 95% CI 0.05-0.64), an interaction between community socioeconomic distress and not getting needed medicine (aOR 1.06, 95% CI 1.00-1.13), and having no community socioeconomic distress while not facing challenges related to a place to stay/live (aOR 0.93, 95% CI 0.87-0.99) were statistically associated with an individual being symptomatic at the first test visit.

Conclusions: This study addresses critical limitations to the current COVID-19 testing literature, which almost exclusively uses population-level disease screening data to inform public health responses.

(JMIR Public Health Surveill 2024;10:e52762) doi:10.2196/52762

KEYWORDS

COVID-19; testing; symptomatic; RADx; rural; health disparities; regression analysis; surveillance; SARS-CoV-2; United States; asymptomatic; survey; demographic; clinical; behavioral; logistic regression; bivariate map; child; children; youth; adolescent; adolescents; teen; teens; teenager; teenagers; public health; machine learning; mHealth; mobile health; digital health; cross-sectional study; physical health; mental health; Rapid Acceleration of Diagnostics

Introduction

Pandemic Spread

The SARS-CoV-2 pandemic has had several waves of infection driven by the introduction and spread of multiple variants. In the United States, the first case of COVID-19 appeared in January 2020 [1]. The Alpha variant, first introduced in December 2020, comprised the second wave of the COVID-19 pandemic, occurring at the same time as vaccination campaigns were being rolled out in many states [2]. This was a critical
moment in the SARS-CoV-2 timeline, as Alpha was the first variant of concern with adapted mutations, increasing the risk of person-to-person transmission [3]. The pattern of spread and speed of the third and fourth waves involving Omicron variants (BA1-3 and BA4-5) were wider reaching and faster than previous variants [4]. Unfortunately, this was problematic with reduced testing efforts making it more difficult to monitor SARS-CoV-2 infections in populations [5].

Health Disparities

Urban US areas experienced the greatest burden of cases early in the epidemic with West Virginia, a largely rural state, being the last state to identify a confirmed case of COVID-19 in late January 2020. This is problematic as 50% of rural residents are at high risk of serious illness and hospitalization if they contract SARS-CoV-2 [6]. In West Virginia, the Alpha variant cases peaked in April 2021, 4 months after its initial detection in the United States and after vaccination campaigns were already underway, with subsequent peaks in October 2021 for the Delta variant and February 2022 for the first Omicron peak. Due to unequal access to health resources, the impacts of the disease vary throughout the state, particularly in southwest West Virginia where there are already health disparities [7].

Prior Work

By August 2020, testing in the United States had peaked at around 1 million tests per day, at which point COVID-19 had become the third leading cause of death [1]. COVID-19 testing data have been used in machine learning models and spatial epidemiological studies to help identify disparities in testing and outcomes for COVID-19, and guide public health policies [8-10]. Previous studies have analyzed socioeconomic status, race, comorbidities, mental health, and substance abuse effects to identify disparities and seasonal impacts [7,9-14]. Furthermore, other epidemic models have been tested using COVID-19 data to help forecast future SARS-CoV-2 waves and look at the impact of testing itself [11,12]. However, previous research was conducted during times with high-to moderate community testing, when data more accurately reflected the general population’s risk of disease and allowed for standard epidemiological models of study [1,13-15]. The pandemic is now in a phase where those seeking testing are largely symptomatic. Lower testing and increasing reliance on home testing for COVID-19 have created a situation in which traditional epidemiological measures are suboptimal [16].

Study Objectives

Subsequent waves of SARS-CoV-2 infection are now monitored with varying surveillance efforts, which have dwindled from the population to the community level as the focus shifted away from testing to preventing severe infection and health care strain [5]. This situation has resulted in a heightened reliance on testing data from people who are symptomatic and seek out testing. Currently, few studies have examined the demographics, clinical factors, or barriers to testing among people who are symptomatic and seeking COVID-19 testing [15,17]. As such, the objectives of this study were to identify factors that increased the prevalence of individuals who were symptomatic at testing locations and assess whether there was spatial autocorrelation among the rate of tested people who were symptomatic and their residential zip code. Spatial autocorrelation, or clustering of tested people who were symptomatic, was assessed to better understand if geographic differences in hospital or doctor referrals potentially biased the number of people who were symptomatic who visited testing locations [18,19]. The results address current literature gaps concerning which factors are associated with test seeking and have the potential to inform public health policy to ensure COVID-19 testing services remain available to vulnerable populations living in the rural United States.

Methods

Data Source and Management

This cross-sectional study utilized questionnaire data collected for phase 2 of the Rapid Acceleration of Diagnostics for Underserved Populations program (RADx-UP2). Detailed information on RADx-UP2, including program aims and research projects, has been described elsewhere [20]. Briefly, RADx-UP is a multisite National Institutes of Health–funded project, developed to disseminate testing resources within communities of varying social or economic vulnerability [21]. All sites are required to include common data elements (CDEs) in their data collection instruments to harmonize data across states [21]. Project CDEs include an individual’s address, demographics, clinical comorbidities, signs and symptoms at the time of testing, behavioral data, and more [21]. For the West Virginia site, data were collected using ArcGIS Survey 123 (Esri) at testing events. The benefits of using syndromic surveillance for public health programming and response have been described in previous public health research elsewhere [16,22-24].

Ethical Considerations

Approval for this study was given by the West Virginia University Institutional Review Board (protocol 2202534378A001). Informed consent was requested on the survey used for data collection in the form of an opt-out question. All included data has been deidentified for analysis and publication. Site-level incentives were developed in conjunction with community partners that assisted in hosting testing events. As such, these incentives varied by site and community and consisted of monthly raffles for all who participated in a survey during that month for each testing site, with prizes including technology, tickets to sporting events, and outdoor items. Other incentives included t-shirts and plastic reusable cups for participation.

Inclusion Criteria

The study inclusion criteria were any individual tested at a West Virginia RADx-UP2 testing site from May 2022 until November 2022. Testing sites included pharmacies, hospitals, and homeless shelters in underserved areas throughout the state manned by RADx-UP2 staff as well as at testing events such as Solutions Oriented Addiction Response (SOAR) meetings and other community-sponsored events. Individuals seeking COVID-19 testing paid for by RADx-UP funding and consented to have their information collected. For this study, the analytic sample

https://publichealth.jmir.org/2024/1/e52762
was limited to information collected at each individual’s first test, including the individual’s demographics, signs and symptoms, history of chronic disease, receipt of two vaccine doses, and challenges or motivators to seek care. The study outcome was the odds of an individual who was symptomatic (vs asymptomatic) being seen at the time of first testing. Individuals who were symptomatic presented with one of the following at the time of testing: fever or chills, cough, shortness of breath or difficulty breathing, lack of energy or general tired feeling, muscle or body aches, headache, new loss of taste or smell, sore throat/congestion or runny nose, feeling sick to your stomach or vomiting/diarrhea, abdominal pain, or skin rash.

**Predictor Covariates**

Predictor covariates in the analysis included age categories (<18, 19 – 29, 30 – 39, 40 – 49, 50 – 59, and ≥60 years), race, sex at birth, whether a person is an essential worker, whether a person is fully vaccinated (eg, received two doses of Moderna/Pfizer or one dose of Johnson & Johnson), presence/absence of physical or mental health conditions (yes/no), six challenges to health (yes/no), specific barriers to testing (yes/no), and a measure of economic distress based on the individual’s zip code of residence to adjust for nonrandom community-level effects. Physical health conditions, mental health conditions, and barriers to testing were combined into their groupings due to the small sample size of the individual and missingness in the subgroups. Physical health conditions included immunocompromised condition, autoimmune disease, hypertension, diabetes, chronic kidney disease, cancer diagnosis or treatment within the past 12 months, cardiovascular disease, asthma, chronic obstructive pulmonary disease, other chronic lung disease, and sickle cell anemia. Mental health conditions included depression, alcohol or substance use disorder, intravenous drug use, and other mental health disorders. The six challenges to health included access to mental/physical health care, having a place to stay/live, getting enough food to eat, having clean water to drink, getting the medicine needed, and having transportation from one place to another. Barriers to testing included protected time off to visit a testing site; out-of-pocket costs for test; out-of-pocket costs for transportation, childcare, or time off work to get tested; knowledge of where testing is done in their community; pain or discomfort from the test or saliva collection; and concern about others handling their personal data. All predictor covariates, except for the economic distress score, were collected as CDEs required for all funded RADx-UP2 projects [21]. The zip code–level Distressed Communities Index (DCI) was linked to survey data by individual zip code of residence to adjust for nonrandom selection of underserved communities for testing. The DCI is a measurement of community economic disparities that consists of seven measures obtained from the US Census Bureau’s American Community Survey: no high school diploma, housing vacancy rate, adults not working, poverty rate, median income ratio, changes in employment, and changes in establishments. This was critical as RADx-UP2 nonrandomly selects communities for testing based on whether they are underserved. DCI was a continuous variable, where higher numbers indicated more distress [25]. This index of socioeconomic deprivation has been utilized in previous social epidemiology literature to characterize health disparities [26,27].

**Statistical Analysis**

Data were analyzed by multivariable logistic regression to evaluate the association between the odds of an individual being symptomatic at the time of the first test and each of the predictors. Interaction effects between the six challenges to health and the DCI were also included in the multivariable logistic regression model, and backward selection with Akaike information criterion was used to ensure the best covariates model was used adjusting for age, gender, and race after selection. Statistical significance was evaluated using adjusted odds ratios (aORs) and corresponding 95% CIs at an α level of .05. Tests for global spatial autocorrelation (clustering) of individuals who were symptomatic were conducted using a global Moran I value. Statistical spatial dependence was evaluated using the tests’ computed z score and P value [28]. All data management and regression analyses were conducted in R (The R Foundation for Statistical Computing). Spatial analysis and thematic maps displaying zip code–level relationships between the rate of people who were symptomatic per 10 individuals tested and the distressed communities score were created in ArcGIS Pro 2.9.2 (Esri).

**Results**

**Data Source and Management**

Of the 2103 testing questionnaires completed between May 7 and November 14, 2022, 1423 unique individuals were identified as having self-reported as being symptomatic at the time of their first test (Table 1). In the overall sample, 24.5% (n=348) were 60 years or older, 85.5% (n=1217) were White, and 51.7% (n=735) were female. The majority of individuals were vaccinated (n=975, 68.5%), did not report any physical (n=773, 54.3%) or mental (n=1120, 78.7%) health conditions, and did not have any of the six challenges to health: access to health care (n=1155, 81.2%), place to stay/live (n=1162, 81.7%), enough food to eat (n=1198, 84.2%), clean water to drink (n=1228, 86.3%), getting needed medication (n=1163, 81.7%), and having transportation (n=1165, 81.9%). Among individuals who were symptomatic, 26.6% (n=219) were 60 years or older, 87.1% (n=717) were White, and 55.8% (n=459) were female. Similar to the overall sample, the majority of individuals who were symptomatic were vaccinated (n=581, 70.6%), did not report any physical (n=418, 50.8%) or mental (n=656, 79.7%) health issues, and did not have any of the six challenges to health: access to health care (n=736, 89.4%), place to stay/live (n=746, 90.6%), enough food to eat (n=759, 92.2%), clean water to drink (n=773, 93.9%), getting needed medication (n=741, 90%), and having transportation (n=747, 90.8%).
### Table
Demographics and clinical characteristics of individuals tested for SARS-CoV-2 during phase 2 of the Rapid Acceleration of Diagnostics for Underserved Populations program. The program took place between May 7 and November 14, 2022, and tested 1423 unique individuals who self-reported as being symptomatic at the time of their first test.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Individuals, n (%)</th>
<th>Clinical symptoms, n (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nonsymptomatic (n=600)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Symptomatic (n=823)</td>
<td></td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤18</td>
<td>285 (20.0)</td>
<td>110 (18.3)</td>
<td>175 (21.3)</td>
</tr>
<tr>
<td>19-29</td>
<td>143 (10.0)</td>
<td>51 (8.5)</td>
<td>92 (11.2)</td>
</tr>
<tr>
<td>30-39</td>
<td>208 (14.6)</td>
<td>100 (16.7)</td>
<td>108 (13.1)</td>
</tr>
<tr>
<td>40-49</td>
<td>188 (13.2)</td>
<td>92 (15.3)</td>
<td>96 (11.7)</td>
</tr>
<tr>
<td>50-59</td>
<td>164 (11.5)</td>
<td>74 (12.3)</td>
<td>90 (10.9)</td>
</tr>
<tr>
<td>≥60</td>
<td>348 (24.5)</td>
<td>129 (21.5)</td>
<td>219 (26.6)</td>
</tr>
<tr>
<td>Missing</td>
<td>87 (6.1)</td>
<td>44 (7.3)</td>
<td>43 (5.2)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1217 (85.5)</td>
<td>500 (83.3)</td>
<td>717 (87.1)</td>
</tr>
<tr>
<td>Black/African American</td>
<td>114 (8.0)</td>
<td>61 (10.2)</td>
<td>53 (6.4)</td>
</tr>
<tr>
<td>Other</td>
<td>67 (4.7)</td>
<td>33 (5.5)</td>
<td>34 (4.1)</td>
</tr>
<tr>
<td>Missing</td>
<td>25 (1.8)</td>
<td>6 (1.0)</td>
<td>19 (2.3)</td>
</tr>
<tr>
<td><strong>Sex at birth</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>735 (51.7)</td>
<td>276 (46.0)</td>
<td>459 (55.8)</td>
</tr>
<tr>
<td>Male</td>
<td>651 (45.7)</td>
<td>307 (51.2)</td>
<td>344 (41.8)</td>
</tr>
<tr>
<td>Missing</td>
<td>37 (2.6)</td>
<td>17 (2.8)</td>
<td>20 (2.4)</td>
</tr>
<tr>
<td><strong>Essential worker</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>990 (69.6)</td>
<td>429 (71.5)</td>
<td>561 (68.2)</td>
</tr>
<tr>
<td>Yes</td>
<td>319 (22.4)</td>
<td>118 (19.7)</td>
<td>201 (24.4)</td>
</tr>
<tr>
<td>Missing</td>
<td>114 (8.0)</td>
<td>53 (8.8)</td>
<td>61 (7.4)</td>
</tr>
<tr>
<td><strong>Vaccinated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>392 (27.5)</td>
<td>182 (30.3)</td>
<td>210 (25.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>975 (68.5)</td>
<td>394 (65.7)</td>
<td>581 (70.6)</td>
</tr>
<tr>
<td>Missing</td>
<td>56 (3.9)</td>
<td>24 (4.0)</td>
<td>32 (3.9)</td>
</tr>
<tr>
<td><strong>Physical health condition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>773 (54.3)</td>
<td>355 (59.2)</td>
<td>418 (50.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>472 (33.2)</td>
<td>168 (28.0)</td>
<td>304 (36.9)</td>
</tr>
<tr>
<td>Missing</td>
<td>178 (12.5)</td>
<td>77 (12.8)</td>
<td>101 (12.3)</td>
</tr>
<tr>
<td><strong>Mental health condition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1120 (78.7)</td>
<td>464 (77.3)</td>
<td>656 (79.7)</td>
</tr>
<tr>
<td>Yes</td>
<td>233 (16.4)</td>
<td>110 (18.3)</td>
<td>123 (14.9)</td>
</tr>
<tr>
<td>Missing</td>
<td>70 (4.9)</td>
<td>26 (4.3)</td>
<td>44 (5.3)</td>
</tr>
<tr>
<td><strong>Challenges to health</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Access to health care</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1155 (81.2)</td>
<td>419 (69.8)</td>
<td>736 (89.4)</td>
</tr>
<tr>
<td>Yes</td>
<td>232 (16.3)</td>
<td>173 (28.8)</td>
<td>59 (7.2)</td>
</tr>
<tr>
<td>Missing</td>
<td>36 (2.5)</td>
<td>8 (1.3)</td>
<td>28 (3.4)</td>
</tr>
<tr>
<td><strong>Place to stay/live</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Individuals, n (%)</td>
<td>Clinical symptoms, n (%)</td>
<td>Non-symptomatic (n=600)</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1162 (81.7)</td>
<td>416 (69.3)</td>
<td>746 (90.6)</td>
</tr>
<tr>
<td>Yes</td>
<td>216 (15.2)</td>
<td>171 (28.5)</td>
<td>45 (5.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>45 (3.2)</td>
<td>13 (2.2)</td>
<td>32 (3.9)</td>
</tr>
<tr>
<td><strong>Enough food to eat</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1198 (84.2)</td>
<td>439 (73.2)</td>
<td>759 (92.2)</td>
</tr>
<tr>
<td>Yes</td>
<td>185 (13.0)</td>
<td>150 (25.0)</td>
<td>35 (4.3)</td>
</tr>
<tr>
<td>Missing</td>
<td>40 (2.8)</td>
<td>11 (1.8)</td>
<td>29 (3.5)</td>
</tr>
<tr>
<td><strong>Clean water to drink</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1228 (86.3)</td>
<td>455 (75.8)</td>
<td>773 (93.9)</td>
</tr>
<tr>
<td>Yes</td>
<td>159 (11.2)</td>
<td>136 (22.7)</td>
<td>23 (2.8)</td>
</tr>
<tr>
<td>Missing</td>
<td>36 (2.5)</td>
<td>9 (1.5)</td>
<td>27 (3.3)</td>
</tr>
<tr>
<td><strong>Getting needed medicine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1163 (81.7)</td>
<td>422 (70.3)</td>
<td>741 (90.0)</td>
</tr>
<tr>
<td>Yes</td>
<td>212 (14.9)</td>
<td>162 (27.0)</td>
<td>50 (6.1)</td>
</tr>
<tr>
<td>Missing</td>
<td>48 (3.4)</td>
<td>16 (2.7)</td>
<td>32 (3.9)</td>
</tr>
<tr>
<td><strong>Transportation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1165 (81.9)</td>
<td>418 (69.7)</td>
<td>747 (90.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>206 (14.5)</td>
<td>165 (27.5)</td>
<td>41 (5.0)</td>
</tr>
<tr>
<td>Missing</td>
<td>52 (3.7)</td>
<td>17 (2.8)</td>
<td>35 (4.3)</td>
</tr>
<tr>
<td><strong>Barriers to testing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>519 (36.5)</td>
<td>195 (32.5)</td>
<td>324 (39.4)</td>
</tr>
<tr>
<td>Yes</td>
<td>460 (32.3)</td>
<td>203 (33.8)</td>
<td>257 (31.2)</td>
</tr>
<tr>
<td>Missing</td>
<td>444 (31.2)</td>
<td>202 (33.7)</td>
<td>242 (29.4)</td>
</tr>
</tbody>
</table>

**Statistical Analysis**

In the parsimonious model, backward selection with Akaike information criterion dropped the following variables: essential worker, vaccinated, and access to health care. Among persons at the time of first testing, all age groups, races, sex at birth, barriers to testing, enough food to eat, clean water to drink, and the DCI and clean water to drink interaction were not statistically associated with the odds of seeing an individual who was symptomatic at a testing location (all P values >.05). Individuals with a physical health condition and challenges related to a place to stay/live were statistically more likely to seek testing while being symptomatic, and mental health condition and the DCI and clean water to drink interaction were moderately so. Those reporting physical health conditions were 53% more likely to have reported being symptomatic (aOR 1.53, 95% CI 0.96-2.48), and those living in a high DCI zip code while also not getting needed medicine were 6% more likely to have reported being symptomatic (aOR 1.06, 95% CI 1.00-1.13). Individuals with a challenge getting needed medicine and transportation as well as the DCI and challenges in having a place to stay/live interaction were statistically less likely to seek testing while symptomatic, and living in a high DCI zip code was moderately so. Participants who had challenges in getting needed medication were 99% less likely to report being symptomatic (aOR 0.01, 95% CI 0.00-0.82). Those who had challenges with transportation were 77% less likely to report being symptomatic (aOR 0.23, 95% CI 0.05-0.64). Those living in a high DCI zip code and facing challenges of having a place to stay/live were 7% less likely to report being symptomatic (aOR 0.93, 95% CI 0.87-0.99), and those living in a high DCI zip code were 1% less likely to seek testing as a symptomatic individual (aOR 0.99, 95% CI 0.98-1.00). Complete results for the logistic regression are displayed in Table 2.
Table. Adjusted odds ratios and corresponding 95% CIs for logistic regression models. Full discussion of results can be found in the Results: Statistical Analysis section. The parsimonious model was derived using backward selection from the original model. A full description of the procedure can be found in the Methods: Statistical Analysis section.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Original model</th>
<th></th>
<th></th>
<th>Parsimonious model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted odds ratio</td>
<td>95% CI</td>
<td>Adjusted odds ratio</td>
<td>95% CI</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤18</td>
<td>1.6</td>
<td>0.9-2.9</td>
<td>1.64</td>
<td>0.94-2.92</td>
<td></td>
</tr>
<tr>
<td>19-29 (reference)</td>
<td>1</td>
<td>— a</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>30-39</td>
<td>0.83</td>
<td>0.51-1.37</td>
<td>0.85</td>
<td>0.52-1.4</td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>0.97</td>
<td>0.57-1.67</td>
<td>0.98</td>
<td>0.58-1.68</td>
<td></td>
</tr>
<tr>
<td>50-59</td>
<td>0.75</td>
<td>0.42-1.33</td>
<td>0.73</td>
<td>0.42-1.3</td>
<td></td>
</tr>
<tr>
<td>≥60</td>
<td>0.72</td>
<td>0.44-1.17</td>
<td>0.69</td>
<td>0.43-1.1</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Black/African American</td>
<td>0.67</td>
<td>0.39-1.16</td>
<td>0.68</td>
<td>0.4-1.17</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.7</td>
<td>0.37-1.37</td>
<td>0.71</td>
<td>0.37-1.37</td>
<td></td>
</tr>
<tr>
<td>Sex at birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.87</td>
<td>0.64-1.17</td>
<td>0.87</td>
<td>0.65-1.17</td>
<td></td>
</tr>
<tr>
<td>Essential worker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.1</td>
<td>0.76-1.59</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Vaccinated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.88</td>
<td>0.61-1.27</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Physical health condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.87</td>
<td>1.31-2.68</td>
<td>1.85</td>
<td>1.3-2.65</td>
<td></td>
</tr>
<tr>
<td>Mental health condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.59</td>
<td>1-2.6</td>
<td>1.53</td>
<td>0.96-2.48</td>
<td></td>
</tr>
<tr>
<td>Challenges to health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access to health care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.52</td>
<td>0-1510</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Place to stay/live</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>142</td>
<td>0.08-1,151,363</td>
<td>307.13</td>
<td>1.46-106,372</td>
<td></td>
</tr>
<tr>
<td>Enough food to eat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.01</td>
<td>0-1793</td>
<td>2.71</td>
<td>0.79-10.27</td>
<td></td>
</tr>
<tr>
<td>Clean water to drink</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>41</td>
<td>0.06-71,082</td>
<td>66</td>
<td>0.12-112,368</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Original model</td>
<td>Parsimonious model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adjusted odds ratio</td>
<td>95% CI</td>
<td>Adjusted odds ratio</td>
<td>95% CI</td>
<td></td>
</tr>
<tr>
<td>Getting needed medicine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Transportation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>321</td>
<td>0.01-22,493,690</td>
<td>0.23</td>
<td>0.08-0.64</td>
<td></td>
</tr>
<tr>
<td>Barriers to testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (reference)</td>
<td>1</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.31</td>
<td>0.95-1.8</td>
<td>1.29</td>
<td>0.94-1.77</td>
<td></td>
</tr>
<tr>
<td>Distress score</td>
<td>0.99</td>
<td>0.98-1</td>
<td>0.99</td>
<td>0.98-1</td>
<td></td>
</tr>
</tbody>
</table>

*Not applicable.

Geospatial Analysis

A bivariate map of the zip code–level rate of individuals who were symptomatic per 10 individuals seeking COVID-19 testing and the DCI is displayed in Figure 1. Visually, there appear to be overlapping trends in the DCI and rate of people who are symptomatic per 10 people served at testing locations in the southern and northern regions of West Virginia. In particular, southern West Virginia had more zip codes where the rate of tested people who were symptomatic was low and the DCI was low, indicating fewer people who were symptomatic from non-distressed communities when compared to the rest of the state. There was only 1 zip code in the northern region of West Virginia that followed this trend. However, both regions had zip codes where the rate of tested people who were symptomatic per 10 individuals was high and the DCI score was high, indicating a high number of people who were symptomatic coming from distressed communities. In the southern region, there were many zip codes with a high DCI score but a low rate of people who were symptomatic. This visual observation supports findings from the logistic regression that the DCI was statistically associated with a lower rate of tested people who were symptomatic, particularly for persons in southern West Virginia. When assessing spatial autocorrelation, global Moran I did not detect any statistically significant clustering in the rate of people who were symptomatic per 10 individuals tested throughout the RADx-UP study area. Statistically significant clustering was evaluated incrementally across distance thresholds of varying diameters (smallest: 357 km, Moran I=0.002, P=.63; largest: 784 km, Moran I=0.001, P=.06) without indication of statistical significance.
Discussion

Principal Results

This study identified several factors associated with test seeking among individuals who were symptomatic at RADx-UP2 COVID-19 testing locations. Our study found that individuals with a physical or mental health condition, those facing a challenge in having a place to stay/live, or those with the interaction of living in a high DCI zip code while also not getting needed medicine were more likely to appear at a testing location with symptoms. Additionally, individuals from less distressed communities, who were able to access needed medicine or transportation, and those with the interaction of living in a high DCI zip code and facing challenges in having a place to stay/live were less likely to be symptomatic at the time of their testing. Importantly, we found no statistically significant geographic pattern in the rate of people who were symptomatic per 10 individuals tested. This could suggest that differences observed for persons less likely to be symptomatic by a higher DCI were not due to geographic contexts, such as urban or rural, and perhaps more related to social determinants of health such as facing a challenge in having a place to stay/live within the individual’s zip code of residence. Importantly, these findings address a gap in the existing literature, particularly among studies that utilize recent testing data within epidemiological investigations by looking at underserved areas.
and the reasoning behind individuals seeking testing [15,17]. Recent testing data reflect a shift toward symptomatic populations who are more likely to struggle with a stable living situation and experience multiple physical or mental health conditions. This is an important consideration, as new information from this study provides an idea of the extent to which the generalizability of testing data is restricted to vulnerable populations or those separate from the general population.

Physical and Mental Health Conditions
Physical and mental health conditions were found to be associated with individuals presenting with symptoms for COVID-19 testing. Physical health conditions, such as autoimmune disease, hypertension, diabetes, chronic kidney disease, and cardiovascular disease, can cause impediments to the immune system and leave individuals more susceptible to severe illnesses, including COVID-19 [29,30]. These individuals may be more willing to seek out testing when they become symptomatic due to their increased risk of serious illness [4,15,27,29,30]. Altogether these individuals would be more likely to be symptomatic when reporting to testing facilities, whether due to the increased risk of the physical conditions themselves or as a preventative measure taken by the individuals. Additionally, mental health conditions, such as alcohol and substance use, can also increase an individual’s susceptibility to infection from and exposure to COVID-19 [22,31,32]. Mental health conditions can lead to impediments in the immune system, which make an individual more susceptible to COVID-19 and increase situations of greater exposure to COVID-19 [22,31,32]. This is particularly relevant to individuals experiencing homelessness, who are a vulnerable population at high risk for mental health conditions and must undergo COVID-19 testing to gain entrance to shelters [11,15,31-33].

Challenges to Health and Economic Distress
Those individuals who have challenges in having a place to stay/live and those with the interaction of DCI and getting needed medication were more likely to be symptomatic at the time of testing. These socioeconomic issues could be associated with these individuals being more vulnerable to exposures, leading to more chances of respiratory disease spread due to related aspects such as homelessness or not being able to afford health care such as medication. It was found that individuals who reported challenges in getting needed medicine or transportation, those who lived in distressed communities, and those with the interaction of living in a distressed community while having challenges in having a place to stay/live were less likely to be symptomatic at the time of testing. These associations with not having issues of getting needed medicine or transportation challenges to health could indicate there are fewer travel/access obstacles to the health of an individual as well as fewer people experiencing homelessness in these socioeconomic groups. This could indicate that individuals who are not impeded by these socioeconomic drivers are more likely to seek testing when becoming symptomatic. Coinciding with having no challenges to health, living in higher areas of greater economic distress was associated with lower odds of being symptomatic (Figure 1: pink areas). These findings are interesting because these components measure socioeconomic challenges at both the individual and community levels. These findings give insight into the behaviors of underserved communities that exist across West Virginia when compared to previous studies that look at population-level data and collection methods that would otherwise limit these underserved communities [15,17].

Limitations
This study has several limitations. First, data for the study comes from questionnaires that are self-reported by the individuals. Due to recall bias or social desirability bias, individuals may be misclassified according to symptomatic status or the presence of a potential predictor [34,35]. Next, many of these symptoms that individuals reported could also be present in the transmission of other pathogens [36]. However, we believe that this did not impact the validity of the study, as the goal was to better understand which factors were associated with the use of testing services in any individual who was symptomatic. Third, individuals who were symptomatic faced challenges to health, such as getting to a testing site or not knowing about available testing, and may not have sought testing. Fourth, the sample size does not indicate confirmed COVID-19 cases—only those who were symptomatic and seeking COVID-19 testing. Finally, the study population is only a subset of the total underserved areas of West Virginia, and some study variables had small sample representation or missing data.

Conclusions
Overall, this study of symptomatic factors associated with COVID-19 testing in West Virginia emphasized the urgent need to better understand barriers to testing. Despite limitations, this research addresses gaps in the current COVID-19 testing research. This is especially important in underserved areas experiencing disparities, such as the southwestern part of West Virginia (Figure 1). Critical to future public health policy creation is determining why individuals who are symptomatic in high-distress areas are less likely to seek free COVID-19 testing. While factors such as a lack of transportation are possible, there may be other reasons such as belief in the presence of ongoing SARS-CoV-2 transmission or belief in effective prevention (eg, vaccines or quarantine) or treatment.

Acknowledgments
Melanie Taylor, Triston Nutter, Caitlin Herdman, Raven Allen, Katherine Belcher, and Andrew Meyer of West Virginia Clinical and Translational Science Institute all provided research support. Research support was received from the National Institute of General Medical Sciences of the National Institutes of Health (NIH NIGMS) under award 2U54GM104942-07 for the West Virginia Clinical and Translational Science Institute. Additional support was provided by the National Institute on Minority
Health and Health Disparities’ Rapid Acceleration of Diagnostics Among Underserved Populations (RADx-UP: 1U01MD017419-01) and NIH NGMS RADx-UP (3U54GM104942-06S1). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Data Availability**

The original data sets contain personal health information and small identifiable subgroups, and cannot be shared publicly. For questions about data availability, contact the corresponding author.

**Conflicts of Interest**

None declared.

**References**


21. About RADx-UP. RADx-UP. URL: https://radx-up.org/about/ [accessed 2024-07-04]


Abbreviations

aOR: adjusted odds ratio
CDE: common data element
DCI: Distressed Communities Index
RADx-UP: Rapid Acceleration of Diagnostics for Underserved Populations
SOAR: Solutions Oriented Addiction Response

Edited by A Mavragani; submitted 14.09.23; peer-reviewed by R Chandrasekaran, S Thawillarp; revised version received 03.06.24; accepted 06.06.24; published 18.07.24.

Please cite as:
Dotson T, Price B, Witrick B, Davis S, Kemper E, Whanger S, Hodder S, Hendricks B
Factors Associated With Surveillance Testing in Individuals With COVID-19 Symptoms During the Last Leg of the Pandemic: Multivariable Regression Analysis
JMIR Public Health Surveill 2024;10:e5276
URL: https://publichealth.jmir.org/2024/1/e5276
doi:10.2196/5276
Sex Differences in Clustering Unhealthy Lifestyles Among Survivors of COVID-19: Latent Class Analysis

Lan T H Le1,2,3*, MD, PhD; Thi Ngoc Anh Hoang4*, MPhil; Tan T Nguyen5,6, MD, PhD; Tien D Dao7, MD, PhD; Binh N Do8,9, MD, PhD; Khue M Pham10, MD, PhD; Vinh H Vu11, MD, PhD; Linh V Pham12, MD, PhD; Lien T H Nguyen12, MD, MSc; Hoang C Nguyen1,13, MD, PhD; Tuan V Tran14, MD, PhD; Trung H Nguyen15, MD, PhD; Anh T Nguyen16, MD; Hoan V Nguyen11,17, MD; Phuoc B Nguyen18, MD, PhD; Hoai T T Nguyen19, MD; Thu T M Pham10,20, MD, PhD; Thuy T Le21,22, MD, PhD; Thao T P Nguyen23, MD, MSc; Cuong Q Tran24, MD, PhD; Ha-Linh Quach25, MPH, MPhil; Kien T Nguyen26, MPH; Tuyen Van Duong27,28, PhD

1Director Office, Thai Nguyen National Hospital, Thai Nguyen, Vietnam
2Training and Direction of Healthcare Activity Center, Thai Nguyen National Hospital, Thai Nguyen, Vietnam
3Biochemistry Department, Thai Nguyen National Hospital, Thai Nguyen, Vietnam
4Faculty of Medicine, Phenikaa University, Hanoi, Vietnam
5Department of Orthopedics, Can Tho University of Medicine and Pharmacy, Can Tho, Vietnam
6Director Office, Can Tho University of Medicine and Pharmacy Hospital, Can Tho, Vietnam
7Institute of Oncology and Nuclear Medicine, Military Hospital 175, Ho Chi Minh City, Vietnam
8Department of Infectious Diseases, Vietnam Military Medical University, Ha Noi, Vietnam
9Department of Military Science, Vietnam Military Medical University, Ha Noi, Vietnam
10Faculty of Public Health, Hai Phong University of Medicine and Pharmacy, Hai Phong, Vietnam
11Infectious and Tropical Diseases Department, Viet Tiep Hospital, Hai Phong, Vietnam
12Department of Pulmonary and Cardiovascular Diseases, Hai Phong University of Medicine and Pharmacy Hospital, Hai Phong, Vietnam
13President Office, Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam
14Neurology Department, Thai Nguyen University of Medicine and Pharmacy, Thai Nguyen, Vietnam
15Director Office, Gang Thep Hospital, Thai Nguyen, Vietnam
16Director Office, Hospital for Tropical Diseases, Hai Duong, Vietnam
17Department of Infectious Diseases, Hai Phong University of Medicine and Pharmacy, Hai Phong, Vietnam
18Director Office, Kien An Hospital, Hai Phong, Vietnam
19Training and Direction of Healthcare Activity Center, Kien An Hospital, Hai Phong, Vietnam
20School of Public Health, College of Public Health, Taipei Medical University, Taipei, Taiwan
21President Office, Da Nang University of Medical Technology and Pharmacy, Da Nang, Vietnam
22Faculty of Medical Laboratory Science, Da Nang University of Medical Technology and Pharmacy, Da Nang, Vietnam
23Institute for Community Health Research, University of Medicine and Pharmacy, Hue University, Hue, Vietnam
24Faculty of Public Health, Pham Ngoc Thach University of Medicine, Ho Chi Minh City, Vietnam
25Centre for Ageing Research & Education, Duke-NUS Medical School, National University of Singapore, Singapore, Singapore
26Department of Health Promotion, Faculty of Social and Behavioral Sciences, Hanoi University of Public Health, Ha Noi, Vietnam
27School of Nutrition and Health Sciences, Taipei Medical University, Taipei, Taiwan
28International Master/Ph.D. Program in Medicine, Taipei Medical University, Taipei, Taiwan
*these authors contributed equally

Corresponding Author:
Tuyen Van Duong, PhD
School of Nutrition and Health Sciences
Taipei Medical University
No. 250, Wuxing Street
Taipei, 11031
Taiwan
Phone: 886 227361661 ext 6545
Fax: 886 227373112
Email: tvduong@tmu.edu.tw
Abstract

Background: The COVID-19 pandemic has underscored the significance of adopting healthy lifestyles to mitigate the risk of severe outcomes and long-term consequences.

Objective: This study focuses on assessing the prevalence and clustering of 5 unhealthy lifestyle behaviors among Vietnamese adults after recovering from COVID-19, with a specific emphasis on sex differences.

Methods: The cross-sectional data of 5890 survivors of COVID-19 in Vietnam were analyzed from December 2021 to October 2022. To examine the sex differences in 5 unhealthy lifestyle behaviors (smoking, drinking, unhealthy diet, physical inactivity, and sedentary behavior), the percentages were plotted along with their corresponding 95% CI for each behavior. Latent class analysis was used to identify 2 distinct classes of individuals based on the clustering of these behaviors: the “less unhealthy” group and the “more unhealthy” group. We examined the sociodemographic characteristics associated with each identified class and used logistic regression to investigate the factors related to the “more unhealthy” group.

Results: The majority of individuals (male participants: 2432/2447, 99.4% and female participants: 3411/3443, 99.1%) exhibited at least 1 unhealthy behavior, with male participants being more susceptible to multiple unhealthy behaviors. The male-to-female ratio for having a single behavior was 1.003, but it escalated to 25 for individuals displaying all 5 behaviors. Male participants demonstrated a higher prevalence of combining alcohol intake with sedentary behavior (949/2447, 38.8%) or an unhealthy diet (861/2447, 35.2%), whereas female participants tended to exhibit physical inactivity combined with sedentary behavior (375/3443, 39.9%) or an unhealthy diet (1260/3443, 36.6%). Married male participants had increased odds of falling into the “more unhealthy” group compared to their single counterparts (odds ratio [OR] 1.45, 95% CI 1.14-1.85), while female participants exhibited lower odds (OR 0.65, 95% CI 0.51-0.83). Female participants who are underweight showed a higher likelihood of belonging to the “more unhealthy” group (OR 1.11, 95% CI 0.89-1.39), but this was not observed among male participants (OR 0.6, 95% CI 0.41-0.89). In both sexes, older age, dependent employment, high education, and obesity were associated with higher odds of being in the “more unhealthy” group.

Conclusions: The study identified notable sex differences in unhealthy lifestyle behaviors among survivors of COVID-19. Male survivors are more likely to engage in unhealthy behaviors compared to female survivors. These findings emphasize the importance of tailored public health interventions targeting sex-specific unhealthy behaviors. Specifically, addressing unhealthy habits is crucial for promoting post–COVID-19 health and well-being.

(JMIR Public Health Surveill 2024;10:e50189) doi:10.2196/50189

KEYWORDS
sex difference; cluster; lifestyle behavior; COVID-19 recovery; latent class analysis; sex; unhealthy; lifestyle; adult; long COVID-19; infected; survivor; public health; intervention; promote; well-being; COVID-19; adults; mobile phone

Introduction

The COVID-19 pandemic has a significant impact on population health and causes widespread disruption globally [1]. Beyond the public health crisis, it has triggered substantial alternations in people’s lifestyles, including poorer nutrition intake, sedentary lifestyle due to prolonged lockdown, sleep disturbance, and mental health problems [2-4]. These unhealthy habits are associated with a higher risk of noncommunicable diseases (NCDs) [5-7] and greater vulnerability to SARS-CoV-2 infection, resulting in more severe COVID-19 outcomes [8-10].

In Vietnam, the COVID-19 control policies have exerted a considerable impact on people’s lifestyles [11]. The stringent isolation measures have resulted in significant changes in daily routines. For example, symptomatic patients underwent isolation for a minimum of 14 days in designated facilities such as medical camps or field hospitals, where living conditions were limited. This was followed by an additional 14 days of quarantaine at home after recovery [12]. These changes are particularly noticeable in alterations to diet and physical activities [13]. Even asymptomatic individuals with COVID-19 (those testing positive for SARS-CoV-2 without symptoms) quarantined for at least 14 days at home are susceptible to a sedentary lifestyle, primarily due to the limited physical activity options and dependence on provided foods [14]. Indeed, several publications have indicated that survivors of COVID-19 experience a wide range of health problems after recovery [15], and these problems are often linked to an individual’s lifestyle and health behaviors [16,17]. Additionally, studies have shown that unhealthy lifestyle behaviors tend to cluster, with individuals who engage in 1 unhealthy behavior being more likely to engage in others [18,19]. For example, the co-occurrence of a sedentary lifestyle with excessive substance use, alcohol consumption, and smoking can lead to worse health conditions [18-20], especially for survivors of COVID-19 who are already vulnerable. While several publications have examined the long-term mental and physical health of COVID-19 infection [21,22], there is limited evidence on the unhealthy lifestyles among survivors of COVID-19, that is, who are more likely to engage in these lifestyles. It would be of interest for policy makers to provide timely and targeted health promotion interventions to survivors of COVID-19 to avoid further complications of long-term COVID-19 health problems.

Several studies have found significant differences in the clustering of unhealthy lifestyle behaviors between sex [18,19,23-25]. For instance, one study in the United States...
investigated the sex differences in lifestyle behaviors among adults with diabetes between 1999 and 2018. The findings revealed that both male participants and female participants with diabetes displayed unhealthy lifestyle behaviors, but the specific behaviors differed by sex. Male patients with diabetes were more likely to smoke and engage in excessive drinking, whereas their counterparts were more likely to be physically inactive and have poor dietary habits [25]. In another study focusing on Indian adults, researchers explored that the prevalence of multiple unhealthy lifestyle behaviors is significant, with a higher tendency for male adults to exhibit the clustering of multiple factors compared to female ones [26].

In Vietnam, sex differences in lifestyle behaviors arise from societal expectations. Male participants often exhibit a higher tendency to participate in risky behavior influenced by socialization, whereas female participants benefit from their cautiousness and adherence to healthy behaviors and health education [27-29]. For instance, within Vietnamese tradition, smoking and drinking might be perceived as integral to the male gender role, serving to demonstrate masculinity and foster social connections [30]. In contrast, societal expectations generally discourage women from engaging in these behaviors, aligning with gender role norms.

Although previous research has extensively investigated various patterns of lifestyle behaviors, little is known about whether these patterns differ among adult survivors of COVID-19. Additionally, previous studies have examined the impact of sociodemographic characteristics such as race, ethnicity, culture, age, socioeconomic status, employment, and education on unhealthy behaviors [31-33]. However, a sex-specific analysis to understanding and measuring these behaviors is still lacking, except in studies conducted on adolescents [23,34]. Therefore, this study aims to use latent class analysis (LCA) to identify distinct patterns of unhealthy lifestyle behaviors, including cigarette smoking, alcohol consumption, unhealthy diet, physical inactivity, and sedentary behaviors, among survivors of COVID-19 after their recovery, focusing on sex differences. Furthermore, this study seeks to identify the factors associated with “more unhealthy” behaviors in this population.

**Methods**

**Study Settings**

Vietnam first confirmed the presence of the SARS-CoV-2 virus in January 2020, with initial cases being mostly from other countries. However, local transmission began to develop in February and March 2020. To curb the spread, the Vietnamese government implemented a zero-COVID-19 strategy throughout 2020, which involved contact tracing, mass testing, quarantining, and lockdowns. This approach was largely successful, but since April 2021, the country has been facing its largest outbreak yet. As a result, lockdowns have been implemented in one-third of provinces and cities, affecting roughly one-third of the population. The emergence of the Omicron variant in the first quarter of 2022 led to a sharp increase in infections, although Vietnam’s high vaccination rates resulted in fewer fatalities [35] (Figure 1).

Our study collection period spanned from December 1, 2021, to October 31, 2022. The research was carried out across 17 hospitals and health centers located in 9 provinces across the country. We included 5 provinces with 8 hospitals in the Northern region, 3 provinces with 3 health centers in the Central region, and 2 provinces with 6 hospitals and health centers in the Southern region (Multimedia Appendix 1).

**Figure 1.** COVID-19 situation in Vietnam from April 2021 to December 2022.
Study Design and Participants
This is a nationwide cross-sectional, web-based survey conducted from 2021 to 2022 in Vietnam. We enrolled Vietnamese adults (aged 18 years or older) who had been infected with SARS-CoV-2, which is confirmed by available positive test results for SARS-CoV-2 (either real-time polymerase chain reaction or rapid antigen test), and afterward recovered from the infection, which is confirmed by available negative test results for SARS-CoV-2 (either real-time polymerase chain reaction or rapid antigen test). Participants who did not consent, were nonpermanent residents of Vietnam, were unable to understand the survey questions, or had cognitive or mental health issues that may have affected their responses to the survey were excluded from the study (n=87).

Participants from 17 hospitals and COVID-19 health centers across 9 provinces representative of 3 geographical regions of Vietnam were purposively selected. We recruited 3450 participants in 4 Northern region provinces and cities, 1130 participants in 3 Central region provinces, and 1310 participants in 2 Southern region cities. In total, we collected data from 5890 participants. Details of the number of participants recruited per site of the 3 geographical regions can be found in Figure 2 and Multimedia Appendix 1.

Figure 2. Distribution of participants in the study.

Sample Size
A sample of 5890 participants provided sufficient data for inclusion in the analysis, representing 50.1% of the total 11,761 survivors of COVID-19 from 17 study sites and 0.05% of the total 10,607,166 survivors of COVID-19 in Vietnam until the end of data collection on October 31, 2022 [35].

Data Collection Procedures
The data were collected from December 1, 2021, to October 31, 2022. The survey form was prepared on Microsoft Forms (Microsoft Corp). Initially, we secured agreements to collaborate with 17 target sites for the research. Before data collection, senior researchers provided the enumerators, who were health care workers at each site, with an introduction to the data collection procedures. The enumerators reached out to survivors of COVID-19 through various contact channels (phone, smartphone app, and email) and distributed the survey link. The reminders were sent out to achieve the response rate. Data were collected, extracted, and coded for analysis anonymously.

Assessments and Measurements
LCA Indicators: Unhealthy Lifestyle Behaviors
Based on previous literature, we assessed 5 most common unhealthy lifestyle behaviors including the following binary indicators:

- Cigarette smoking: Participants were asked about their tobacco use status and were given 3 options: “currently smoking,” “used to smoke and stopped,” and “never smoke.” We coded the variable “cigarette smoking” as “nonsmokers” if participants answered “never smoke” or “used to smoke and stopped.” Participants who answered “currently smoking” were categorized as “smokers.”
- Alcohol consumption: We asked participants whether they had consumed at least 1 standard drink of alcoholic beverage in the past 30 days. The variable was binary coded as “yes” or “no.”
- Unhealthy diet: We used the 5-item healthy eating score to assess the healthy diet. The 5-item healthy eating score is comparable with the 2015 health eating index in assessing the overall diet quality [36] and has been validated and used in Vietnam [37]. Participants were asked to respond to 5
questions about their consumption frequency of fruits, vegetables, whole grains, dairy, and fish over the past 30 days using a 6-point scale ranging from 0=rarely or never to 5=3 or more times per day. The total score can range from 0 to 25, with a higher score indicating a healthier diet. We used a cut-off point as the median of 15 (IQR 13-17) to create a binary variable, with a score of less than 15 indicating the “unhealthy diet.”

• Physical inactivity: Physical activity is defined as any bodily movement that requires the contraction of skeletal muscles and increases energy expenditure above the resting metabolic rate [38]. In this study, we used the short version of the International Physical Activity Questionnaire to assess participants’ physical activity levels [39]. Participants who reported engaging in vigorous-intensity activity on 3 or more days per week, or moderate-intensity activity or walking for at least 30 minutes per day on 5 or more days, were considered to have met the criteria for being “physically active,” and the opposite was labeled “physically inactive.”

• Sedentary behaviors: Sedentary behaviors are defined as any activities that involve sitting or reclining and consume an energy expenditure equal to or below 1.5 metabolic equivalents. Sitting time is usually the main indicator used to quantify the time devoted to sedentary behaviors [40]. To assess sedentary behavior, we asked survivors of COVID-19 about the average number of hours they spent sitting for nonwork purposes (eg, watching television, playing computer games, using social media, or engaging in other sitting activities) on a typical day in the past 7 days. Participants who reported spending more than 2 hours per day on these activities were classified as having sedentary behavior [34].

**Clinical Parameters**

Participants were asked to report their height in meters and weight in kilograms. The self-reported weight is concordant with image-captured weight in web-based research [41]. The BMI was calculated by dividing their weight by the square of their height in meters. Participants were then classified as underweight (BMI<18.5 kg/m²), normal weight (18.5 kg/m²≤BMI<23 kg/m²), overweight (23 kg/m²≤BMI<25 kg/m²), or obese (BMI≥25 kg/m²) [42]. In addition, comorbid conditions other than COVID-19 were also assessed.

**Sociodemographic Characteristics**

Sociodemographic characteristics were also measured, including age (years), sex (male or female), marital status (single; married; or widow, divorced, or separated), education levels (illiterate or elementary, secondary or high school, vocational or college, or university or higher), employment status (dependent, independent, or unemployed), whether the participant was a health care worker (yes or no), and self-reported social status (low, middle, or high).

**Data Analysis**

Data were analyzed on Stata (version 17.0; StataCorp). First, we tabulated sociodemographic characteristics and nutritional status for the overall sample separately by sex and differentiated by Wilcoxon rank-sum tests, Pearson chi-square tests, and Fisher exact tests. Second, the prevalence and 95% CIs of 5 unhealthy lifestyle behaviors were calculated for male and female participants. The 95% CIs were calculated using the formula CI=proportion mean±1.96×SE of proportion mean [43]. We then used the UpSet diagrams to describe the combinations of these behaviors [44]. The UpSet diagrams can help display complex intersections of a lifestyle behaviors matrix, where the rows represent different sets of combinations and the columns represent the number of male and female participants who had these combinations.

LCA is a statistical method used to identify subgroups of individuals based on observed variables [45]. Previous studies have used LCA to identify subgroups with distinct patterns of lifestyle behaviors and explore the factors associated with these patterns [31,34]. We used the expectation-maximization algorithm of LCA with 30 iterations—a general optimization technique for deriving maximum likelihood estimates in the presence of latent variables from 5 indicators [46]. This algorithm alternated between the expectation step—imputing latent variables based on current parameter estimates, and the maximization step—updating parameters to maximize the likelihood of the observed data from these 5 indicators [46,47]. Our approach involved systematically testing a series of models with an increasing number of latent classes to pinpoint the most fitting model. We considered Akaike information criterion (AIC) and Bayesian information criterion (BIC) to choose the optimal number of classes. The class with the smallest AIC and BIC was considered a good fit [33].

We defined 2 distinct classes, namely, the “less unhealthy” group and the “more unhealthy” group. Within each latent group, we assessed the predicted probabilities for categorical indicators and assigned each participant to the group with the highest probability. The distribution of indicators and covariates was compared among each group membership, and logistic regression analysis was performed to estimate the association between covariates and being in the “more unhealthy” group, stratified by sex. The selection of covariates for this model was informed by our hypothesis of correlation, guided by the findings in the literature [4,16,18,19,23,25,29]. Odds ratio (OR) and 95% CIs were reported. Likelihood ratio tests were used to determine the effect of each variable in the models.

**Ethical Considerations**

This study was reviewed and approved by the institutional ethical review committee of Hanoi University of Public Health, Vietnam (IRB 400/2021/YTCC-HD3 and 45/2022/YTCC-HD3). Participants have consented their participation. Data were confidentially collected and analyzed. No compensation was provided for participation.

**Results**

**Participants’ Characteristics**

Table 1 provides a summary of participants’ characteristics, stratified by sex. Among 5890 survivors of COVID-19, the median age was 31 (IQR 23-40) years. More than half of them were female (n=3443, 58.5%), married (n=3402, 57.8%), and
had attained a university education or higher (n=3034, 51.5%). In total, 3475 (59%) participants were dependent workers, and 1928 (32.7%) were health care workers. Most participants belonged to the middle social status category (n=4822, 81.9%) and had no comorbidities other than COVID-19 (n=4177, 70.9%).

Table 1. Distribution of sociodemographic characteristics among 5890 survivors of COVID-19, stratified by sex.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All participants (N=5890)</th>
<th>Male participants (n=2447, 41.5%)</th>
<th>Female participants (n=3443, 58.5%)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), median (IQR)</td>
<td>31 (23-40)</td>
<td>30 (23-40)</td>
<td>32 (24-41)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2447 (41.5)</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Female</td>
<td>3443 (58.5)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Single</td>
<td>2346 (39.8)</td>
<td>1156 (47.2)</td>
<td>1190 (34.6)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>3402 (57.8)</td>
<td>1259 (51.5)</td>
<td>2143 (62.2)</td>
<td></td>
</tr>
<tr>
<td>Widow or divorce or separate</td>
<td>142 (2.4)</td>
<td>32 (1.3)</td>
<td>110 (3.2)</td>
<td></td>
</tr>
<tr>
<td>Education levels, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Illiterate or elementary</td>
<td>161 (2.7)</td>
<td>62 (2.5)</td>
<td>99 (2.9)</td>
<td></td>
</tr>
<tr>
<td>Secondary or high school</td>
<td>1264 (21.5)</td>
<td>569 (23.3)</td>
<td>695 (20.2)</td>
<td></td>
</tr>
<tr>
<td>Vocational or college</td>
<td>1431 (24.3)</td>
<td>425 (17.4)</td>
<td>1006 (29.2)</td>
<td></td>
</tr>
<tr>
<td>University or higher</td>
<td>3034 (51.5)</td>
<td>1391 (56.8)</td>
<td>1643 (47.7)</td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.003</td>
</tr>
<tr>
<td>Dependent worker</td>
<td>3475 (59)</td>
<td>1382 (56.5)</td>
<td>2093 (60.8)</td>
<td></td>
</tr>
<tr>
<td>Independent worker</td>
<td>1461 (24.8)</td>
<td>656 (26.8)</td>
<td>805 (23.4)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>954 (16.2)</td>
<td>409 (16.7)</td>
<td>545 (15.8)</td>
<td></td>
</tr>
<tr>
<td>Health care workers, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>3962 (67.3)</td>
<td>1790 (73.2)</td>
<td>2172 (63.1)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1928 (32.7)</td>
<td>657 (26.8)</td>
<td>1271 (36.9)</td>
<td></td>
</tr>
<tr>
<td>Social status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Low</td>
<td>767 (13)</td>
<td>302 (12.3)</td>
<td>465 (13.5)</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>4822 (81.9)</td>
<td>1974 (80.7)</td>
<td>2848 (82.7)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>301 (5.1)</td>
<td>171 (7)</td>
<td>130 (3.8)</td>
<td></td>
</tr>
<tr>
<td>Comorbidity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.09</td>
</tr>
<tr>
<td>No</td>
<td>4177 (70.9)</td>
<td>1764 (72.1)</td>
<td>2413 (70.1)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1713 (29.1)</td>
<td>683 (27.9)</td>
<td>1030 (29.9)</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Underweight</td>
<td>709 (12)</td>
<td>133 (5.4)</td>
<td>576 (16.7)</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>3573 (60.7)</td>
<td>1261 (51.5)</td>
<td>2312 (67.2)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>1043 (17.7)</td>
<td>671 (27.4)</td>
<td>372 (10.8)</td>
<td></td>
</tr>
<tr>
<td>Obese</td>
<td>565 (9.6)</td>
<td>382 (15.6)</td>
<td>183 (5.3)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>P value from Wilcoxon rank-sum, Pearson chi-square, and Fisher exact test compared the distribution of covariates between male and female participants.

<sup>b</sup>N/A: not applicable.

There were significant differences in the distribution of all covariates between male and female participants except for comorbidities. A higher percentage of female participants were underweight (female: 576/3443, 16.7% vs male: 133/2447, 5.4%). In contrast, male participants had higher rates of overweight (male: 671/2447, 27.4% vs female: 372/3443, 10.8%) and obesity (male: 382/2447, 15.6% vs female: 183/3443, 5.3%) than female participants.

**Distribution of Unhealthy Lifestyle Behaviors**

Table 2 displays the proportion of unhealthy lifestyle behaviors in survivors of COVID-19, stratified by sex. Among both sexes,
sedentary behaviors were the most prevalent behavior (male: 1871/2447, 76.5% and female: 2629/3443, 76.4%), followed by unhealthy diet (male: 1508/2447, 61.6% and female: 2109/3443, 61.3%). The prevalence of smoking and drinking was significantly higher among male participants compared to their female counterparts (307/2447, 12.5% vs 12/3443, 0.3%; \( P < .001 \) and 1217/2447, 49.7% vs 816/3443, 23.7%; \( P < .001 \), respectively). Conversely, the proportion of physical inactivity among female participants was higher than among male participants (female: 1928/3443, 56% vs male: 988/2447, 40.4%; \( P < .001 \).

### Table 2. Distribution of 5 unhealthy lifestyle behaviors, stratified by sex (N=5890).

<table>
<thead>
<tr>
<th>Unhealthy lifestyle behaviors</th>
<th>All participants (N=5890), n (%)</th>
<th>Male participants (n=2447, 41.5%), n (%)</th>
<th>Female participants (n=3443, 58.5%), n (%)</th>
<th>( P ) value (^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cigarette smoking</td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>5571 (94.6)</td>
<td>2140 (87.5)</td>
<td>3431 (99.7)</td>
<td></td>
</tr>
<tr>
<td>Smokers</td>
<td>319 (5.4)</td>
<td>307 (12.5)</td>
<td>12 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>3857 (65.5)</td>
<td>1230 (50.3)</td>
<td>2627 (76.3)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2033 (34.5)</td>
<td>1217 (49.7)</td>
<td>816 (23.7)</td>
<td></td>
</tr>
<tr>
<td>Unhealthy diet</td>
<td></td>
<td></td>
<td></td>
<td>.77</td>
</tr>
<tr>
<td>No</td>
<td>2273 (38.6)</td>
<td>939 (38.4)</td>
<td>1334 (38.7)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3617 (61.4)</td>
<td>1508 (61.6)</td>
<td>2109 (61.3)</td>
<td></td>
</tr>
<tr>
<td>Physical inactivity</td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>2974 (50.5)</td>
<td>1459 (59.6)</td>
<td>1515 (44)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2916 (49.5)</td>
<td>988 (40.4)</td>
<td>1928 (56)</td>
<td></td>
</tr>
<tr>
<td>Sedentary behavior</td>
<td></td>
<td></td>
<td></td>
<td>.93</td>
</tr>
<tr>
<td>No</td>
<td>1390 (23.6)</td>
<td>576 (23.5)</td>
<td>814 (23.6)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4500 (76.4)</td>
<td>1871 (76.5)</td>
<td>2629 (76.4)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) \( P \) value from Pearson chi-square and Fisher exact test compared the distribution of covariates between male and female participants.

Table 3 illustrates the distribution of co-occurring unhealthy lifestyle behaviors among male and female participants. Almost all survivors of COVID-19 had participated in 1 behavior (male: 2432/2447, 99.4% and female: 3411/3443, 99.1%). In all categories, there was a higher prevalence of male participants than female participants who had 2 (male: 1918/2447, 78.4% vs female: 2551/3443, 74.1%), 3 (male: 1091/2447, 44.6% vs female: 1243/3443, 36.1%), 4 (male: 331/2447, 13.5% vs female: 282/3443, 8.2%), and all 5 (male 61/2447, 2.5% vs female: 3/3443, 0.1%) behaviors. The absolute and relative differences between the prevalence of male and female participants also increased as the number of behaviors increased. The ratio of male to female participants participating in 1 behavior was 1.003. This ratio increased gradually with a higher number of behaviors and surged to 25 at 5 unhealthy behaviors.

Figure 3 illustrates different combinations of unhealthy behaviors, stratified by sex. The most frequent combination observed in 1123 (45.9%) of 2447 male participants and 1584 (46%) of 3443 female participants was sedentary behavior and unhealthy diet. Male participants showed a high prevalence of combining alcohol consumption with sedentary behaviors (949/2447, 38.8%) and an unhealthy diet (861/2447, 35.2%), followed by the combination of all 3 abovementioned behaviors (663/2447, 27.1%). On the other hand, female participants tended to combine physical inactivity with sedentary behavior (1305/3443, 37.9%) or an unhealthy diet (1260/3443, 36.6%), as well as all 3 behaviors (885/3443, 25.7%).

---

https://publichealth.jmir.org/2024/1/e50189
Table 3. Distribution of unhealthy lifestyle behavior co-occurrence, stratified by sex (N=5890).

<table>
<thead>
<tr>
<th>Unhealthy lifestyle behavior co-occurrence</th>
<th>All participants (N=5890), n (%)</th>
<th>Male participants (n=2447, 41.5%), n (%)</th>
<th>Female participants (n=3443, 58.5%), n (%)</th>
<th>% Difference (absolute; relative)</th>
<th>P value(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>47 (0.8)</td>
<td>15 (0.6)</td>
<td>32 (0.9)</td>
<td></td>
<td>0.3 (1.003)</td>
</tr>
<tr>
<td>Yes</td>
<td>5843 (99.2)</td>
<td>2432 (99.4)</td>
<td>3411 (99.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two risks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.3 (1.06)</td>
</tr>
<tr>
<td>No</td>
<td>1421 (24.1)</td>
<td>529 (21.6)</td>
<td>892 (25.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4469 (75.9)</td>
<td>1918 (78.4)</td>
<td>2551 (74.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three risks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.5 (1.24)</td>
</tr>
<tr>
<td>No</td>
<td>3556 (60.4)</td>
<td>1356 (55.4)</td>
<td>2200 (63.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2334 (39.6)</td>
<td>1091 (44.6)</td>
<td>1243 (36.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four risks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.3 (1.65)</td>
</tr>
<tr>
<td>No</td>
<td>5277 (89.6)</td>
<td>2116 (86.5)</td>
<td>3161 (91.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>613 (10.4)</td>
<td>331 (13.5)</td>
<td>282 (8.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Five risks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.4 (25.0)</td>
</tr>
<tr>
<td>No</td>
<td>5826 (98.9)</td>
<td>2386 (97.5)</td>
<td>3440 (99.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>64 (1.1)</td>
<td>61 (2.5)</td>
<td>3 (0.1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)P value from Pearson chi-square and Fisher exact test compared the distribution of covariates between male and female participants.

Figure 3. UpSet diagrams for different combinations of unhealthy lifestyle behaviors of participants, stratified by sex. This figure does not illustrate the combination with a prevalence of less than 5% in both sexes. The table underneath depicts the co-occurrence of unhealthy lifestyle behavior combination, where the rows represent different sets of combinations and the columns represent the number of male and female participants who had these combinations.

The Cluster of Unhealthy Lifestyle Behaviors

Table 4 reports the model fit indices of LCA models from 2 to 6 classes fitting to 5 behavior indicators. The 2-class model has the lowest AIC and BIC. In addition, the 2-class model is easier to interpret than the other models. Therefore, we proceed with the 2-class model.
Table 4. The model fit indices of latent class analysis models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Two-class model</th>
<th>Three-class model</th>
<th>Four-class model</th>
<th>Five-class model</th>
<th>Six-class model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood</td>
<td>−16,050.62</td>
<td>−15,787.33</td>
<td>−15,769.49</td>
<td>−15,768.81</td>
<td>Not convergent</td>
</tr>
<tr>
<td>AIC</td>
<td>31,584.99</td>
<td>31,608.67</td>
<td>32,123.24</td>
<td>31,595.62</td>
<td>Not convergent</td>
</tr>
<tr>
<td>BIC</td>
<td>31,722.26</td>
<td>32,196.75</td>
<td>31,738.67</td>
<td>31,789.39</td>
<td>Not convergent</td>
</tr>
</tbody>
</table>

*AIC: Akaike information criterion.*

*BIC: Bayesian information criterion.*

Table 5 presents the 2-class membership, distributed among the prevalence of 5 unhealthy lifestyle behaviors and stratified by sex. The 2 classes were labeled based on the distribution of each behavior in each class. The “less unhealthy” group included 64.7% (3810/5890) of total participants, and among them, there were 48.4% (1184/2447) of male participants and 76.3% (2626/3443) of female participants. Sedentary behavior was the most common behavior exhibited in both male and female participants in this group (887/1184, 74.9% and 1954/2626, 74.4%, respectively), followed by an unhealthy diet and physical inactivity. None of the participants in this membership had participated in 4 or 5 behaviors, and none of them reported any alcohol consumption. More than 43% of both male (521/1184, 43.8%) and female (1143/2626, 43.5%) “less unhealthy” participants engaged in 2 behaviors.

Table 5. Distribution of unhealthy lifestyle behaviors among survivors of COVID-19 by 2-class memberships.

<table>
<thead>
<tr>
<th>Unhealthy lifestyle behaviors</th>
<th>All participants (N=5890)</th>
<th>Male participants (n=2447)</th>
<th>Female participants (n=3443)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less unhealthy (n=3810, 64.7%), n (%)</td>
<td>More unhealthy (n=2080, 35.3%), n (%)</td>
<td>Less unhealthy (n=1263, 51.6%), n (%)</td>
</tr>
<tr>
<td>Cigarette smoking</td>
<td>22 (0.6)</td>
<td>297 (14.3)</td>
<td>19 (1.6)</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>0 (0)</td>
<td>2036 (97.7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Unhealthy diet</td>
<td>2119 (55.6)</td>
<td>1498 (72)</td>
<td>602 (50.7)</td>
</tr>
<tr>
<td>Physical inactivity</td>
<td>1977 (51.9)</td>
<td>939 (45.1)</td>
<td>509 (43)</td>
</tr>
<tr>
<td>Sedentary behaviors</td>
<td>2841 (74.6)</td>
<td>1659 (79.8)</td>
<td>887 (74.9)</td>
</tr>
</tbody>
</table>

The “more unhealthy” group included 35.3% (2080/5890) of total participants, and among them, there were 51.6% (1263/2447) of male participants and 23.8% (819/3443) of female participants. The most prevalent behavior practiced in both male and female participants was alcohol consumption (1217/1263, 96.4% and 816/819, 99.9%, respectively), followed by sedentary behaviors (984/1263, 77.9% and 675/819, 82.6%, respectively). Male participants in the “more unhealthy” group exhibited a significantly higher prevalence of cigarette smoking compared to their female counterparts (288/1263, 22.8% vs 9/819, 1.1%, respectively). There were 599 (47.4%) of 1263 male participants and 359 (43.9%) of 819 female participants in this group who participated in 3 unhealthy behaviors. Approximately 5% (61/1263) of male participants engaged in all 5 behaviors, while this proportion was 0.4% (3/819) among female participants in the same class.

Table 6 displays the distribution of covariates by class membership. There were significant differences in the distribution of all covariates between “less unhealthy” and “more unhealthy” groups. Nearly half (1010/2080, 48.6%) of the participants in the “more unhealthy” group were single, 60.7% (1263/2080) of them were male, 67.3% (1400/2080) of them had at least a university education, and 71.9% (1495/2080) of “more unhealthy” participants were dependent workers. Regarding “less unhealthy” group, nearly 69% (2626/3810) of the participants were female, and 62.4% (2377/3810) of them were married. More than half (314/2080, 52%) of “less unhealthy” participants were dependent workers, and 42.9% (1364/3810) of them had at least a university education. The prevalence of obesity was almost 3 times higher in the “more unhealthy” group compared to the “less unhealthy” group (314/2080, 15.1% vs 251/3810, 6.6%, respectively).

Table 6. Distribution of covariates by class membership.

The prevalence of obesity was almost 3 times higher in the “more unhealthy” group compared to the “less unhealthy” group (314/2080, 15.1% vs 251/3810, 6.6%, respectively).
Table 6. Distribution of covariates by class membership (N=5890).

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Less unhealthy group (n=3810, 64.7%)</th>
<th>More unhealthy group (n=2080, 35.3%)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), median (IQR)</td>
<td>32.0 (23.0-44.0)</td>
<td>29.0 (23.0-37.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>1184 (31.1)</td>
<td>1263 (60.7)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2626 (68.9)</td>
<td>817 (39.3)</td>
<td></td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Single</td>
<td>1336 (35.1)</td>
<td>1010 (48.6)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>2377 (62.4)</td>
<td>1025 (49.3)</td>
<td></td>
</tr>
<tr>
<td>Widow or divorce or separate</td>
<td>97 (2.5)</td>
<td>45 (2.2)</td>
<td></td>
</tr>
<tr>
<td>Education levels, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Illiterate or elementary</td>
<td>140 (3.7)</td>
<td>21 (1)</td>
<td></td>
</tr>
<tr>
<td>Secondary or high school</td>
<td>1025 (26.9)</td>
<td>239 (11.5)</td>
<td></td>
</tr>
<tr>
<td>Vocational or college</td>
<td>1011 (26.5)</td>
<td>420 (20.2)</td>
<td></td>
</tr>
<tr>
<td>University or above</td>
<td>1634 (42.9)</td>
<td>1400 (67.3)</td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Dependent worker</td>
<td>1980 (52)</td>
<td>1495 (71.9)</td>
<td></td>
</tr>
<tr>
<td>Independent worker</td>
<td>1139 (29.9)</td>
<td>322 (15.5)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>691 (18.1)</td>
<td>263 (12.6)</td>
<td></td>
</tr>
<tr>
<td>Health care workers, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>2699 (70.8)</td>
<td>1263 (60.7)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1111 (29.2)</td>
<td>817 (39.3)</td>
<td></td>
</tr>
<tr>
<td>Social status, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Low</td>
<td>529 (13.9)</td>
<td>238 (11.4)</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>3115 (81.8)</td>
<td>1707 (82.1)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>166 (4.4)</td>
<td>135 (6.5)</td>
<td></td>
</tr>
<tr>
<td>Comorbidity, n (%)</td>
<td></td>
<td></td>
<td>.004</td>
</tr>
<tr>
<td>No</td>
<td>2654 (69.7)</td>
<td>1523 (73.2)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1156 (30.3)</td>
<td>557 (26.8)</td>
<td></td>
</tr>
<tr>
<td>Nutritional status, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Underweight</td>
<td>495 (13)</td>
<td>214 (10.3)</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>2453 (64.4)</td>
<td>1120 (53.8)</td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>611 (16)</td>
<td>432 (20.8)</td>
<td></td>
</tr>
<tr>
<td>Obese</td>
<td>251 (6.6)</td>
<td>314 (15.1)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>P value from Pearson chi-square and Fisher exact test.

Factors Related to the More Unhealthy Group

Table 7 presents multivariable logistics regression to examine the association between covariates and being in the “more unhealthy” group among male and female participants. In both sexes, younger individuals had lower odds of being in the “more unhealthy” group (male: P=.001; OR 0.99, 95% CI 0.98-0.99 and female: P<.001; OR 0.98, 95% CI 0.97-0.99). Marital status showed a significant association with the “more unhealthy” group in both models. Married male participants had higher odds of being in the “more unhealthy” group compared to single ones (OR 1.45, 95% CI 1.14-1.85), while female participants exhibited lower odds (OR 0.65, 95% CI 0.51-0.83). Independent workers and unemployed participants had lower odds of being in the “more unhealthy” group compared to dependent workers of both sexes. In both models, higher educational levels were found to be associated with increased odds of belonging to the “more unhealthy” group.
### Table 7. Multivariable models of factors related to more unhealthy group among survivors of COVID-19.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Male participants</th>
<th>Female participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>P value(^a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.99 (0.98-0.99)</td>
<td>.001</td>
</tr>
<tr>
<td>Marital status (reference: single)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1.45 (1.14-1.85)</td>
<td>.006</td>
</tr>
<tr>
<td>Widow or divorce or separate</td>
<td>2.11 (0.94-4.75)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Education levels (reference: illiterate or elementary)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary or high school</td>
<td>1.05 (0.59-1.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vocational or college</td>
<td>1.92 (1.06-3.47)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>University or above</td>
<td>2.4 (1.34-4.31)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Current employment status (reference: dependent workers)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent workers</td>
<td>0.65 (0.51-0.82)</td>
<td>.52</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.55 (0.42-0.72)</td>
<td>.36</td>
</tr>
<tr>
<td>Being health care workers (reference: no)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.08 (0.86-1.36)</td>
<td>.96</td>
</tr>
<tr>
<td>Social status (reference: low)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>1.03 (0.80-1.33)</td>
<td>.98</td>
</tr>
<tr>
<td>High</td>
<td>1.06 (0.71-1.58)</td>
<td>.98</td>
</tr>
<tr>
<td>Comorbidity (reference: no)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.21 (0.99-1.47)</td>
<td>.05</td>
</tr>
<tr>
<td>Nutritional status (reference: normal)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>0.6 (0.41-0.89)</td>
<td>1.11 (0.89-1.39)</td>
</tr>
<tr>
<td>Overweight</td>
<td>1.04 (0.85-1.26)</td>
<td>1.42 (1.07-1.87)</td>
</tr>
<tr>
<td>Obese</td>
<td>2.03 (1.58-2.62)</td>
<td>1.73 (1.21-2.47)</td>
</tr>
</tbody>
</table>

\(^a\)OR: odds ratio.  
\(^b\)P value from likelihood ratio test.

Furthermore, individuals who are overweight or obese were more likely to fall into the “more unhealthy” category when compared to participants with a normal weight. Among male participants, the odds of being in the “more unhealthy” group slightly increased for overweight individuals who are overweight (OR 1.04, 95% CI 0.85-1.26), while significantly higher odds were found for individuals who are obese (OR 2.03, 95% CI 1.58-2.62). For female participants, both individuals who are overweight (OR 1.42, 95% CI 1.07-1.87) and obese (OR 1.73, 95% CI 1.21-2.47) had an increased likelihood of being in the “more unhealthy” group. Female participants who are underweight showed a higher likelihood of falling into the “more unhealthy” group (OR 1.11, 95% CI 0.89-1.39), but this was not the case for male participants (OR 0.6, 95% CI 0.41-0.89). Being a health care worker, level of social status, and having comorbidities did not show a statistically significant association with higher odds of being in the “more unhealthy” group.

### Discussion

#### Principal Findings

This national study examines the prevalence and clustering patterns of 5 unhealthy lifestyle behaviors among survivors of COVID-19 in Vietnam. This study provides a distinct pattern of unhealthy behaviors exhibited by male and female individuals after their COVID-19 recovery. We found that sedentary behaviors and unhealthy diets are common unhealthy behaviors in both sexes. While physical inactivity was more commonly practiced among female participants, male participants tended to engage in smoking and alcohol consumption. Nearly all participants had engaged in at least 1 unhealthy behavior, and male participants had a higher likelihood of engaging in multiple behaviors than female participants. Male participants also tended to practice alcohol intake with others such as sedentary behaviors and unhealthy diet, whereas female participants engaged in physical inactivity with sedentary behaviors and unhealthy diet.

Two unhealthy behavior classes were identified using LCA, with the “more unhealthy” group featured by a greater...
proportion of male participants. Certain common factors associated with this group were observed in both male and female participants, including older age, dependent employment, high educational levels, and obesity. However, sex-specific differences were found. For male participants, being married was an additional factor linked to a higher likelihood of being in the “more unhealthy” group. On the other hand, for female participants, being single and underweight were additional factors associated with this group.

**Prevalence of Unhealthy Lifestyle Behaviors by Sex**

The most prevalent unhealthy behaviors among survivors of COVID-19 were sedentary behaviors and unhealthy diets, which is consistent with findings from other studies among students and younger adults [23,34,48]. While there were significant differences between the percentage of male and female participants engaged in almost all behaviors, there were no significant differences between the prevalence of unhealthy diet and sedentary behavior in male and female participants in our sample. Our findings are in line with those from a study conducted among 1058 first-year students in Greece in 2016, which also found no significant differences in the prevalence of sedentary behavior (defined by screen time of more than 2 hours) and fruit and vegetable intake between male and female participants [23]. Since the study collected data when several provinces and cities were under lockdown across Vietnam, one explanation for the absence of these significant sex differences in terms of sedentary behavior and an unhealthy diet could be that both sexes were impacted similarly in this period. The lockdown might have caused significant changes in daily routines and lifestyles for the whole population, such as reduced physical activity, increased sedentary behavior, and changes in dietary patterns. Similar changes were observed in previous studies that took place during the cordon sanitaire period globally [49-51]. These changes can then exacerbate the prevalence of both sexes engaging in unhealthy behaviors. Indeed, there was a considerable increase in snacking and meal numbers or in unfavorable food choices and dietary habits during the COVID-19 lockdown reported in 2020, and there were no significant differences in sex participation rate as well [52].

Our evidence highlights the significant sex disparities in cigarette smoking and alcohol consumption among survivors of COVID-19, skewing higher in male participants. This observation aligns with prior research, where male participants tend to engage in more hazardous behaviors than female participants [53-56]. Male participants were also more likely to combine alcohol consumption with other factors such as sedentary lifestyles and unhealthy diets. With nearly 30% (683/2447) of male participants in our sample with chronic conditions, the high prevalence of co-occurring unhealthy behaviors coupled with the possible long-term effect of COVID-19 infection might increase the risk of NCDs in this population [57]. The observed sex differences in cigarette smoking and alcohol consumption in this study might be explained by a range of social, cultural, and economic factors. Vietnamese tradition perceived smoking and drinking as a part of the male gender role, used to display masculinity and establish social connections [30], while women are generally not discouraged to engage in these behaviors due to gender role expectations. Following such a drastic health event such as a cancer diagnosis or in our case—COVID-19 infection, lifestyle behaviors might also differ between sexes [58]. While women are more likely to adhere to health education and healthy behaviors after diagnosis, men are less likely to change their behaviors [28]. In addition, possible stress and anxiety induced by both the infection and the pandemic might cause men to smoke or drink more as a coping mechanism [59]. To address this gap, targeted interventions for male survivors of COVID-19 are recommended. These interventions may include a health education program for health problems following COVID-19 infection focused on the consequences associated with smoking and alcohol intake or a health promotion campaign highlighting the double-burden aspects of long COVID-19 symptoms and unhealthy behaviors.

**Clustering of Multiple Unhealthy Lifestyle Behaviors by Sex**

We highlighted significant differences in the number and combination of lifestyle-compromising behaviors between male and female participants. Surprisingly, the prevalence of male participants engaging in 4 or 5 unhealthy lifestyle behaviors simultaneously was nearly twice and 25 times higher compared to female participants. These findings are consistent with evidence from the adult population in India, which suggests that more adult male participants tend to exhibit the clustering of multiple NCD risk factors than female participants [26]. However, these findings contradict previous research indicating that female participants tend to engage in multiple unhealthy behaviors more frequently than male participants [23,34]. For example, in a study involving 3495 adolescents in Vietnam, it was found that the prevalence of practicing 2 unhealthy behaviors was similar between male and female participants. However, there was a sex discrepancy in the co-occurrence of 3 factors (male: 27.3% and female: 31.5%) and 4 factors (male: 13.7% and female: 15.1%) [34]. Similarly, a study conducted with 1058 first-year students in Greece found that the co-occurrence of 4 unhealthy factors was twice as high in female participants compared to male participants (male: 2.9% and female: 5.2%) [23]. It is important to note that the percentages of co-occurring behaviors may vary across studies due to differences in the number and types of behaviors measured as well as the specific cut-off points used. This is particularly relevant in cases where clear guidelines are lacking [32,60].

Among survivors of COVID-19 in Vietnam, we identified 2 distinct behavior patterns: “less unhealthy” and “more unhealthy” groups. These groups differed significantly in various factors. In the “less unhealthy” group, which represented around 65% (3810/5890) of the sample, a sedentary lifestyle was most prevalent, and there were a higher proportion of female participants. Conversely, the “more unhealthy” group, constituting over one-third of the population, was predominantly male and exhibited higher rates of alcohol intake and sedentary behavior. Additionally, the 2 groups differed in the number of concurrent unhealthy lifestyle behaviors. None of the “less unhealthy” participants reported engaging in more than 3 unhealthy behaviors simultaneously, while over 70% (male: 930/1263, 73.6% and female: 642/819, 78.4%) of the “more unhealthy” ones reported 3 or more behaviors. These distinctions
were also reflected in the sex distribution. This finding is in line with a previous study documenting the clustering of health-related behaviors among Indian adults and US young adults [26], in which the authors also found a higher percentage of men grouped in the higher-risk cluster. Overall, the findings reaffirm that (1) there were distinct patterns of unhealthy lifestyle behaviors practiced by survivors of COVID-19, and (2) these patterns differ significantly by sex.

Our findings also indicate that several factors were associated with the “more unhealthy” pattern with sex-specific differences. In this study, male participants who were married displayed a higher likelihood of engaging in unhealthy behaviors compared to those who were single. This contradicts previous research conducted on healthy behaviors among young Korean adults [61] and survivors of cancer [62], which found that unmarried men or men without a partner were more likely to practice unhealthy behaviors compared to their married counterparts. Our evidence suggests that dependent workers, especially male participants, were more likely to engage in unhealthy behaviors. This aligns with previous research indicating that men in dependent jobs, particularly in Asian contexts, tend to participate in social gatherings at work that often involve smoking and drinking [63]. The drinking culture after work is a common practice in many Asian countries, including Vietnam [64], South Korea [65,66], and Japan [67]. On the other hand, self-employment was positively associated with healthier behaviors. This can be attributed to the flexibility self-employed individuals have in managing their work schedule, allowing them to allocate time for health promotion activities [68,69].

Additionally, we found that obesity was significantly associated with the “more unhealthy” group. This highlights the need to address the growing public health challenge of a higher prevalence of obese individuals, which has been exacerbated by the COVID-19 pandemic [70-73]. Given that our adults recovering from COVID-19 are already at risk of postrecovery complications, the engagement in unhealthy behaviors due to obesity conditions further places them in a vulnerable position. We strongly recommend targeted public health interventions and services for this population, focusing on promoting an active lifestyle and a balanced diet and discouraging unhealthy behaviors to regain and maintain their pre-COVID-19 infection health status.

Limitations and Strengths

This study has several limitations that should be acknowledged. First, the reliance on self-reported data for assessing lifestyle behaviors introduces the possibility of information bias. Participants may have underreported or overreported their engagement in certain behaviors. Second, the cross-sectional design of this study restricts our ability to establish causal relationships between lifestyle behaviors and different demographic factors. Third, the use of web-based surveys may have introduced selection bias, as only individuals who were contactable by phone were included in the sample, and the overall response rate was 50.1% (5890/11,761). It is important to consider the generalizability of our findings, as they may not apply to other vulnerable populations experiencing different chronic health impacts, given the variations in definitions, methodologies, and targeted populations in measuring lifestyle behaviors [32,60]. Fourth, the changes in lifestyle behaviors before and amid the pandemic were not investigated in this study. Fifth, we were unable to access data on Vietnamese societal and cultural dynamics to provide implications for the observed sex differences in lifestyle behaviors. Future research should address these limitations by using other empirical measures of unhealthy lifestyle behaviors and using longitudinal study designs to provide more robust evidence.

Despite certain limitations, this study possesses notable strengths. First, the data were derived from a large population-based cohort, enhancing the representativeness of our findings. Additionally, we used validated questionnaires to assess unhealthy lifestyle behaviors and associated factors, bolstering the generalizability of our study and facilitating comparisons with prior research. Moreover, this study contributes to the existing body of knowledge by shedding light on the distinct patterns of lifestyle behaviors observed among adults recovering from COVID-19, including sex differences, through the advanced application of LCA [55]. As survivors of COVID-19 constitute a growing population facing various health challenges, this study provides valuable insights into their sex-specific engagement in unhealthy behaviors. This, in turn, highlights the importance for public health agencies and policy makers to recognize the inadvertent detrimental impact of the COVID-19 pandemic, as well as the infection itself, on individuals’ lifestyle behaviors and overall health outcomes.

Conclusions

The study revealed distinct patterns of unhealthy lifestyle behaviors among survivors of COVID-19 in Vietnam, with sedentary lifestyles and unhealthy diets being the most prevalent. Notably, these patterns differed significantly between sexes, with male participants exhibiting higher unhealthy behaviors and engaging in a greater number of unhealthy behaviors compared to female counterparts. These findings have important implications for public health initiatives, highlighting the need for tailored educational interventions that address sex-specific lifestyle behaviors. Specifically, efforts should be made to reduce unhealthy tendencies, particularly among married male adults recovering from COVID-19 who are obese, to promote a healthy and active life after COVID-19.

Acknowledgments

This work was supported by Thai Nguyen National Hospital; Military Hospital 103; Hai Phong University of Medicine and Pharmacy; University of Medicine and Pharmacy; Hue University; Can Tho University of Medicine and Pharmacy, Vietnam; and Taipei Medical University, Taiwan. The authors also appreciate and acknowledge the participation of patients from selected hospitals and health centers. This study was supported by Thai Nguyen National Hospital, National Science and Technology Council, Taiwan (MOST 111-2410-H-038-008-MY2), and Taipei Medical University (112-3805-005-111 and TMU108-AE1-B37).
Authors' Contributions
LTHL, TNAH, and TVD analyzed the data and drafted the paper. LTHL, TNAH, TTN, TDD, BND, KMP, VHV, LVP, LTHN, HCN, TVT, THN, ATN, HVN, PBN, HTTN, TTMP, TTL, TTPN, CQT, HLQ, KTN, and TVD contributed to conceptualization, investigation, methodology, validation, and revising the paper. LTHL, TNAH, TTN, BND, KMP, VHV, LVP, LTHN, TVT, THN, HVN, PBN, HTTN, TTMP, TTL, TTPN, and CQT conducted data curation. All authors approved the final version of the paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Participant distribution across study sites.

References
5. Lifestyle risk factors. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/nceh/track/Topics/LifestyleRiskFactors.htm#-r4=Four personal behaviors that can control and increase excessive alcohol use [accessed 2024-03-05]


36. Shams-White MM, Chui K, Deuster PA, McKeown NM, Must A. Investigating items to improve the validity of the five-item healthy eating score compared with the 2015 healthy eating index in a military population. Nutrients 2019;11(2):251 [FREE Full text] [doi: 10.3390/nu11020251] [Medline: 30678099]


Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
LCA: latent class analysis
NCD: noncommunicable disease
OR: odds ratio

©Lan T H Le, Thi Ngoc Anh Hoang, Tan T Nguyen, Tien D Dao, Binh N Do, Khue M Pham, Vinh H Vu, Linh V Pham, Lien T H Nguyen, Hoang C Nguyen, Tuan V Tran, Trung H Nguyen, Anh T Nguyen, Hoan V Nguyen, Phuoc B Nguyen, Hoai T T Nguyen, Thu T M Pham, Thu Y T Le, Thao T P Nguyen, Cuong Q Tran, Ha-Linh Quach, Kien T Nguyen, Tuyen Van Duong. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 02.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Digital Dietary Behaviors in Individuals With Depression: Real-World Behavioral Observation

Yue Zhu, PhD; Ran Zhang, PhD; Shuluoj Sun, MS; Yihui Sun, MS; Fay Womer, MD; Rongxun Liu, MS; Sheng Zeng, MS; Xizhe Zhang, PhD; Fei Wang, MD

1Early Intervention Unit, Department of Psychiatry, The Affiliated Brain Hospital of Nanjing Medical University, Nanjing, China
2Functional Brain Imaging Institute of Nanjing Medical University, Nanjing, China
3School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China
4School of Computer Science and Engineering, Northeastern University, Shenyang, China
5Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN, United States
6Henan Key Laboratory of Immunology and Targeted Drug, Henan Collaborative Innovation Center of Molecular Diagnosis and Laboratory Medicine, School of Laboratory Medicine, Xinxiang Medical University, Xinxiang, China

*these authors contributed equally

Corresponding Author:
Xizhe Zhang, PhD
Early Intervention Unit
Department of Psychiatry
The Affiliated Brain Hospital of Nanjing Medical University
264 Guangzhou Road, Gulou District, Nanjing City, China
Nanjing, 210000
China
Phone: 1 86 02583295953
Fax: 1 86 02583295953
Email: zhangxizhe@njmu.edu.cn

Abstract

Background: Depression is often accompanied by changes in behavior, including dietary behaviors. The relationship between dietary behaviors and depression has been widely studied, yet previous research has relied on self-reported data which is subject to recall bias. Electronic device–based behavioral monitoring offers the potential for objective, real-time data collection of a large amount of continuous, long-term behavior data in naturalistic settings.

Objective: The study aims to characterize digital dietary behaviors in depression, and to determine whether these behaviors could be used to detect depression.

Methods: A total of 3310 students (2222 healthy controls [HCs], 916 with mild depression, and 172 with moderate-severe depression) were recruited for the study of their dietary behaviors via electronic records over a 1-month period, and depression severity was assessed in the middle of the month. The differences in dietary behaviors across the HCs, mild depression, and moderate-severe depression were determined by ANCOVA (analyses of covariance) with age, gender, BMI, and educational level as covariates. Multivariate logistic regression analyses were used to examine the association between dietary behaviors and depression severity. Support vector machine analysis was used to determine whether changes in dietary behaviors could detect mild and moderate-severe depression.

Results: The study found that individuals with moderate-severe depression had more irregular eating patterns, more fluctuated feeding times, spent more money on dinner, less diverse food choices, as well as eating breakfast less frequently, and preferred to eat only lunch and dinner, compared with HCs. Moderate-severe depression was found to be negatively associated with the daily 3 regular meals pattern (breakfast-lunch-dinner pattern; OR 0.467, 95% CI 0.239-0.912), and mild depression was positively associated with daily lunch and dinner pattern (OR 1.460, 95% CI 1.016-2.100). These changes in digital dietary behaviors were able to detect mild and moderate-severe depression (accuracy=0.53, precision=0.60), with better accuracy for detecting moderate-severe depression (accuracy=0.67, precision=0.64).
Conclusions: This is the first study to develop a profile of changes in digital dietary behaviors in individuals with depression using real-world behavioral monitoring. The results suggest that digital markers may be a promising approach for detecting depression.

JMIR Public Health Surveill 2024 | vol. 10 | e47428
doi: 10.2196/47428

KEYWORDS
dietary behaviors; digital marker; depression; mental health; appetite disturbance; behavioral monitoring; eating pattern; electronic record; digital health; behavioral; surveillance

Introduction

The mental health of students has become the forefront of concerns, particularly since the onset of the COVID-19 pandemic. Approximately 45% of college students in China reported experiencing mental health issues during the outbreak [1]. Depression screening typically involves self-reported data, but there is a lack of objective markers to promptly identify individuals experiencing depression. Early identification and intervention are crucial for mitigating the impact of depression during critical periods for the academic and occupational functioning of students [2].

Appetite disturbance or changes in dietary behaviors are common symptoms of depression and may serve as objective indicators of the condition in a large population [3]. Dietary behavior can exert an influence on mental health through a variety of pathways, including circadian rhythms, oxidative stress, and the gut microbiota [4]. Based on the time of meals, intervals between meals, daily eating window, and food intake of the day, dietary behavior patterns can be categorized into morningness, intermediate, and eveningness chronotypes [5]. Eveningness chronotype commonly exhibits a higher tendency to skip breakfast, eat dinner later, and allocate a greater proportion of their daily food intake to later hours of the day [6]. Prior research has additionally demonstrated that the eveningness chronotype, coupled with social jetlag, constitutes a risk factor for depression [7,8]. Besides, individuals with more pronounced fluctuations in their eating windows often display heightened emotional vulnerability [9]. However, previous studies of diet relied on retrospective questionnaires or interviews, which may fail to accurately reflect real-world behavior [10]. Additionally, these methods typically assess only 1 aspect of dietary behavior, such as diet quality or eating habits [8,11,12]. To fully understand dietary behaviors, it is necessary to use multiple scales to assess multiple dimensions, such as diet quality, emotional eating, and chronotype of eating habits. However, using multiple scales can lead to participants taking too long to complete the questionnaire, reducing its validity. Furthermore, understanding daily behavior features requires repeated behavioral monitoring over an extended period, while retrospective reports from a single point in time may not accurately reflect true behavior [13].

The use of electronic platforms for behavioral monitoring allows for real-time assessment of human behavior and can trigger an alert if measured behavior deviates from healthy norms [14,15]. Additionally, these platforms enable the collection of large amounts of high-frequency, high-dimensional continuous data, which can be used to identify typical multidimensional behavior features over an extended period based on naturalistic situations [16,17]. The growing body of literature leveraging behavioral monitoring for depression prediction has gained traction, spurred by the profound shifts in lifestyle behavior patterns, especially during the COVID-19 pandemic [18-20]. Nonetheless, predominant inquiries have predominantly concentrated on probing the correlation between physical activity, social network engagement, and mental well-being facilitated through mobile devices [21-23]. However, an evident void persists in comprehensively exploring the nexus between the surveillance of dietary behavior patterns and depression. On university campuses, meals are often paid for using electronic transactions linked to a student account, providing the opportunity to collect digital dietary behavior data. In this study, real-world monitoring was used to track dietary behavior for a month, and data on time, expenditure, and location patterns were collected.

To the best of our knowledge, this is the first study to use electronic device–based behavioral monitoring instead of retrospective self-reported data to examine the digital dietary behaviors of individuals with depression compared with controls, and to investigate the relationship between these behaviors and depression. It is common for depression to occur alongside other symptoms, but the relationship between digital dietary behaviors and these comorbid symptoms of depression remains underexplored. Therefore, we also aim to determine whether these comorbid symptoms are associated with dietary behaviors and to clarify the role these symptoms play in the relationship between changes in dietary behavior and depression. The final data analysis will involve using digital dietary behavior features to detect depression.

Methods

Recruitment of Participants

A total of 3678 medical students from Xinxiang Medical University willingly engaged in this study, responding to the institution’s mental health survey notification. As part of this engagement, they underwent a cross-sectional mental status survey from October 6 to 12, 2020. Concurrently, during the survey, these participants also consented to furnish records of their eating behaviors for the period spanning from October 1 to 31, 2020. These records were sourced from electronic transactions linked to their respective student accounts. All participants completed questionnaires via WeChat (Tencent Corp) and signed web-based informed consent.

Web-Based Measurements

Basic sociodemographic characteristics, such as gender, age, BMI, and educational level, were collected using the WeChat
official account platform. All participants also completed the following psychological assessments: the Patient Health Questionnaire-9 (PHQ-9), the Generalized Anxiety Disorder Questionnaire-7 (GAD-7), the Perceived Stress Scale-14 (PSS-14), and the Insomnia Severity Index (ISI). All the details of these psychological assessments can be found in Multimedia Appendix 1.

**Study Participants and Inclusion Criteria**

Based on the outcomes of the psychological survey, the inclusion criteria for the mild depression group, the moderate-severe depression group, and the healthy control (HC) group, as well as the exclusion criteria, were defined as follows. Inclusion criteria were established as follows: individuals in the mild depression group had PHQ-9 scores between 5 and 9, while those in the moderate-severe depression group had PHQ-9 scores of 10 or higher. The HC group had PHQ-9 scores below 5, GAD-7 scores below 5, PSS-14 scores below 29, and ISI scores below 8. Exclusion criteria for all participants included PHQ-9 scores below 5 and any of the following conditions: GAD-7 scores of 5 or higher, PSS-14 scores of 29 or higher, or ISI scores of 8 or higher. The participant inclusion process for this study is depicted in Figure 1 (Step 1).

**Ethical Considerations**

Prior to their involvement in the study, all participants provided their informed consent through a formally endorsed consent form. The study was approved by the Biomedical Ethics Committee of Xinxiang Medical University (XYLL-2020235).

**Dietary Data Collection**

During the COVID-19 pandemic in China, students were required to stay on campus, resulting in most students eating at the campus cafeterias on a daily basis. These meals are often paid for using electronic transactions linked to a student’s account, and there are 3 cafeterias at the school. The dietary data preprocess can be found in Multimedia Appendix 1. Previous studies have shown that diet time, or chronotype, is related to circadian rhythms and mood [5], and location and periodicity of meals are related to depression symptoms [24]. Therefore, we evaluated 6 dietary behavior features, including time, location, expenditure, daily dietary items and frequency of meals, and all-day dietary behavior patterns, to objectively assess dietary behaviors in relation to depression in this study (Figure 1, Step 2).

**Calculation of Dietary Features**

**Time Patterns**

To determine the number of daily meals, we counted the number of records starting from the first recorded meal. Dietary behaviors within 2 hours of this initial record were part of a single meal, which could be further divided into up to 1, 2, or 3 meals per day: breakfast, lunch, and dinner. In addition, based on the cafeteria’s hours of operation, the 3 meals are divided into time slots as follows: breakfast from 6:30 AM to 8:30 AM, lunch from 11 AM to 1 PM, and dinner from 5:30 PM to 7:30 PM. If there were multiple electronic transactions within a single meal, we used the timestamp of the first transaction as the time of the meal for analysis.

**Time Intervals Between Meals**

We used the times of breakfast, lunch, and dinner calculated from the time patterns to determine the time intervals between each pair of meals: breakfast and lunch, lunch and dinner, and breakfast and dinner.

**Expenditure on Meals**

After dividing meals into breakfast, lunch, and dinner based on the time patterns, we calculated the total cost for each meal by...
summing up multiple electronic transactions for a single meal. Expenditure on a single meal may indicate an individual’s appetite as it reflects the amount of food purchased.

**Location Patterns**

To determine location patterns of dietary behavior, we calculated the frequency of visits to each cafeteria by counting total number of cafeterias visited in October and calculated the frequency of the student’s visits to the cafeteria for breakfast, lunch, and dinner using the following formula:

\[ \text{Frequency} = \frac{\text{Number of visits}}{\text{Total number of days}} \]

**Daily Dietary Diversity and Frequency of Meals**

The average number of different foods consumed per day in October was calculated as a measure of food diversity. The frequency of breakfast, lunch, and dinner in the month was also calculated as a measure of dietary habits. These measures were used to understand the dietary behaviors of participants in the study.

**All-Day Dietary Behavior Patterns**

People may have regular or irregular dietary behavior patterns, such as consistently or selectively eating breakfast, lunch, and dinner. There are 7 possible dietary behavior patterns: eating only breakfast, only lunch, only dinner, breakfast-lunch, lunch-dinner, and all 3 meals (breakfast-lunch-dinner pattern). All-day dietary behavior patterns are defined as the meals eaten by a participant on a given day. The frequency of a participant’s daily dietary behavior patterns can be calculated based on their dietary records for the entire month of October. To do this, the daily records are taken as a unit and the frequency of each dietary behavior pattern is calculated using the following formula:

\[ \text{Frequency} = \frac{\text{Number of days}}{\text{Total number of records}} \]

**Statistical Indices of Dietary Features**

Statistical indices were calculated for dietary features such as time patterns, intervals between meals, and expenditure on meals, including mean, median, median absolute deviation (MAD), and maximum and minimum values for the entire month of October. The study also separated weekday and weekend behavior by calculating dietary features separately for the 2 time periods. More information can be found in Table S1 in Multimedia Appendix 1.

**Statistical Analysis**

ANOVA and chi-square tests were used to analyze demographic and psychological characteristics. ANCOVA (analyses of covariance) was used to compare dietary behavior features among groups, with age, gender, BMI, and educational level as covariates. All variables in the 6 dietary features were transformed into Z-scores to give equal weight and minimize the impact of outliers in the study. Logistic regression analyses were used to examine associations between all-day dietary behavior patterns and depression severity (mild depression, and moderate-severe depression), as well as associations between other psychological symptoms in depression groups or all participants and all-day dietary behavior patterns, with age, gender, BMI, and educational level as covariates. The all-day dietary behavior patterns were categorized into tertiles: Rare (less than 1 SD), Normal (within 1 SD), and Always (more than 1 SD) in logistic regression analyses. Mediation analysis was performed using Model 4 with 1 independent variable (all-day dietary behavior patterns), 1 dependent variable (groups), and 1 moderator (other psychological symptoms) in the PROCESS. A bootstrapping procedure with 95% CIs was used to measure the moderating effect, with 5000 bootstrap samples. Age, gender, BMI, and educational level were included as covariates in the model. All analyses were conducted using SPSS (version 25.0; IBM Corp).

**Results**

**Demographic and Clinical Characteristics of Participants**

A total of 3310 students, consisting of 2222 HC, 916 mild depression, and 172 moderate-severe depression, met the inclusion criteria and were ultimately included in the study. And 368 students exhibited psychological symptoms, notably anxiety, insomnia, or abnormal stress, while concurrently not manifesting depressive symptoms. As a result, they were excluded from the study cohort. In the students with mild depression, the following rates were observed: 37.7% (n=345) with anxiety, 17.8% (n=163) with perceived abnormal stress, and 23.8% (n=218) with insomnia. In moderate-severe depressive students, the corresponding percentages were: 75% (n=129) with anxiety, 70.3% (n=121) with perceived abnormal stress, and 53.5% (n=92) with insomnia. The demographic and clinical characteristics of the participants can be found in Table 1.
Table 1. Demographic and clinical characteristics of the study population.

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>HC[^a]</th>
<th>Mild depression</th>
<th>Moderate-severe depression</th>
<th>F test or chi-square test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants, n</td>
<td>2222</td>
<td>916</td>
<td>172</td>
<td>N/A[^b]</td>
<td>N/A</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>19.46 (1.49)</td>
<td>19.48 (1.35)</td>
<td>19.64 (1.52)</td>
<td>1.209 (2)</td>
<td>.3</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td>74 (43.02)</td>
<td>315 (34.39)</td>
<td>74 (43.02)</td>
<td>7.631 (2)</td>
<td>.02</td>
</tr>
<tr>
<td>Male</td>
<td>864 (38.88)</td>
<td>315 (34.39)</td>
<td>98 (56.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1358 (61.12)</td>
<td>601 (65.61)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI, n (%)</td>
<td>22 (12.79)</td>
<td>142 (15.50)</td>
<td>281 (12.65)</td>
<td>4.783 (4)</td>
<td>.31</td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>281 (12.65)</td>
<td>142 (15.50)</td>
<td>22 (12.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.5-24</td>
<td>1625 (73.13)</td>
<td>653 (71.29)</td>
<td>126 (73.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;24</td>
<td>316 (14.22)</td>
<td>121 (13.21)</td>
<td>24 (13.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education level[^c], n (%)</td>
<td>2190 (98.56)</td>
<td>908 (99.13)</td>
<td>170 (98.84)</td>
<td>2.342 (4)</td>
<td>.67</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>16 (0.72)</td>
<td>3 (0.33)</td>
<td>1 (0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postgraduate</td>
<td>7 (0.32)</td>
<td>2 (0.22)</td>
<td>1 (0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9[^d] score, mean (SD)</td>
<td>1.42 (1.41)</td>
<td>6.58 (1.33)</td>
<td>12.94 (3.31)</td>
<td>700.113 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>GAD-7[^e] score, mean (SD)</td>
<td>0.49 (0.92)</td>
<td>3.60 (2.74)</td>
<td>7.47 (4.26)</td>
<td>1712.361 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>GAD-7&gt;5 (Yes), n (%)</td>
<td>0</td>
<td>345 (37.66)</td>
<td>129 (75)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>PSS-14[^f] score, mean (SD)</td>
<td>15.77 (6.79)</td>
<td>23.08 (6.50)</td>
<td>32.41 (6.63)</td>
<td>767.972 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>PSS score &gt; 28 (Yes), n (%)</td>
<td>0</td>
<td>163 (26.33)</td>
<td>121 (70.35)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>ISI[^g] score, mean (SD)</td>
<td>1.78 (1.86)</td>
<td>5.42 (3.46)</td>
<td>8.93 (5.79)</td>
<td>992.727 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ISI score &gt; 8 (Yes), n (%)</td>
<td>0</td>
<td>218 (23.80)</td>
<td>92 (53.49)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\[^a\]HC: healthy control.
\[^b\]N/A: not applicable.
\[^c\]Information that was missing for some participants.
\[^d\]PHQ-9: Patient Health Questionnaire-9.
\[^e\]GAD-7: Generalized Anxiety Disorder Questionnaire-7.
\[^f\]PSS-14: Perceived Stress Scale-14.
\[^g\]ISI: Insomnia Severity Index.

Dietary Features Among HC, Mild Depression, and Moderate-Severe Depression

The ANCOVA analyses showed that there were significant differences between the groups in terms of time patterns, the intervals between meals, expenditure on meals, daily dietary diversity and frequency of meals, and all-day dietary behavior patterns (details in Table S2 in Multimedia Appendix 1). There was no significant difference in location patterns among the 3 groups.

In terms of time patterns, post hoc analyses showed that compared with mild depression and HC, moderate-severe depression had significantly increased MAD of lunchtime (P=.04 and P=.004, separately), latest lunchtime (P=.02 and P=.007, separately) and MAD of dinner time (P=.02 and P=.01, separately) on weekdays, and there was no significant difference between mild depression and HC (Figure 2A). In terms of time intervals between meals, post hoc analyses that compared mild depression and HC, moderate-severe depression had increased MAD (P=.001 and P<.001, separately) and maximum (P=.02 and P=.005, separately) time intervals between lunch and dinner on weekdays, respectively, but mild depression and HC had no significant difference with each other (Figure 2B). On weekends, post hoc analyses that compared with HC, mild depression had a significant decrease in the mean time interval between breakfast and lunch (P=.01), while mild depression and moderate-severe depression had significant decreases in the maximum time interval between breakfast and lunch (P=.02 and P=.03 separately; Figure 2B). The MAD of lunchtime on weekdays was higher in moderate-severe depression (18 minutes 58 seconds) than mild depression (17 minutes 38 seconds) and HC (17 minutes 5 seconds). Similarly, the MAD of dinner time on a weekday was higher in moderate-severe depression (26
minutes 11 seconds) than in mild depression (23 minutes 39 seconds) and HC (23 minutes 26 seconds). Additionally, the MAD time interval between lunch and dinner on weekdays was larger in moderate-severe depression (34 minutes 20 seconds) than in mild depression (29 minutes 25 seconds) and HC (29 minutes 23 seconds; Table S2 in Multimedia Appendix 1).

These findings indicate a temporally erratic lunch and dinner dietary pattern in moderate-severe depression.

Figure 2. Alternations of dietary pattern features in mild depression and moderate-severe depression. a: $P < .05$; b: $P < .01$; c: $P < .001$; and d: significance at $P < .05$, after false discovery rate correction. The values of dietary pattern features in the graph were transformed into Z-scores. Bre: breakfast; Din: dinner; HC: healthy control; Lun: lunch; MAD: median absolute deviation; Max: maximum; MED: median; Min: minimum.

In terms of expenditure on meals, compared with HC, both students with mild and moderate-severe depression had significant increases in the mean ($P = .02$ and $P = .02$, separately) and median ($P = .002$ and $P = .006$, separately) of dinner expenditure on weekdays (Figure 2C). However, there was no significant difference between mild depression and moderate-severe depression. On weekends, the mild depressive students had a significant increase in the minimum lunch expenditure compared with HC ($P = .03$) and moderate-severe depression ($P = .01$), but HC and moderate-severe depression had no significant difference from each other (Figure 2C).

In terms of daily dietary diversity, post hoc analyses that moderate-severe depression had significant decreases compared with mild depression and HC both on weekdays ($P = .02$ and $P = .01$, separately) and weekends ($P = .03$ and $P = .01$, separately). However, there was no significant difference between mild depression and HC. We also found that compared with HC, moderate-severe depression had a significant reduction in breakfast frequency on weekdays ($P = .02$) and weekends ($P = .005$). Additionally, compared with mild depression, moderate-severe depression had a significant reduction in breakfast frequency on weekends ($P = .05$). However, there was no significant difference between mild depression and HC in breakfast frequency on weekdays and weekends (Figure 2D).

Finally, in terms of all-day dietary behavior patterns, compared with HC and mild depression, moderate-severe depression had a reduction in the breakfast-lunch-dinner pattern on weekdays ($P = .001$ and $P = .002$, separately) and enrichment in the lunch-dinner pattern on weekdays ($P = .005$ and $P = .03$ separately; Figure 2E). However, there was no significant difference between HC and mild depression. On weekends, we observed that moderate-severe depression had an enrichment in the lunch-dinner pattern compared with HC ($P = .008$; Figure 2E).

Associations Between All-Day Dietary Behavior Patterns and Depression

To gain a deeper understanding of the relationship between comprehensive indicators of all-day dietary behavior patterns and depression severity. Using multiple logistic regression, we analyzed the associations between alterations in all-day dietary behavior patterns and depression severity in Figures 3A and 3B. After adjusting for age, gender, BMI, and educational level, we found that the Normal and Always lunch-dinner patterns on weekdays were positively associated with mild depression (Exp(B), 95% CI 1.360, 1.050-1.761; 1.460, 1.016-2.100; respectively), and the Always breakfast-lunch-dinner pattern on weekdays was negatively associated with moderate-severe depression (Exp(B), 95% CI 0.467, 0.239-0.912). The comparison of tertile levels of these all-day dietary behavior patterns among 3 groups can be found in Table S3 in Multimedia Appendix 1, and the percent of tertile levels of these all-day dietary behavior patterns can be found in Multimedia Appendix Figure S1.
Figure 3. (A,B) Associations between all-day dietary behavior patterns and depression. (C) Associations between Bre-Lun-Din patterns on weekdays and comorbidity of depression and anxiety in individuals. (D) Associations between Bre-Lun-Din patterns on weekday and anxiety in individuals. (E) The mediating role of anxiety between groups (HC, mild depression, and moderate-severe depression) and Bre-Lun-Din patterns. a: $P<.05$, b: $P<.01$, and c: $P<.001$. AOR: adjusted odds ratio; Bre: breakfast; Din: dinner; HC: healthy control; Lun: lunch.

Associations Between All-Day Dietary Behavior Patterns and Other Clinical Symptoms

Our findings revealed a positive correlation between anxiety and adherence to the consistent always breakfast-lunch-dinner pattern on weekdays among individuals with depression as well as all participants (Figures 3C and 3D). However, no significant associations were discovered between other all-day dietary patterns and other clinical symptoms (Table S4 in Multimedia Appendix 1). In the mediation model, anxiety served as a mediator ($R^2=0.570$, $P<.001$). A bootstrapped 95% CI confirmed that the indirect effect of groups (HC, mild depression, and moderate-severe depression) had an impact of 0.042 that was produced by anxiety as a mediator on breakfast-lunch-dinner pattern (Figure 3E).

The Detection of Dietary Patterns for Depression

Our results showed that alterations in dietary behaviors of depression had an accuracy of 0.53, $F_1$-score of 0.52, precision of 0.60, recall of 0.62, and an area under curve of 0.59 for detecting mild depression and moderate-severe depression. For detecting moderate-severe depression specifically, the accuracy was 0.67, $F_1$-score was 0.60, precision was 0.64, recall was 0.65, and area under curve was 0.69 (Table 2).

Table 2. The detection of depression in classification models.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>$F_1$-score</th>
<th>Precision</th>
<th>Recall</th>
<th>AUCa</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCb versus mild depression versus moderate-severe depression</td>
<td>0.53</td>
<td>0.52</td>
<td>0.60</td>
<td>0.62</td>
<td>0.59</td>
</tr>
<tr>
<td>HC versus moderate-severe depression</td>
<td>0.67</td>
<td>0.60</td>
<td>0.64</td>
<td>0.65</td>
<td>0.69</td>
</tr>
</tbody>
</table>

aAUC: area under the curve.
bHC: healthy control.
Discussion

Principal Results

This study is the first to investigate digital dietary patterns of individuals with mild and moderate-severe depression using electronic device–based monitoring. The results indicate that individuals with moderate-severe depression exhibit more irregular eating time patterns, greater fluctuations in their feeding window, higher expenditure on dinner, lower food diversity, and a decreased frequency of consuming breakfast, as well as more irregular lunch-dinner patterns and less regular breakfast-lunch-dinner patterns compared with HC. The study also found that maintaining a regular breakfast-lunch-dinner pattern is negatively associated with moderate-severe depression, and maintaining an irregular lunch-dinner pattern is positively associated with mild depression. Additionally, the presence of anxiety is positively associated with the breakfast-lunch-dinner pattern, and the severity of anxiety has an indirect effect on the relationship between depression and the breakfast-lunch-dinner pattern. Importantly, the study suggests that digital dietary features can be used to detect depression, particularly moderate-severe depression, indicating that quantified digital behavior could be a promising approach to the detection of depression.

Comparison With Prior Work in Dietary Behavior of Depression

We found that individuals with moderate-severe depression exhibited a reduction in dietary diversity, a decreased frequency of consuming breakfast, and irregular timing for lunch and dinner. The loss of interest in pleasurable activities, including eating, is a core symptom of depression, which may explain the decreased dietary diversity in individuals with moderate-severe depression. The findings of decreased breakfast frequency in moderate-severe depression align with previous research, which has identified a significant association between skipped or infrequent breakfast and an increased risk for depression [25,26]. The frequency of lunch and dinner did not significantly contribute to depression [26,27]. This discrepancy may be attributed to the fact that individuals with depression tend to have worse moods in the morning, which may negatively impact their appetite for breakfast. Considering these findings, it is important to consider the specific meal that is skipped when implementing dietary interventions to prevent depression. The study found that individuals with mild and moderate-severe depression exhibited significant fluctuations in the timing of lunch and dinner, as well as an irregular time interval between these meals. These findings align with previous research that has identified the importance of the feeding window on mental health [28,29], with irregular eating time patterns being associated with an increased risk of mental health distress [30]. Additionally, this study is the first to investigate the expenditure on 3 meals in individuals with depression, which provides insight into their food intake and appetite for each meal. The results indicate that individuals with mild and moderate-severe depression spent more money on dinner, which is consistent with the observation that individuals with major depressive disorder tend to have higher food intake during dinner compared with breakfast and lunch [31]. Notably, this study is the first to report that depression is associated with a preference for higher expenditure on dinner, while also being accompanied by a low frequency of breakfast.

This study is the first to objectively quantify all-day dietary behavior patterns by integrating time patterns, frequency of meals, and other parameters, and it revealed that moderate-severe depression is associated with an increase in lunch-dinner patterns and a decrease in breakfast-lunch-dinner patterns. Previous research has demonstrated that consuming breakfast, lunch, and dinner every day can reduce the prevalence of first-onset depression in a 5-year follow-up [32]. The lunch-dinner pattern is associated with eating later in the day and a meta-analysis has indicated that depressed patients are more inclined to spend on food later in the day, a phenomenon known as eveningness chronotype [33]. The relationship between eveningness chronotype and depressive symptoms aligns with pre-existing theories of chronobiology, which suggest that circadian dysfunction can have adverse effects on psychological well-being [34,35]. Eveningness chronotype is associated with a higher likelihood of regularly skipping or postponing breakfast [36] and it is well recognized that eating breakfast plays an important role in lowering blood cortisol levels and disturbances in glucose metabolism, which may affect serotonin levels [37].

Prior research in the dietary patterns field, whether dietary nutrition, dietary frequency, or dietary chronotype studies, predominantly relied on questionnaires [38,39]. The use of electronic behavioral monitoring for evaluating dietary behavioral patterns is still in its nascent technological phase. The self-monitoring is a prevalent method for assessing electronic dietary behaviors in the study of mental health [40,41]. This approach predominantly entails participants proactively documenting their daily dietary intake using an electronic device, often an app. However, it is important to acknowledge that excessive self-recording of behaviors could potentially lead to fatigue and monotony. While a frequency ranging from 2 to 3 times per week proved to be acceptable and reasonable [42], it once again hinders the inherent issue of recall bias that was previously encountered. Moreover, the field of wearable device-based monitoring for tracking dietary behavior has witnessed notable advancements [43,44]. However, its application in the field of psychiatry remains unexplored to date. The study marks the pioneering use of objective electronic device–based monitoring, unveiling anomalous eating patterns among individuals with depression.

The Associations Between Dietary Behavioral Patterns and Depression

The study found that maintaining a lunch-dinner pattern was positively associated with mild depression. Furthermore, maintaining a breakfast-lunch-dinner pattern was negatively associated with moderate-severe depression. These findings align with previous research that has identified eating breakfast as a health-promoting behavior [45] and a positive association between skipping breakfast and depressive symptoms [46]. Additionally, it highlights the importance of meal substitution for regular eating patterns as a positive association with
emotional disorders [47]. Overall, these findings suggest that maintaining a regular dietary pattern can be considered a dietary strategy for depression prevention.

Furthermore, the study found that anxiety is positively associated with the breakfast-lunch-dinner pattern. We can infer from the result that individuals with anxiety tend to choose to regularly eat 3 meals in their daily life and that depression with increasing anxiety may lead to an increase in the breakfast-lunch-dinner eating behavior. Literature has also indicated that individuals with anxiety tend to engage in more overeating situations [48] and have a positive association between anxiety symptoms and emotional and external eating [49]. However, the study also indicates that those with depression tend to intake more in dinner and then postpone or skip breakfast the next day. The study also found that anxiety could mediate the relationship between depression and breakfast-lunch-dinner pattern, which indicates that students with anxiety tend to maintain a regular breakfast-lunch-dinner pattern rather than postponing or skipping breakfast. This highlights the need for careful consideration when implementing dietary interventions for the comorbidity of depression and anxiety.

The Potential Value of Digital Dietary Behaviors in Depression

This study is the first to use digital dietary behaviors based on real-world behavioral monitoring to detect mild depression and moderate-severe depression. The results indicate that digital dietary behaviors could distinguish between moderate-severe depression and HC. Depression is a disorder characterized by brain-based dysfunction that is expressed through behavioral changes. The diagnosis of depression traditionally relies on structured interviews or questionnaires, which are based on retrospective self-reports and the threshold scores of questionnaires. These methods can be prone to bias and subjectivity. The use of digital markers of continuous daily behavioral monitoring as an objective indicator to detect depression represents a promising supplementary approach [13].

In this study, students’ dietary behaviors were analyzed over a period of 1 month to draw a profile of dietary patterns in depression. This is the first exploration of the relationship between digital dietary behaviors and depression, and the use of digital dietary behaviors to detect depression. Another thing to watch out for is that the results indicate that a deep learning model of dietary behaviors from digital devices holds more accurate detection of moderate-severe depression than mild depression. This is likely due to the observation that moderate-severe depression presents more erratic dietary behaviors than mild depression and HCs. Overall, this study highlights the potential of digital dietary features as a promising manner in the detection of depression, particularly moderate-severe depression.

Limitations

There are several limitations in this study that should be acknowledged. First, the measurement of depression severity in this study was based on self-reported data, future studies should consider clinician-based rating scales to reduce bias. Additionally, this study used a cross-sectional survey of depression, which does not capture changes in symptomology over time. A longitudinal psychological survey combined with continuous daily behavioral monitoring could provide further insight into dietary progression markers related to the severity of depressive symptoms over time. Finally, the study did not assess daily dietary structure and nutrient intake.

Conclusions

This study represents a pioneering endeavor in objectively characterizing the digital dietary behaviors of individuals experiencing depression, using real-world monitoring as opposed to self-reported retrospective data. Our findings indicate that students with depression experience disruptions in various aspects, including time patterns, the intervals between meals, expenditure on meals, daily dietary diversity and frequency of meals, and all-day dietary behavior patterns. Notably, individuals with moderate-severe depression showcase greater irregularities in eating time patterns, fluctuated feeding windows, decreased food diversity, higher expenditure on dinner, and a preference for consuming only lunch and dinner. Furthermore, maintaining a regular breakfast-lunch-dinner eating pattern exhibits a negative correlation with moderate-severe depression. Keeping a lunch-dinner pattern is positively associated with mild depression. This research not only fills a critical gap in the existing academic literature but also sheds light on the promising potential of digital dietary behaviors as objective makers to the detection of depression, particularly moderate-severe depression.

Acknowledgments

This study was supported by grants from National Science Fund for Distinguished Young Scholars (81725005 to FW), NSFC-Guangdong Joint Fund (U20A6005 to FW), Jiangsu Provincial Key Research and Development Program (BE2021167 to FW), National Natural Science Foundation of China (62176129 to XZ), National Key Research and Development Program (2022YFC2405603 to XZ), and Henan Provincial Research and Practice Project for Higher Education Teaching Reform (2021SJGLX189Y to RL). The authors would like to thank all participants who took part in this study. The authors would also like to thank the support to recruitment by personnel at Xinxiang Medical University.

Data Availability

The data sets generated and analyzed during this study are available from the co-corresponding authors (FW: fei.wang@yale.edu; XZZ: zhangxizhe@njmu.edu.cn) on reasonable request.
Authors' Contributions
FW and XZZ are co-corresponding authors. XZZ and FW designed the study. RXL and YZ enrolled participants. YHS, YZ, SLY, and SZ analyzed data. YZ, RZ, and FYW wrote the paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental materials.

References


40. Hahn SL, Kaciroti N, Eisenberg D, Weeks HM, Bauer KW, Sonneville KR. Introducing dietary self-monitoring to undergraduate women via a calorie counting app has no effect on mental health or health behaviors: results from a randomized


44. Dimitratos SM, German JB, Schaefer SE. Wearable technology to quantify the nutritional intake of adults: validation study. JMIR Mhealth Uhealth 2020;8(7):e16405 [FREE Full text] [doi: 10.2196/16405] [Medline: 32706729]


Abbreviations

- **ANCOVA**: analyses of covariance
- **GAD-7**: Generalized Anxiety Disorder Questionnaire-7
- **HC**: healthy control
- **ISI**: Insomnia Severity Index
- **MAD**: median absolute deviation
- **PHQ-9**: Patient Health Questionnaire-9
- **PSS-14**: Perceived Stress Scale-14

© Yue Zhu, Ran Zhang, Shuluo Yin, Yihui Sun, Fay Womer, Rongxun Liu, Sheng Zeng, Xizhe Zhang, Fei Wang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 22.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Understanding Risk Factors for Oropharyngeal Gonorrhea Among Sex Workers Attending Sexual Health Clinics in 2 Australian Cities: Mixed Methods Study

Tiffany R Phillips1,2, BSc, PhD; Christopher K Fairley1,2, MD, PhD; Kate Maddaford3, MPH; Anna McNulty3, MBBS; Basil Donovan4, MD; Rebecca Guy4, PhD; Ruthy McIver3, MPH; Rebecca Wigan2, MPH; Rick Varma4, MBChB; Jason J Ong1,2, PhD; Denton Callander4,5,6, PhD; Gabrielle Skelsey7; Mish Pony8; Dylan O'Hara9; Jade E Bilardi1,2,10*, PhD; Eric PF Chow1,2,11*, PhD

1Monash University, Clayton, Australia
2Alfred Health, Melbourne Sexual Health Centre, Carlton, Australia
3Sydney Sexual Health Centre, Sydney, Australia
4The Kirby Institute, University of New South Wales, Sydney, Australia
5Operational Centre Geneva, Medecins Sans Frontiers, Mombasa, Kenya
6SexTech Lab, The New School, New York, NY, United States
7Resourcing Health & Education (RhED), St Kilda, Australia
8Scarlet Alliance, Australian Sex Workers Association, Sydney, Australia
9Vixen, Victoria's Peer Sex Worker Organisation, Melbourne, Australia
10Department of General Practice, University of Melbourne, Melbourne, Australia
11Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

* these authors contributed equally

Corresponding Author:
Tiffany R Phillips, BSc, PhD
Monash University
Wellington Road
Clayton, 3800
Australia
Phone: 61 99054000
Email: tiffany.phillips@monash.edu

Abstract

Background: The risk factors for oropharyngeal gonorrhea have not been examined in sex workers despite the increasing prevalence of gonorrhea infection.

Objective: This study aims to determine the risk factors for oropharyngeal gonorrhea in female and gender-diverse sex workers (including cisgender and transgender women, nonbinary and gender fluid sex workers, and those with a different identity) and examine kissing, oral sex, and mouthwash practices with clients.

Methods: This mixed methods case-control study was conducted from 2018 to 2020 at 2 sexual health clinics in Melbourne, Victoria, and Sydney, New South Wales, Australia. We recruited 83 sex workers diagnosed with oropharyngeal gonorrhea (cases) and 581 sex workers without (controls). Semistructured interviews with 19 sex workers from Melbourne were conducted.

Results: In the case-control study, the median age of 664 sex workers was 30 (IQR 25-36) years. Almost 30% of sex workers (192/664, 28.9%) reported performing condomless fellatio on clients. Performing condomless fellatio with clients was the only behavior associated with oropharyngeal gonorrhea (adjusted odds ratio 3.6, 95% CI 1.7-7.6; P=.001). Most participants (521/664, 78.5%) used mouthwash frequently. In the qualitative study, almost all sex workers reported kissing clients due to demand and generally reported following clients’ lead with regard to kissing style and duration. However, they used condoms for fellatio because they considered it a risky practice for contracting sexually transmitted infections, unlike cunnilingus without a dental dam.
**Conclusions:** Our study shows that condomless fellatio is a risk factor for oropharyngeal gonorrhea among sex workers despite most sex workers using condoms with their clients for fellatio. Novel interventions, particularly targeting the oropharynx, will be required for oropharyngeal gonorrhea prevention.

(JMir Public Health Surveill 2024;10:e46845) doi:10.2196/46845

**KEYWORDS**
case-control; qualitative; oral sex; condoms; transactional sex

**Introduction**

**Background**
There has been an increase in gonorrhea incidence in several countries over the past decade [1-3], which is of particular concern given increasing antimicrobial resistance (AMR) [4,5]. The oropharynx has been implicated in Neisseria gonorrhoeae transmission, possibly through saliva exchanged during oral sex practices and tongue kissing [6-9]. The oropharynx is considered a crucial site for AMR given difficulties in treating oropharyngeal gonorrhea and the increased propensity for N gonorrhoeae to develop resistance in the oropharynx compared to other anatomical sites [4,10,11]. Sex workers have been identified as an important priority population that bears a disproportionate burden of gonorrhea worldwide, yet data on the prevalence of oropharyngeal gonorrhea in this population are sparse [12]. Therefore, there is an urgent need for research into the risk factors for oropharyngeal gonorrhea, particularly among sex workers, given its importance in AMR [13]. In 2017, an Australian study of female sex workers attending a sexual health clinic in Melbourne, Victoria, Australia, found that female sex workers had a higher prevalence of oropharyngeal gonorrhea (2%) than of genital gonorrhea (1%) [14]. That study also found that oropharyngeal infection was often independent of genital infection [14]. A retrospective study of 42 Australian sexual health clinics showed a 200% increase in oropharyngeal gonorrhea among female sex workers from 2009 to 2015 [15]. Among gay, bisexual, and other men who have sex with men, a case-control study found that oropharyngeal gonorrhea was associated with the number of kissing and receptive fellatio casual sex partners in the preceding 3 months [9]. Due to the overlapping nature of tongue kissing and oral sex practices, it was not possible to undertake an adjusted analysis to determine the risk for any one practice in isolation for the transmission of oropharyngeal gonorrhea [9]. To our knowledge, there has been no similar case-control study conducted to identify risk factors for oropharyngeal gonorrhea in sex workers.

A 2018 study of 180 female sex workers attending a sexual health clinic in Melbourne found that 149 (83.7%) tongue kissed and 175 (97.2%) performed fellatio on at least one male client in an average working week [16]. In addition to the lack of data on risk factors for oropharyngeal gonorrhea among sex workers, there is a lack of qualitative data exploring how sex workers decide whether to engage in tongue kissing and oral sex practices (ie, cunnilingus and fellatio) with male clients. There has been considerable research indicating an increased demand by male clients of female sex workers for the “girlfriend experience,” wherein sex workers and clients engage in sexual practices typical of those in intimate, noncommercial relationships (eg, kissing and cunnilingus) [17,18]. Research on the girlfriend experience tends to focus on the male clients who seek this interaction rather than on the sex workers who provide the service, with no studies exploring the decision to provide these services. A Finnish qualitative study (2008) explored sexual pleasure among female sex workers, a factor very rarely considered in the literature [19]. The study found that, while sex work sometimes required mental distancing or disengagement from the work, which decreases sexual pleasure, the work may empower women by giving them control over the sexual experience and enhancing their own pleasure. No study to our knowledge has explored which factors, including pleasure, may influence the decision of sex workers to engage in kissing and oral sex practices with clients.

Since the mid-2010s, the role of mouthwash has been and continues to be investigated as a novel intervention for gonorrhea prevention in at-risk populations such as men who have sex with men [20-23]. Using mouthwash as a means for harm reduction has been previously recommended for sex workers to reduce sexually transmitted infections (STIs) [24], and a Melbourne-based study has shown that 83% of female sex workers reported using mouthwash daily or weekly [16]. However, it is unclear why sex workers used mouthwash, be it driven by hygiene or a desire to reduce STI transmission, though recent research has shown that Listerine use does not reduce reinfection with gonorrhea [22,23]. In the event that recommendations are made incorporating mouthwash into an STI prevention strategy for sex workers, it would be beneficial to better understand what is driving mouthwash use among sex workers.

**Objectives**
The primary aim of this study was to identify the risk factors for oropharyngeal gonorrhea among sex workers (including female and nonbinary sex workers) using a case-control study design. The secondary aim was to explore why sex workers engage in tongue kissing and oral sex and use mouthwash using a qualitative approach.

**Methods**

**Study Design**
The Health Research in Sex Workers (HERS) study was a mixed methods study comprising an unmatched case-control study and semistructured interviews (Multimedia Appendix 1). Topics from the quantitative data were expanded on and explored in more detail in the semistructured interviews. The interviews were conducted simultaneously with the case-control survey and sought to clarify how and why sex workers may engage in oral sex practices involving saliva with clients.

https://publichealth.jmir.org/2024/1/e46845
Study Setting

The HERS case-control study was conducted at two sexual health clinics in Australia: (1) Melbourne Sexual Health Centre (MSHC) in Melbourne, Victoria; and (2) the Thai and Chinese clinics at Sydney Sexual Health Centre (SSHC) in Sydney, New South Wales (NSW). Participants were recruited for the case-control study at MSHC from November 2018 to March 2020 and at SSHC from November 2018 to December 2019. Both recruitment sites are large public sexual health clinics providing free sexual health services in inner urban settings. Sex work in Australia varies in legality by jurisdiction [25]. At the time of writing (January 2022), sex work was regulated by the Sex Work Act 1994 in Victoria [26], which included the criminalization of all sexual activities that include oral, vaginal, or anal penetration without a condom. In the state of NSW, sex work is decriminalized, there is no legislation criminalizing condomless sex, and risk is managed through Work Health and Safety legislation. High rates of condom use among female sex workers with male clients in both states have previously been reported, including >90% and >78% consistent use for vaginal sex and fellatio in an average working week, respectively [16-27,29].

During the study period, sex workers were required by law to receive mandatory HIV and STI testing every 3 months in Victoria [30]. There is no mandatory HIV and STI testing in NSW; rather, STI testing frequency recommendations are based on individual risk as per the Australian STI Management Guidelines [31].

Participants for the qualitative study were recruited from the MSHC from March 2019 to January 2020. Participants were unable to be recruited from the SSHC due to practical constraints as the interview team was based in Melbourne.

Female sex workers (cisgender and transgender) and sex workers who selected a different identity who were assigned male at birth, were aged ≥18 years, attended either clinic during the study period, and were working in the sex industry at the time of consultation were eligible for the HERS case-control study, and those English-speaking sex workers who attended the MSHC were also eligible for the semistructured interviews.

Case-Control Study

Participants

Sex workers attending the MSHC and SSHC for STI screening had a clinician-collected oropharyngeal swab to test for N gonorrhoeae. Cases were defined as sex workers with a positive oropharyngeal swab for gonorrhea, and controls were sex workers who had a negative oropharyngeal swab for gonorrhea.

Recruitment

Eligible and interested sex workers attending the MSHC or the Thai and Chinese clinics at the SSHC were given a paper-based questionnaire with a prelabeled study ID by the recruiting clinicians (Multimedia Appendix 2). In addition, sex workers attending either clinic for treatment of oropharyngeal gonorrhea were approached by a research nurse if they had not already completed the questionnaire and asked if they would like to participate in the study. In this way, cases were purposively recruited.

Data Collection

The questionnaire was designed by the study investigators with feedback sought from organizations that provide services for sex workers (Resourcing Health & Education [RhED] and the Sex Workers Outreach Project [SWOP], both services for sex workers in Victoria and NSW, respectively). The questionnaire was offered in English, Thai, and Chinese at both recruitment sites (translated by a professional translation company but checked for readability by bilingual clinicians). Participants’ unique patient identifier was recorded on a separate log with the corresponding questionnaire study ID. This log was collected daily by the research staff. Consent was implied by questionnaires being returned completed. No payment was given to participants for returning the questionnaire.

The HERS case-control quantitative questionnaire collected data on demographic characteristics (eg, age, sex, country of birth, length of time in Australia, and languages spoken at home), mouthwash practices, location of sex work, sex practices performed with male clients in an average working week, and sex practices with noncommercial male sexual partners in the previous 7 days. Sex practices with female partners were not ascertained given that most clients are male and the risk of gonorrhea among women who have sex with women is lower than among those who have sex with men [32]. Sex practices included tongue kissing, oral sex (fellatio with and without ejaculation and cunnilingus), vaginal sex, anal sex, sex involving toys, and using saliva (either theirs or a partner’s) as lubricant during sex (vaginal sex, anal sex, and sex involving toys).

Gonorrhea was diagnosed using nucleic acid amplification tests performed on the Aptima Combo 2 assay (Hologic) at the MSHC and the Roche cobas CT and NG assay (Roche Diagnostics) at the SSHC. The Aptima and Roche assays have similar sensitivity and specificity for oropharyngeal gonorrhea detection [33]. Returned questionnaires were linked (via the questionnaire study ID) to the patient identifier number recorded in the log by research staff to match their oropharyngeal gonorrhea test results with their questionnaire data.

Study Size

Sample size was calculated using OpenEpi [34] using the estimate that 25% of female sex workers have condomless fellatio [35], an assumed minimum risk factor prevalence of 25% in controls, and a minimum odds ratio (OR) of 2. It should be noted that our original aim was to conduct a 1:4 age-matched case-control study, which would have required 92 cases to 366 controls assuming the same parameters. However, due to the COVID-19 pandemic, recruitment had to be terminated before 92 cases were recruited given that all sex work was required by law to cease during the COVID-19 lockdown nationally (beginning March 25, 2020, in Australia) [36] and because sex practices with clients (and, thus, risk factors for oropharyngeal gonorrhea) might have changed during the pandemic after sex work resumed. For an unmatched case-control study with a ratio of 1:7 cases to controls, at 80% power and a .05 significance, there needed to be 83 cases to 579 controls. There were 83 cases...
and 581 controls recruited before the COVID-19 lockdown; thus, we were able to conduct an unmatched analysis with 80% power to detect a difference using a significance level of 5%.

This study was reported as per the Strengthening the Reporting of Observational Studies in Epidemiology 2008 guidelines for case-control studies.

**Statistical Methods**

Descriptive statistics were used to calculate the median age of participants. The median number of years in Australia for those born overseas was calculated, and participants were categorized as newly arrived or not based on whether they were in Australia for less than or equal to the median number of years or more than the median number of years. The median was calculated for the number of male clients seen in an average working week. The Mann-Whitney U test was used to compare the differences between continuous variables between cases and controls. A chi-square test was used to compare differences in the proportion of categorical variables, and those who declined to report were excluded from the chi-square analysis.

Univariable and multivariable unconditional logistic regression was conducted to identify (1) the factors associated with declining to report sexual practices and (2) the factors associated with oropharyngeal gonorrhea. There were some differences in patient characteristics between the MSHC and SSHC, so we adjusted for site of recruitment in the logistic regression. ORs and the corresponding 95% CIs were reported. Variables with \( P < .10 \) in the univariable analyses were included in the multivariable analyses. Given the different assays for gonorrhea detection used between the MSHC and SSHC, the site of recruitment was adjusted in the multivariable logistic regression for associations with oropharyngeal gonorrhea. All statistical analyses were conducted using Stata (version 14; StataCorp).

**Qualitative Study**

**Overview**

A qualitative descriptive approach was used for the qualitative study component. Qualitative description is a pragmatic rather than theory-driven approach that aims to provide a description of participants’ views and experiences rather than an interpretive, theory-driven analysis [37]. This approach is commonly used in health care when there are specific questions of clinical interest that the study seeks to answer [37].

**Participants**

Sex workers attending the MSHC who completed a HERS English-language version of the questionnaire for the case-control study were eligible for the semistructured interviews.

**Recruitment**

During the recruitment period, participants at the MSHC were shown an invitation at the end of the English-language version of the quantitative HERS questionnaire to participate in a one-on-one semistructured interview on the topics covered in the questionnaire. All participants who responded “yes” to taking part in the interviews were contacted via SMS text message (from the phone numbers listed on their medical records at the MSHC), and interviews were arranged either face-to-face or over the phone, according to participant preference, with those still wanting to participate. Written or verbal (if the interview was conducted over the phone) informed consent was obtained from participants after the study information was discussed with them.

**Data Collection**

The interview schedule was designed by the research team at the MSHC and then reviewed by the ACON Research Ethics Review Committee; the SWOP in NSW; and RhED, a service for sex workers in Victoria.

The interview schedule contained questions regarding sexual practices they performed at work that involved saliva exchange (either theirs or the clients’), as well as their use of mouthwash at work. The interview schedule was based on the case-control survey questions, allowing for further exploration and depth of understanding regarding oral sex practices that involve saliva (specifically, adding depth to the data in terms of the “what,” “how,” and “why” of these practices). Participants were given an Aus $50 (US $32.08) gift voucher after the completion of their interview (either in person or by mail in an unmarked envelope) to reimburse them for their time. All the interviews were audio recorded with the participants’ permission and transcribed verbatim. Transcripts were deidentified. Audio files and transcripts were stored on secure servers and will be destroyed after 7 years in line with Alfred Health ethics requirements.

Interviews were conducted in English by 1 of 2 researchers (TRP and KM), both of whom have experience with conducting interviews on sensitive sexual health issues. TRP sat in on the first 2 interviews conducted by KM, and the 2 researchers reviewed and discussed the interviews afterward to ensure consistency in the approach to them. TRP and KM met weekly to discuss the interviews, the interview schedule, and the developing themes. Upon completion of 12 interviews, the wider qualitative research team members (TRP, KM, and JB) met to discuss the data, the developing themes, and any further lines of questioning required. At this point, one question was added to the interview schedule to further explore a topic that some participants had touched on in the interviews. The new question concerned hypothetical willingness to change sexual practices at work if saliva was shown to transmit gonorrhea.

A total of 19 interviews had been completed when participant recruitment was interrupted by the COVID-19 pandemic. The research team met to discuss preliminary findings; it was decided that recruitment would not continue in late 2020 and 2021 due to the likely differing sex practices and risk perceptions with regard to COVID-19 as well as the ongoing COVID-19 lockdowns and restrictions causing continued disruptions in sex work and research in Australia.

According to Braun and Clarke [38], the concept of data saturation and its meaning and use depend on the purpose and goals of a study and the data analysis approach being used. While Braun and Clarke [38] argue that data saturation is not the best rationale for sample size, particularly when using a reflective thematic data analysis approach, when studies are...
conducted for largely pragmatic reasons and follow a fairly structured deductive approach to data analysis in which the data generated are relatively concrete, participants are largely recruited from a particular setting, coding largely relates to the broad topics or interview questions, and themes act largely as a summary of participants’ differing responses with little further interpretation regarding meaning, the concept of data saturation, or the point at which no “new” insights are provided, may be useful [38]. Given the deductive nature of the analysis and that the themes were largely guided by the research questions, the concept of data saturation was applied, with themes saturated among a broad sample with a range of ages, years working in the sex industry, and experiences in the context of brothel-based sex work in Melbourne. However, further insights into 2 particular groups of sex workers would have been beneficial, including those who worked outside of brothels or who were cases (diagnosed with oropharyngeal gonorrhea at the time of recruitment). For pragmatic reasons previously outlined (recruitment had to cease due to the COVID-19 pandemic), this was not possible.

**Data Analysis**

Data were analyzed using deductive thematic analysis, a “top down” approach wherein coding and analysis are largely informed by the ideas and concepts the researcher brings to the data rather than being created from the data themselves. In the case of this study, coding and themes were largely guided by the interview schedule questions and topics, which had been informed by literature in the field and the specific questions of clinical interest that the research team sought to answer [39]. Each transcript was initially read and coded by TRP. After all transcripts were coded, they were imported into NVivo (version 12; QSR International) for data management. The codes were grouped and labeled into preliminary themes before each transcript was read again, and the themes were further revised, refined, and compared for similarities and differences. A subset of transcripts was read and coded by JB and KM, after which all 3 researchers (TRP, JB, and KM) met to discuss and reach a final consensus on the themes. There were no major differences in interpretation evident.

This study is reported as per the RATS guidelines for qualitative research [40].

**Ethical Considerations**

This study was approved by the Alfred Hospital Ethics Committee, Melbourne, Victoria, Australia (project 596/17), and the South Eastern Sydney Local Health District (reference: 18/G/166). This study was also approved by the ACON Research Ethics Review Committee (reference number: 2018/17) with support from the SWOP NSW. This project was also reviewed by RhED, a service for sex workers in Victoria. As stated above, informed consent was given for the interviews and implied consent was considered by a returned completed questionnaire for the case-control study. Data presented is anonymous. Participants in the qualitative interviews were compensated Aus $50 (US $32.08). There was no compensation for participants in the case-control study.

**Results**

**Case-Control Study**

There were 386 surveys included from the MSHC, of which 23 (6%) were from cases and 363 (94%) were from controls. There were 278 surveys included from the SSHC, of which 60 (21.6%) were from cases and 218 (78.4%) were from controls (Multimedia Appendix 2). Thus, of the 664 surveys, the total number of cases was 83 (12.5%), and the total number of controls was 581 (87.5%). Of the 664 surveys included in the analysis, there were 542 (81.6%) that were returned with every section completed, 12 (1.8%) that were returned with every section except the demographics and mouthwash sections completed, and 110 (16.6%) that were returned with every section except the sex practice with clients and noncommercial partners sections completed. Being born in China or other countries aside from Thailand was associated with returning the survey with incomplete sections for sexual practices (Multimedia Appendix 3).

The median age of participants was 30 (IQR 25-36) years, and participants at the SSHC were older than those from the MSHC (median age 33 vs 28 y; \( P < .001 \); Table 1). Cases were older than controls (median age 32 vs 30 y; \( P = .01 \)).
Table 1. Risk factors for oropharyngeal gonorrhea among female sex workers recruited from 2 sexual health clinics in Australia (N=664).

<table>
<thead>
<tr>
<th>Site</th>
<th>Cases (n=83), n (%)</th>
<th>Controls (n=581), n (%)</th>
<th>OR(^a) (95% CI)</th>
<th>aOR(^b) (95% CI)</th>
<th>P value for aOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSHC(^c)</td>
<td>23 (27.7)</td>
<td>363 (62.5)</td>
<td>Reference</td>
<td>Reference</td>
<td>— (d)</td>
</tr>
<tr>
<td>SSHC(^e)</td>
<td>60 (72.3)</td>
<td>218 (37.5)</td>
<td>4.3 (2.6-7.2)(^f)</td>
<td>3.1 (1.5-6.2)</td>
<td>.001</td>
</tr>
<tr>
<td>Gender identity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>79 (95.2)</td>
<td>531 (91.4)</td>
<td>Reference</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Transgender and gender diverse</td>
<td>4 (4.8)</td>
<td>50 (8.6)</td>
<td>0.5 (0.2-1.5)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤24</td>
<td>8 (9.6)</td>
<td>121 (20.8)</td>
<td>Reference</td>
<td>Reference</td>
<td>—</td>
</tr>
<tr>
<td>25-34</td>
<td>43 (51.8)</td>
<td>289 (49.7)</td>
<td>2.3 (1.0-4.9)(^f)</td>
<td>1.2 (0.5-2.9)</td>
<td>.70</td>
</tr>
<tr>
<td>≥35</td>
<td>32 (38.6)</td>
<td>171 (29.4)</td>
<td>2.8 (1.3-6.4)(^f)</td>
<td>1.1 (0.4-2.8)</td>
<td>.85</td>
</tr>
<tr>
<td>Newly arrived in Australia (within 3 years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>34 (41)</td>
<td>370 (63.7)</td>
<td>Reference</td>
<td>Reference</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>49 (59)</td>
<td>211 (36.3)</td>
<td>2.5 (1.6-4.0)(^f)</td>
<td>1.2 (0.6-2.1)</td>
<td>.63</td>
</tr>
<tr>
<td>Smoke daily</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>60 (72.3)</td>
<td>427 (73.5)</td>
<td>Reference</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>20 (24.1)</td>
<td>141 (24.3)</td>
<td>1.0 (0.6-1.7)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Declined to report</td>
<td>3 (3.6)</td>
<td>13 (2.2)</td>
<td>1.6 (0.5-5.9)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Use mouthwash(^g)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrequently</td>
<td>10 (12)</td>
<td>122 (21)</td>
<td>Reference</td>
<td>Reference</td>
<td>—</td>
</tr>
<tr>
<td>Frequently</td>
<td>69 (83.1)</td>
<td>452 (77.8)</td>
<td>1.9 (0.9-3.7)(^f)</td>
<td>1.0 (0.5-2.2)</td>
<td>.97</td>
</tr>
<tr>
<td>Did not disclose</td>
<td>4 (4.8)</td>
<td>7 (1.2)</td>
<td>7.0 (1.7-27.9)(^f)</td>
<td>11.2 (2.2-57.2)</td>
<td>.004</td>
</tr>
<tr>
<td>Type (venue) of sex work</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brothel only</td>
<td>47 (56.6)</td>
<td>289 (49.7)</td>
<td>Reference</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Massage parlor only</td>
<td>12 (14.5)</td>
<td>139 (23.9)</td>
<td>0.5 (0.3-1.0)(^f)</td>
<td>0.2 (0.1-0.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Multiple or other venues</td>
<td>24 (28.9)</td>
<td>150 (25.8)</td>
<td>1.0 (0.6-1.7)</td>
<td>1.0 (0.6-1.9)</td>
<td>.90</td>
</tr>
<tr>
<td>Declined to report</td>
<td>0 (0)</td>
<td>3 (0.5)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Sexual practices with clients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tongue kiss male clients(^h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>9 (10.8)</td>
<td>146 (25.1)</td>
<td>Reference</td>
<td>Reference</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>51 (61.4)</td>
<td>348 (59.9)</td>
<td>2.4 (1.1-5.0)(^f)</td>
<td>1.5 (0.7-3.5)</td>
<td>.30</td>
</tr>
<tr>
<td>Declined to report</td>
<td>23 (27.7)</td>
<td>87 (15)</td>
<td>4.3 (1.9-9.7)(^a)</td>
<td>3.3 (1.2-9.4)</td>
<td>.02</td>
</tr>
<tr>
<td>Perform condomless fellatio on male clients(^h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>17 (20.5)</td>
<td>346 (59.6)</td>
<td>Reference</td>
<td>Reference</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>48 (57.8)</td>
<td>144 (24.8)</td>
<td>6.8 (3.8-12.2)(^f)</td>
<td>3.6 (1.7-7.7)</td>
<td>.001</td>
</tr>
<tr>
<td>Declined to report</td>
<td>18 (21.7)</td>
<td>91 (15.7)</td>
<td>4.0 (2.0-8.1)(^f)</td>
<td>1.1 (0.8-6.0)</td>
<td>.14</td>
</tr>
<tr>
<td>Client ejaculated in mouth(^h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>43 (51.8)</td>
<td>440 (75.7)</td>
<td>Reference</td>
<td>Reference</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>25 (30.1)</td>
<td>56 (9.6)</td>
<td>4.6 (2.6-8.0)(^f)</td>
<td>1.4 (0.7-2.9)</td>
<td>.32</td>
</tr>
</tbody>
</table>
There were 8.1% (54/664) transgender and gender-diverse participants. Most participants were born in Australia or New Zealand (230/664, 34.6%), followed by Thailand (166/664, 25%), China (156/664, 23.5%), and other countries (112/664, 16.9%). Overall, there were 39.2% (260/664) of participants who were newly arrived in Australia. For those cases who were born outside Australia, the median length of time in Australia was 2 (IQR 1-3) years compared to 3 (IQR 2-6) years for controls \((P=.002)\). There was no difference in the proportion of transgender and gender-diverse participants between the 2 sites (28/386, 7.3% from the MSHC vs 26/278, 9.4% from the SSHC; \(P=.33\)). Most participants from the SSHC were born in Thailand (147/278, 52.9%), China (120/278, 43.2%), or Taiwan (5/278, 1.8%); however, there were 2.2% (6/278) who were born in other countries. Most participants from the MSHC were born in Australia or New Zealand (228/386, 59.1%), followed by China (36/386, 9.3%), Thailand (21/386, 5.4%), India (7/386, 1.8%), and Malaysia (7/386, 1.8%). The remaining 22.5% (87/386) of participants from the MSHC were from 39 different countries. For those cases born outside Australia, the median length of time in Australia was 2 (IQR 1-3) years compared to 3 (IQR 2-6) years for controls \((P=.002)\). Most participants worked in only one type or venue of sex work; most worked at brothels only (336/664, 50.6%), followed by massage parlors (151/664, 22.7%) and private (72/664, 10.8%) and street-based sex work (1/664, 0.2%). There were 15.1% (100/664) of participants who worked in more than one type or venue of sex work, of which the most common combination was brothel and private sex work (56/100, 56%).

<table>
<thead>
<tr>
<th>Sexual practices with not-at-work sexual partners</th>
<th>Cases (n=83), n (%)</th>
<th>Controls (n=581), n (%)</th>
<th>OR(^a) (95% CI)</th>
<th>aOR(^b) (95% CI)</th>
<th>(P) value for aOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declined to report</td>
<td>15 (18.1)</td>
<td>85 (14.6)</td>
<td>1.8 (1.0-3.4)(^f)</td>
<td>0.6 (0.3-1.5)</td>
<td>.30</td>
</tr>
<tr>
<td>Tongue kiss noncommercial sexual partner(^i)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No(^j)</td>
<td>48 (57.8)</td>
<td>341 (58.7)</td>
<td>Reference</td>
<td>Reference</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>20 (24.1)</td>
<td>182 (31.3)</td>
<td>0.8 (0.4-1.4)</td>
<td>1.4 (0.7-2.7)</td>
<td>.30</td>
</tr>
<tr>
<td>Declined to report</td>
<td>15 (18.1)</td>
<td>58 (10)</td>
<td>1.8 (1.0-3.5)(^f)</td>
<td>1.6 (0.5-5.4)</td>
<td>.47</td>
</tr>
<tr>
<td>Performed condomless fellatio on not-at-work sexual partner(^i)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No(^j)</td>
<td>57 (68.7)</td>
<td>397 (68.3)</td>
<td>Reference</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>8 (9.6)</td>
<td>104 (17.9)</td>
<td>0.5 (0.2-1.2)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Declined to report</td>
<td>18 (21.7)</td>
<td>80 (13.8)</td>
<td>1.6 (0.9-2.8)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Not-at-work sexual partner ejaculated in mouth(^i)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No(^j)</td>
<td>68 (81.9)</td>
<td>453 (78)</td>
<td>Reference</td>
<td>Reference</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>1 (1.2)</td>
<td>70 (12)</td>
<td>0.1 (0.0-0.7)(^f)</td>
<td>0.1 (0.0-0.6)</td>
<td>.02</td>
</tr>
<tr>
<td>Declined to report</td>
<td>14 (16.9)</td>
<td>58 (10)</td>
<td>1.6 (0.9-3.0)</td>
<td>0.9 (0.3-3.1)</td>
<td>.88</td>
</tr>
</tbody>
</table>

\(^a\) OR: odds ratio.  
\(^b\) aOR: adjusted odds ratio.  
\(^c\) MSHC: Melbourne Sexual Health Centre.  
\(^d\) Not applicable.  
\(^e\) SSHC: Sydney Sexual Health Centre.  
\(^f\) Variables with \(P\) values of <.10 in the univariable analyses were included in the multivariable analyses.  
\(^i\) Participants were asked to report any sex practices with male clients in an average working week.  
\(^j\) Includes 328 controls and 47 cases who did not have a noncommercial sex partner.  

Most participants worked in only one type or venue of sex work; most worked at brothels only (336/664, 50.6%), followed by massage parlors (151/664, 22.7%) and private (72/664, 10.8%) and street-based sex work (1/664, 0.2%). There were 15.1% (100/664) of participants who worked in more than one type or venue of sex work, of which the most common combination was brothel and private sex work (56/100, 56%).

Of the 578 participants who answered the question on number of male clients in an average working week, the median was 10 (IQR 6-20) clients; there was no significant difference in the number of clients between cases and controls \((P=.18; \text{Table 2})\). There were no significant differences in the proportion who performed fellatio on clients, received cunnilingus from clients, had vaginal or anal sex with clients, or had sex involving sex toys between cases and controls (Table 2). However, a significantly higher proportion of cases tongue kissed their clients (51/60, 85% vs 348/494, 70.4%; \(P=.02\)), performed condomless fellatio on clients (48/65, 74% vs 144/490, 29.4%; \(P<.001\)), and had condomless vaginal sex (17/60, 28% vs 78/470, 16.6%; \(P=.03\)) and condomless anal sex (10/18, 56% vs 56/183, 30.6%; \(P=.03\)) with clients in an average working week.
Table 2. Number of male clients seen in an average working week among female sex workers by cases with oropharyngeal gonorrhea and controls (N=664).

<table>
<thead>
<tr>
<th></th>
<th>Cases (n=83)</th>
<th>Controls (n=581)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of male clients seen in an average working week, median (IQR)b</strong></td>
<td>10 (5-15)</td>
<td>10 (6-20)</td>
<td>.18</td>
</tr>
<tr>
<td><strong>Tongue kissed male clients in an average working week, n (%)</strong></td>
<td></td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>No</td>
<td>9 (10.8)</td>
<td>146 (25.1)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>51 (61.4)</td>
<td>348 (59.9)</td>
<td></td>
</tr>
<tr>
<td>Declined to report</td>
<td>23 (27.7)</td>
<td>87 (15)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of clients tongue kissed, median (IQR)</strong></td>
<td>2 (0-9)</td>
<td>2 (0-6)</td>
<td>.17</td>
</tr>
<tr>
<td><strong>Proportion of clients tongue kissed (%), median (IQR)c</strong></td>
<td>40 (10-90)</td>
<td>50 (10-80)</td>
<td>.43</td>
</tr>
<tr>
<td><strong>Perform fellatio on male clients in an average working week, n (%)</strong></td>
<td></td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td>No</td>
<td>2 (2.4)</td>
<td>54 (9.3)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>63 (75.9)</td>
<td>447 (76.9)</td>
<td></td>
</tr>
<tr>
<td>Declined to report</td>
<td>18 (21.7)</td>
<td>80 (13.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of clients performed fellatio on, median (IQR)</strong></td>
<td>5 (2-9)</td>
<td>7 (2-14)</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Proportion of clients performed fellatio on (IQR)c</strong></td>
<td>90 (50-100)</td>
<td>70 (30-99)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Perform condomless fellatio on male clients in an average working week, n (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>17 (20.5)</td>
<td>346 (59.6)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>48 (57.8)</td>
<td>144 (24.8)</td>
<td></td>
</tr>
<tr>
<td>Declined to report</td>
<td>18 (21.7)</td>
<td>91 (15.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of clients performed condomless fellatio on, median (IQR)</strong></td>
<td>1 (0-3)</td>
<td>0 (0-0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Proportion of clients performed condomless fellatio on (IQR)c</strong></td>
<td>50 (40-80)</td>
<td>50 (20-90)</td>
<td>.29</td>
</tr>
<tr>
<td><strong>Receive cunnilingus from male clients in an average working week, n (%)</strong></td>
<td></td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td>No</td>
<td>5 (6)</td>
<td>75 (12.9)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>57 (68.7)</td>
<td>422 (72.6)</td>
<td></td>
</tr>
<tr>
<td>Declined to report</td>
<td>21 (25.3)</td>
<td>84 (14.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of clients received cunnilingus from, median (IQR)</strong></td>
<td>3 (1-7)</td>
<td>4 (1-8)</td>
<td>.62</td>
</tr>
<tr>
<td><strong>Proportion of clients received cunnilingus from (%), median (IQR)c</strong></td>
<td>40 (10-50)</td>
<td>50 (20-70)</td>
<td>.08</td>
</tr>
<tr>
<td><strong>Vaginal sex with clients in an average working week, n (%)</strong></td>
<td></td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>No</td>
<td>9 (10.8)</td>
<td>54 (9.3)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>45 (54.2)</td>
<td>428 (73.7)</td>
<td></td>
</tr>
<tr>
<td>Declined to report</td>
<td>29 (34.9)</td>
<td>99 (17)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of clients had vaginal sex with, median (IQR)</strong></td>
<td>5 (2-15)</td>
<td>8 (3-15)</td>
<td>.11</td>
</tr>
<tr>
<td><strong>Proportion of clients had vaginal sex with (IQR)c</strong></td>
<td>90 (60-100)</td>
<td>95 (80-100)</td>
<td>.69</td>
</tr>
<tr>
<td><strong>Condomless vaginal sex with clients in an average working week, n (%)</strong></td>
<td></td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>No</td>
<td>43 (51.8)</td>
<td>392 (67.5)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>17 (20.5)</td>
<td>78 (13.4)</td>
<td></td>
</tr>
<tr>
<td>Declined to report</td>
<td>23 (27.7)</td>
<td>111 (19.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of clients had condomless vaginal sex with, median (IQR)</strong></td>
<td>0 (0-0)</td>
<td>0 (0-0)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Proportion of clients had condomless vaginal sex with (IQR)c</strong></td>
<td>20 (20-70)</td>
<td>50 (10-90)</td>
<td>.37</td>
</tr>
<tr>
<td><strong>Anal sex with clients in an average working week, n (%)</strong></td>
<td></td>
<td></td>
<td>.40</td>
</tr>
<tr>
<td>No</td>
<td>45 (54.2)</td>
<td>395 (68)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>6 (7.2)</td>
<td>77 (13.3)</td>
<td></td>
</tr>
</tbody>
</table>
### Number of clients had anal sex with, median (IQR)

<table>
<thead>
<tr>
<th></th>
<th>Cases (n=83)</th>
<th>Controls (n=581)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declined to report</td>
<td>32 (38.6)</td>
<td>109 (18.8)</td>
<td>.92</td>
</tr>
<tr>
<td>Number of clients had anal sex with (%)</td>
<td>0 (0-0)</td>
<td>0 (0-0)</td>
<td>.21</td>
</tr>
<tr>
<td>Proportion of clients had anal sex with (%)</td>
<td>35 (5-80)</td>
<td>10 (5-30)</td>
<td>.03</td>
</tr>
</tbody>
</table>

### Condomless anal sex with clients in an average working week, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Cases (n=83)</th>
<th>Controls (n=581)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>127 (21.9)</td>
<td>56 (9.6)</td>
<td>.37</td>
</tr>
<tr>
<td>Yes</td>
<td>20 (14.5)</td>
<td>113 (19.4)</td>
<td>.19</td>
</tr>
<tr>
<td>Declined to report</td>
<td>32 (55.2)</td>
<td>398 (68.5)</td>
<td></td>
</tr>
</tbody>
</table>

### Use sex toys with clients in an average working week, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Cases (n=83)</th>
<th>Controls (n=581)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>127 (21.9)</td>
<td>321 (55.2)</td>
<td>.57</td>
</tr>
<tr>
<td>Yes</td>
<td>56 (9.6)</td>
<td>147 (25.3)</td>
<td>.14</td>
</tr>
<tr>
<td>Declined to report</td>
<td>398 (68.5)</td>
<td>113 (19.4)</td>
<td></td>
</tr>
</tbody>
</table>

### Number of clients had sex with involving toys, median (IQR)

<table>
<thead>
<tr>
<th></th>
<th>Cases (n=83)</th>
<th>Controls (n=581)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>0 (0-0)</td>
<td>0 (0-0)</td>
<td>.37</td>
</tr>
<tr>
<td>Yes</td>
<td>20 (10-50)</td>
<td>10 (5-20)</td>
<td>.19</td>
</tr>
<tr>
<td>Declined to report</td>
<td>0 (0-0)</td>
<td>0 (0-0)</td>
<td></td>
</tr>
</tbody>
</table>

### Proportion of clients had condomless anal sex with (IQR)

<table>
<thead>
<tr>
<th></th>
<th>Cases (n=83)</th>
<th>Controls (n=581)</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>127 (21.9)</td>
<td>321 (55.2)</td>
<td>.57</td>
</tr>
<tr>
<td>Yes</td>
<td>56 (9.6)</td>
<td>147 (25.3)</td>
<td>.14</td>
</tr>
<tr>
<td>Declined to report</td>
<td>398 (68.5)</td>
<td>113 (19.4)</td>
<td></td>
</tr>
</tbody>
</table>

# Qualitative Findings

## Overview

In total, 9.5% (63/664) of sex workers indicated interest in participating in the interviews and were contacted by a research nurse at the MSHC. When contacted, of the 63 sex workers, 28 (44%) at the MSHC agreed to participate and scheduled a time for the interview; however, 9 (32%) did not attend their scheduled interview and could not be reached to reschedule. In total, 19 participants completed interviews before data collection was discontinued. The age of the 19 participants ranged from 18 to 44 years, with a median of 28 (IQR 24-31) years (Table 3). The duration of interviews ranged from 24 to 61 minutes, with a median of 41 minutes. One participant in the qualitative interviews was a case in the case-control group, whereas the rest were from the control group.
Table 3. Participant demographics for the qualitative interviews among female sex workers recruited from the Melbourne Sexual Health Centre (N=19).

<table>
<thead>
<tr>
<th></th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender identity</strong></td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>18 (95)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Oropharyngeal gonorrhea</strong></td>
<td></td>
</tr>
<tr>
<td>Case</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Control</td>
<td>18 (95)</td>
</tr>
<tr>
<td><strong>Age range (y)</strong></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>7 (37)</td>
</tr>
<tr>
<td>25-30</td>
<td>6 (32)</td>
</tr>
<tr>
<td>31-35</td>
<td>6 (32)</td>
</tr>
<tr>
<td><strong>Length of time in the sex industry (y)</strong></td>
<td></td>
</tr>
<tr>
<td>≤1</td>
<td>3 (16)</td>
</tr>
<tr>
<td>2</td>
<td>3 (16)</td>
</tr>
<tr>
<td>≥3</td>
<td>13 (68)</td>
</tr>
<tr>
<td><strong>Country of birth</strong></td>
<td></td>
</tr>
<tr>
<td>Australia or New Zealand</td>
<td>17 (89)</td>
</tr>
<tr>
<td>Overseas</td>
<td>2 (11)</td>
</tr>
<tr>
<td><strong>Location of sex work</strong></td>
<td></td>
</tr>
<tr>
<td>Brothel</td>
<td>16 (84)</td>
</tr>
<tr>
<td>Massage parlor</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Private</td>
<td>1 (5)</td>
</tr>
<tr>
<td>Brothel and private</td>
<td>1 (5)</td>
</tr>
<tr>
<td><strong>Mouthwash use at work</strong></td>
<td></td>
</tr>
<tr>
<td>Before and after clients</td>
<td>8 (42)</td>
</tr>
<tr>
<td>Before clients</td>
<td>4 (21)</td>
</tr>
<tr>
<td>After clients</td>
<td>3 (16)</td>
</tr>
<tr>
<td>Rarely or never</td>
<td>4 (21)</td>
</tr>
</tbody>
</table>

The qualitative data were organized into six descriptive themes related to the kissing and oral sex practices and mouthwash use among sex workers (Textbox 1): (1) the “how” and “why” of kissing clients, (2) always covered—fellatio with clients, (3) not so risky—uncovered cunnilingus from clients, (4) clients’ saliva as lubricant is a “no go,” (5) pleasure with clients—a “perk” or not part of the job, and (6) mouthwash use with clients—a freshener and germ killer.
Textbox 1. An overview of female sex worker practices—the what, the why, and the how from semistructured interviews with 19 female sex workers recruited from the Melbourne Sexual Health Centre.

**Kissing (theme 1: the “how” and “why” of kissing clients)**

- Very common in some form
- Sex worker initiated
  - Typically closed mouth or shallow
  - To improve relationship with regular clients
  - To facilitate a more intimate experience
  - Pressured by client demand and availability
- Client initiated
  - Often negotiated before encounter
  - Commanded higher premium

**Expectations for clients**

- Good oral hygiene
- Sobriety

**Performing oral sex (fellatio; theme 2: always covered—fellatio with clients)**

- Very common
- Condoms used always
  - Perceived as “high risk”
  - Reduce sexually transmitted infection (STI) risks
  - Legal requirements
  - Public health campaigns
- Client pressure
  - Financial incentive
  - Waiting until the encounter had begun

**Receiving cunnilingus (theme 3: not so risky—uncovered cunnilingus from clients)**

- Very common
- Dental dams never used
  - Perceived as “low” or “no risk”
  - Detracts from client experience
  - Difficult to use
  - Not common or industry standard
  - No public health campaigns

**Saliva as lubricant (theme 4: clients’ saliva as lubricant a “no go”)**

- Uncomfortable with client using saliva as lubricant
  - Perceived as “high risk”
  - Disgusted by the idea
  - Not as effective as actual lubricant
- May sometimes use their own saliva as lubricant
  - Not preferred over actual lubricant
The How and Why of Kissing Clients

All participants in the qualitative interviews (19/19, 100%) reported kissing at least some of their clients; however, the style of kissing varied from closed-mouth kissing (“pecks”) and shallow tongue kissing to deep tongue kissing. For most participants, whether they kissed clients depended on whether the client requested the service before the booking as it entails an additional cost, although there were a few participants who reported reserving kissing for regular clients, those they judged to have good oral hygiene (ie, clean teeth), or those not visibly intoxicated.

The style and duration of kissing was most often dictated by the client’s preference and initiation, with almost half of participants describing following the clients’ lead:

- Will use out of convenience

Pleasure with clients (theme 5: pleasure with clients—a “perk” or not part of the job)

- Does not influence which sex practices are engaged in
- Some feel that it is not possible to feel pleasure with clients
  - Compartmentalizing as “work” and not for enjoyment
  - Lack of emotional connection
- Pleasure infrequent but can impact kissing and cunnilingus
  - May kiss for longer or deeper
  - May allow client to kiss or perform cunnilingus despite not paying for the service
- More enjoyment over time
  - More confident and relaxed with clients

Using mouthwash (theme 6: mouthwash use with clients—a freshener and germ killer)

- Routinely used by some female sex workers
  - Freshen up before the booking
  - Kill any germs they may have picked up from the client after the booking
  - As a matter of routine to “clean the slate” before the next client
- Occasionally used by some female sex workers
  - When they feel their breath needs to be freshened up before a booking
  - After a booking if left with unclean feeling
- Some never use mouthwash at work
  - Believe it is bad for oral health
- Most prefer that clients use mouthwash before the booking
  - Freshen breath
  - Perception that it makes client less likely to transmit germs
  - Brothel provides mouthwash
  - Expectation that client uses mouthwash and showers before the booking

So, it’s because they’re paying for it, it’s how they’re wanting that to happen. I’ll generally—if I’m doing the instigation of the kissing, it will usually be more pecking than full tongue. But there will still be saliva on the lips and that sort of thing. Then it’s up to them how they want to take it further. I continue on the line that they’re taking things. [Participant 5; aged 30 years; less than a year in the sex industry; brothel based]

Other participants did not allow the client to dictate kissing style; instead, they reported efforts to avoid deep kissing with clients, such as kissing with closed teeth or telling the client to use less tongue. Some described general discomfort with deep kissing clients, including one who felt that it was too intimate for work:

I just find the sharing of that much saliva is a bit...too...it’s a bit too intimate for just my work...
While participants commonly stated that pay was the only factor in deciding to kiss a client or not, there were several who felt pressured to agree to kiss clients to compete with other sex workers for clients. However, a couple of participants felt that kissing clients enhanced the booking in terms of making the service feel more genuine as well as allowing them to engage with the client in a more natural way.

**Always Covered: Fellatio With Clients**

All participants (19/19, 100%) described performing oral sex (fellatio) on some clients, and all (19/19, 100%) reported always using condoms for this practice. Almost all participants reported that the main reason for using condoms during fellatio was for safety as they did not want to contract an STI in the throat, whereas almost half of participants reported using condoms due to the legal requirement. One participant shared the following:

> Because it’s the law and I want to be really safe. That’s not—if I—I’ve had many, many, many people try to get it without a condom, but I don’t know these people. I don’t know where they’ve been. I need that protection. [Participant 9; aged 18 years; less than a month in the sex industry; brothel based]

Many of these participants described seeing posters in their workplaces that caution against performing fellatio on clients without a condom. Some participants reported receiving pressure from clients to perform oral sex without a condom, most often after the booking had commenced and they were alone in the room. Several participants reported men offering to pay more for a “natural” (condomless) fellatio service, and while reportedly declining these requests, several also said that they knew of other sex workers in their workplaces who accept these offers for more pay.

**Not So Risky: Uncovered Cunnilingus From Clients**

All participants except one reported having some clients perform oral sex (cunnilingus) on them, and none reported using a dental dam for this practice. Reasons for not using a dental dam varied, but the common sentiment among participants was that cunnilingus did not seem to place them at as much risk of STIs compared to them performing condomless fellatio on clients:

> I’ve heard it’s [cunnilingus] more likely to affect them more than me. That’s what I’ve been told anyway. So it’s like their risk, not mine and I know I’m clean, because I get tested all the time. [Participant 6; aged 26 years; 6 years in the sex industry; massage parlor]

Other reasons for not using a dental dam revolved around the dental dam detracting from the experience for the client. Many participants described dental dams as being difficult to use and anticipated that enforcing clients to use one would deter them from booking services with them and, ultimately, lead to less bookings. Several participants commented that it was not industry standard to use dental dams and there were no posters warning sex workers to use them for oral sex like there were for condoms.

**Client’s Saliva as a Lubricant a “No Go”**

With few exceptions, almost all participants were uncomfortable having a client use their own saliva as lubricant for sex, though several described previously having clients spit on their hands and touch the participants’ genitals without asking first and before they could be stopped. Reasons for not being comfortable included a fear of contracting STIs from the client’s saliva, preferring actual lubricant over saliva in general as it is more effective, and being generally disgusted by the client’s saliva. One participant said the following:

> Um, it’s [client’s saliva] kind of gross I guess it’s kind of a bit sticky and not very, yeah, not very effective. [Participant 4; aged 21 years; less than a year in the sex industry; brothel based]

In general, most participants described preferring actual lubricant during sex for its effectiveness over using their own saliva as lubricant with clients. However, there were several participants who were comfortable using their own saliva on occasion. Among these participants, saliva was only used as lubricant because of the convenience, with most still preferring lubricant if it was on hand.

Participants were not explicitly asked to reflect on why they might be comfortable with a client performing cunnilingus on them without a dental dam but not having a client use their saliva as a lubricant for vaginal sex. However, several participants raised this of their own accord. For one participant, the reasoning for this was because the saliva used as a lubricant would go inside her body, whereas she imagined that the client’s saliva during cunnilingus is not necessarily “inserted” into her vagina.

**Pleasure With Clients: A “Perk” or Not Part of the Job**

Most participants felt that their pleasure in a booking would not influence their decision to engage in oral sex practices with clients but rather was seen as a fringe benefit:

> No, not really. If I’m enjoying it, then I’m enjoying it, and that’s just a lucky perk. I wouldn’t stray from my boundaries just because they’ve kind of sprung something on me or whatnot. My service is my service and I stick to it. [Participant 12; aged 26 years; 8 years in the sex industry; brothel based]

Of these participants who said that pleasure was not a factor in their decision to engage in oral sex practices with clients, several felt that actually feeling pleasure from sex work was not possible as they compartmentalized it as work rather than for enjoyment. Others felt that pleasure during sex work was not possible due to the lack of emotional connection:

> I find that I, um, even if I have like the best client in the world, I’m never gonna, like have a pleasurable experience at work, because you’re always just like, on your guard a little bit, and you’re, I’m at work, you know...like I’m like just totally zoned out (laughs) Like they don’t think that, but I’m like thinking about what I’m going to eat for dinner, and, just like whatever, and so...I’ll never be like “I-I really wanna do this thing, and like get pleasure out of this” that
will never happen...But, I will definitely make choices to avoid, like, things that I know I definitely don’t like. [Participant 2; aged 26 years; 2 years in the sex industry; brothel based]

However, there were several participants for whom their pleasure, while infrequent with clients, could influence the duration or style of sex practices with clients, most notably kissing and cunnilingus (but not condom use). These participants described kissing for longer or deeper if they were enjoying the service with the client and, in some cases, allowing the client to kiss or perform cunnilingus on them if they had not paid for the service or simply to perform cunnilingus on them for longer if it was enjoyable:

...like if it was someone where there was a lot of saliva and it wasn't really feeling very good, I probably would let them do that for a while and then I’d just sort of be like “oh, it’s your turn now.” But if I’m having a really good time and it’s feeling really great, I’m not going to interrupt, and stop that person from what they’re doing sooner, if it means that I might get to have a really great time or even reach climax. You just kind of go along with it a bit more when it’s like, a really good time. But not to the point where I would, you know, not to the point where I would have unprotected oral sex on them, or oral, or vaginal sex with me. [Participant 14; aged 32 years; 2 years in the sex industry; brothel based]

Among the participants who felt that their pleasure sometimes influenced their oral sexual practices with clients, all but one stated that sex work became more enjoyable over time after they became more confident in their work:

[I enjoy it more now] Because I’m more relaxed with my clients, I’m more confident and experienced and I like to connect with them. Whereas before, I was like I don’t want to get to know you, but now I pretty much make friends with all my clients, so it’s more enjoyable. [Participant 7; aged 24 years; 4 years in the sex industry; brothel based]

**Mouthwash Use With Clients: A Freshener and Germ Killer**

Almost half of the participants reported using mouthwash before and after clients at work as a matter of routine. The main reason for using mouthwash before a client was to freshen up their breath before the booking. The main reason for using mouthwash after a client was to kill any germs they might have gotten from the client during the booking, particularly germs that cause bad breath and the common cold; however, gonorrhea was also occasionally mentioned. The other reason was to “clean the slate” for the next client; using mouthwash helped them feel refreshed and ready to see the next client:

...when I use mouthwash and redo my make-up, it’s like okay, that client’s finished, done with, go out and I’m ready to present again for the next client. [Participant 9; aged 18 years; less than a month in the sex industry; brothel based]

The remaining half of participants used mouthwash less routinely and tended to use it before a booking if they felt that they needed to freshen their breath or after a client who was a smoker or had left a feeling of uncleanness in their mouth. There were only a few participants who never used mouthwash at work, and these participants did not believe mouthwash was good for their oral health.

Most participants preferred clients to use mouthwash before bookings so that they had fresher breath and were less likely to transmit germs. Generally, for these participants, the brothels provide mouthwash and disposable cups for clients to use, and the mouthwash is poured and waiting for them when they enter the room for the booking. In this way, the participant does not have to ask the client to use mouthwash directly, which several described they would feel too uncomfortable doing as it might be considered rude. However, several others reported that they did not hesitate to ask clients to use the mouthwash if they noticed that it had not been used or if their breath smelled.

One participant, the only participant in our interviews who was a case upon recruitment (had oropharyngeal gonorrhea), made clients use a “cocktail” of mouthwash she brought from home for a full minute that combined Betadine, an antibacterial sore throat gargle, with Cepacol, an antibacterial mouthwash, which she used before and after every client as well. This participant had been using this “cocktail” of mouthwash before every client for 2 months before she became infected with oropharyngeal gonorrhea. This participant did not have outside-of-work sexual partners and did not perform fellatio without a condom with clients, nor did she report any other oral sex practices (ie, rimming or spit play) aside from tongue kissing. She felt that her “cocktail” of mouthwash kept her safe from germs that cause colds and gave her the security to engage in “passionate” tongue kissing that she felt gave her a competitive edge over other sex workers:

Like, if I don’t kiss at all, even lips, they will feel like “oh, this girl, she’s terrible, I’m not coming back” Like how can I have a passionate moment if I don’t kiss, you know? I try to trick them, like, oh it’s a passionate kissing without opening my mouth, you know what I mean? But like, if I do kiss with tongue, they love it. Like, they fall in love, they will always come back, and they will tell their friends, so their friends will come. Like most of my regular clients, I get them because of that, that treatment, you know? But obviously I won’t do that to someone I don’t feel comfortable. But it does help me to make a lot of money, you know. Have a lot of clients, and regular clients. And you know what because I do the mouthwashing, I ask them to do, I put hand sanitizer in their hands, I mean I ask them to use, they have a perception that I am a very clean person as well, and they always feel safe with me as well, you know? [Participant 17; aged 33 years; 2 years in the sex industry; brothel based]
Discussion

Principal Findings

This mixed methods study identified risk factors for oropharyngeal gonorrhea among sex workers in Australia’s 2 most populous cities, and it provided some explanations for the factors and forces that underpin oral sex and mouthwash practices. Our findings show that performing condomless fellatio on clients is associated with oropharyngeal gonorrhea. Furthermore, sex workers frequently tongue kiss clients as previously reported [16,41], yet this practice was not a significant risk factor for oropharyngeal gonorrhea. From our interviews with sex workers in Melbourne, almost all reported kissing clients due to demand for “the girlfriend experience” and generally reported following the client’s lead with regard to kissing style and duration. However, they used condoms for fellatio, often citing that it is illegal not to in Victoria, and they considered condomless fellatio a risky practice for contracting STIs, unlike cunnilingus without a dental dam. Client saliva use as a lubricant was similarly often viewed as risky and “gross” even among those who engaged in cunnilingus with clients without a dental dam.

Our study showed high proportions of condom use among sex workers for fellatio with clients. Previous studies have reported that 79% [16] of female sex workers in Melbourne and 75% of female sex workers in Sydney used condoms for all fellatio activities with clients in an average working week [35] in the previous 3 months. It is possible that sex workers who are not attending sexual health clinics routinely may practice condomless fellatio more often with clients. One study from Sydney in 2017 examining advertisements for private sex work found that half of the female sex workers with web-based profiles were offering condomless fellatio (170/339, 50.2%) [42]; however, it is not clear how often these sex workers actually practice condomless fellatio in an average working week.

In our multivariable analysis, kissing was not a significant risk factor for oropharyngeal gonorrhea; however, it should be noted that the adjusted OR of having oropharyngeal gonorrhea was 1.5 (95% CI 0.7-3.5) among sex workers who tongue kissed male clients compared to those who did not tongue kiss male clients. The OR was >1 even though it was not statistically significant, suggesting that kissing could be a potential risk factor for oropharyngeal gonorrhea and may have some clinical and public health implications. The non–statistically significant result may be due to the limited sample size to have sufficient power to detect the difference as the study was ceased earlier due to the COVID-19 pandemic. Given the overlapping nature of sex practices, particularly kissing, this is a common issue when determining risk factors for oropharyngeal gonorrhea [9]. It is interesting to note that the one case who was interviewed for the qualitative study had only reported tongue kissing clients in the previous 2 weeks as a risk factor for oropharyngeal gonorrhea as she did not have condomless fellatio with clients or engage in other practices with clients that involved saliva and she had no noncommercial sexual partners. However, previous research has found kissing to be a risk factor for oropharyngeal gonorrhea among men who have sex with men, and performing fellatio was not [8]. This is likely due to differences in testing practices and risk perceptions among men who have sex with men compared to heterosexual men, who are more likely to be clients of the sex workers in our study.

There have been previous case reports and epidemiological studies suggesting that tongue kissing can transmit oropharyngeal gonorrhea [8,43,44]. It is also interesting that this participant was diligent about using and having her clients use a particular cocktail of mouthwash (Betadine and Cepacol) before each booking, and after the booking in her case, because she was afraid of catching a cold from a client and having to take time off work. While investigations are being made into the role of Listerine [22,23], no studies to our knowledge have investigated the role of antibacterial mouthwashes such as Betadine and Cepacol in the transmission of gonorrhea. It is possible that the mouthwash cocktail she was using with her clients altered her oral microbiome in an unfavorable way and increased her odds of contracting gonorrhea. A previous study examining the incidence of syphilis among 96 men who have sex with men in Indonesia (2019) found that using antibacterial mouthwash containing chlorhexidine increased the odds of syphilis acquisition, and it was similarly suggested that this could be due in part to changes in the oral microbiome [45]. An Australian clinical trial has shown that daily use of Listerine or Biotene for 12 weeks had no significant effect on the oral microbiome [46]; however, no study to our knowledge has examined the effect of Betadine and Cepacol used in combination on the oral microbiome. In any event, it is clear from our interviews that, for some sex workers, mouthwash use gives the feeling of protecting against bacteria and viruses and is thus used as a safety precaution (in addition to being widely used for hygiene purposes), therefore establishing that the role of mouthwash use with regard to gonorrhea prevention is important to inform best practice.

Our findings from the case-control study that frequent mouthwash use was not associated with oropharyngeal gonorrhea positivity are consistent with those of a previous study among men who have sex with men that found no association between using any mouthwash daily and oropharyngeal gonorrhea positivity [47]. The duration and method of using mouthwash (ie, rinsing or gargling) can vary between individuals, and it remains to be seen whether duration and method could influence the ability of mouthwash to reduce the amount of gonorrhea bacteria in the oropharynx. Among a study of at-risk populations for oropharyngeal gonorrhea, female sex workers used mouthwash for the shortest duration (median 14 seconds); however, this was not a significant difference [48]. Future recommendations with regard to mouthwash use in this population, should it be found to be beneficial for reducing gonorrhea transmission, should take into account duration and method of use.

Our results showed no significant difference between controls and cases in the number of clients on whom they performed fellatio. A previous case-control study among men who have sex with men in Melbourne (2018) found that the number of casual partners in the previous 3 months was an independent risk factor for oropharyngeal gonorrhea [9]. When limiting our
analysis to only those sex workers who reported performing fellatio on clients (excluding those who did not perform this activity in an average working week) during an average working week, cases had fewer clients with whom they engaged in fellatio (median 5 for cases; median 7 for controls; \( P = 0.01 \); data not shown in the Results section). It is possible that female sex workers who perform condomless fellatio on clients see fewer clients in an average working week given that they may make more money per client offering this service compared to only offering fellatio with a condom as some participants in the qualitative interviews reported being offered more money by clients to perform “natural” or condomless fellatio (though none of our participants reported accepting these offers, in part due to concerns over STI transmission). It is also possible that having fewer clients is part of a risk reduction strategy among female sex workers who perform condomless fellatio. Further research is required to investigate why female sex workers who offer condomless fellatio might perform fellatio on fewer clients than those who do not.

There was a higher proportion of cases who reported tongue kissing clients in an average working week compared to controls (51/60, 85% of cases tongue kissed clients in an average working week compared to 348/494, 70.4% of controls). In contrast to performing fellatio on clients, among those who said yes to kissing clients, there were no differences in the number of clients that cases tongue kissed in an average working week compared to controls. In our study, there were 60.1% (399/664) of female sex workers who answered yes to kissing clients in an average working week, which is lower than what a cross-sectional survey of female sex workers from the MSHC in 2018 found (83.7%) [16]. This could be in part due to recruitment for this study occurring at 2 sites and including a higher proportion of non-Australian female sex workers as the 2018 study showed that Asian language–speaking female sex workers were significantly less likely than English-speaking female sex workers to tongue kiss clients [16]. Female sex workers recruited from the SSHC and those born in China, Thailand, or other countries were less likely than those recruited at the MSHC and those from Australia or New Zealand to report tongue kissing (data not shown).

For some participants in our qualitative interviews, kissing and cunnilingus were the only sexual practices that might be influenced by how much pleasure the sex worker was experiencing during the booking, though pleasure impacting these behaviors was generally reported as a rare occurrence. These participants described kissing for longer or deeper if they were enjoying the booking. However, most of the participants in our study described experiencing pleasure at work as an infrequent or nonexistent occurrence. There has been limited research into pleasure for sex workers during sex work. A previous qualitative study of 9 female sex workers in Victoria explored this concept and reported that, for some women, sexual pleasure was possible with a client only after developing intimacy through seeing them multiple times; however, this study was exploratory and specifically recruited women who had positive experiences of sex work [49]. Further research could clarify the extent to which sex workers experience sexual pleasure at work and whether this impacts sexual practice.

The main limitation to this study is that most of the cases were recruited from one clinic (SSHCL) and epidemiology may vary due to environmental and spatial factors. There was also a noted delay in recruiting cases at MSHC compared to SSHCL, which could indicate a higher rate of declining to participate. One reason for this may be the varying laws regarding sex work in the state of Victoria versus NSW, whereby performing condomless oral sex was illegal in Victoria but not in NSW. This may have been a deterrent for taking part in this study as participants may not have felt comfortable reporting any sex practices that were illegal. Previous research has shown that regulated and criminalized sex work often discourages female sex workers to seek health care [50], so it could be that they are less likely to disclose their sex practices to clinicians, which would make it difficult to assess risk. Despite a slower recruitment of cases from the MSHC, among those sex workers who participated in the survey, participants from the MSHC were no more likely to decline to report sexual practices compared to those from the SSHC after adjusting for oropharyngeal gonorrhea diagnosis, age, and country of birth. Only being born in China or other countries aside from Thailand and New Zealand was associated with declining to report sex practices on the survey.

Another limitation of this study was that our convenience sampling may have created a bias toward those who attended sexual health clinics for HIV and STI screening (or presented with symptoms), and this may not be generalizable to the entire sex worker population, including those who did not attend a sexual health clinic. Most of our participants worked in brothels and massage parlors, and only 0.2% (1/664) participated in street-based sex work; thus, the findings may not be generalizable to those engaging in street-based sex work.

A final limitation of this study is that interviews for the qualitative component were only offered in Victoria due to the financial and logistic difficulties of the interviews being conducted only in English and by researchers based only at the MSHC. However, similar to all qualitative data, the qualitative component was not meant to be generalizable to the wider population of sex workers. Rather, this component of our study provided additional depth and understanding to the data collected in the quantitative component among Melbourne-based female sex workers. These interviews were also cut short due to the COVID-19 pandemic; however, sufficient meaning was generated from the data to answer the questions of clinical interest.

The availability of phone interviews in combination with face-to-face interviews can be considered a strength of our study as it allowed participants to freely disclose personal and sensitive information. We found that our interviews over the phone provided rich data that were comparable to, and in some cases deeper, than those from our face-to-face interviews. Allowing the option of phone interviews, particularly asking the participants to select a phone or face-to-face interview upon recruitment, likely encouraged a wider array of participants as it is possible that some participants who would have been uncomfortable with a face-to-face interview opted to share their experiences over the phone rather than declining to participate. Our findings in this case reflect those of other studies that have
shown no difference in data quality between phone interviews and face-to-face interviews [51,52].

Conclusions
Our study shows that condomless fellatio is a risk factor for oropharyngeal gonorrhea among sex workers despite most sex workers using condoms with their clients for fellatio. Novel interventions, particularly targeting the oropharynx, will be required for oropharyngeal gonorrhea prevention.

Acknowledgments
The authors acknowledge the contributions of the participants in the case-control and qualitative components of this study and thank them for their time and insights. The authors would like to acknowledge the Sex Workers Outreach Project and Resourcing Health & Education for their assistance in reviewing the case-control questionnaire and the qualitative interview schedule. They would also like to acknowledge Mark Chung for designing the study graphics, Sabrina Trumpour for assistance with data entry and transcription, Sandra Walker for her help in drafting the study questionnaire, and Jana Sisnowski for her help preparing the study documents for the Sydney Sexual Health Centre (SSHC). The authors would also like to acknowledge Maggie Ma and Birdie Thirapat at the SSHC and the clinicians at the SSHC and the Melbourne Sexual Health Centre for their help recruiting participants for the Health Research in Sex Workers study. EPFC and JJO are supported by an Australian National Health and Medical Research Council (NHMRC) Emerging Leadership Investigator Grant (GNT1172873 and GNT1193955, respectively). CKF is supported by an NHMRC Leadership Investigator Grant (GNT1172900). JEB is supported by an ARC Discovery Early Career Researcher Award grant (DE200100049). BD is supported by an NHMRC Practitioner Fellowship (GNT1154828).

Data Availability
The data sets generated during and analyzed during this study are not publicly available due to ethics requirements considering the confidential nature of patient data. Further information can be obtained by contacting the Alfred Hospital Ethics Committee, Melbourne, Victoria, Australia (project 596/17).

Authors' Contributions
TRP, CKF, AM, BD, RG, RV, DC, JEB, and EPFC were involved in study conceptualization and design. KM, RM, and RW were involved in study recruitment and management. TRP and KM conducted the interviews. TRP analyzed the data and wrote the first draft of the manuscript. JEB and EPFC provided supervision. All authors were involved in manuscript revision.

Conflicts of Interest
None declared.

Multimedia Appendix 1
A map of the data overlaid with variables examined.

Multimedia Appendix 2
Recruitment flowchart.

Multimedia Appendix 3
Associations with declining to answer sexual practice questions in survey among 664 female sex workers.

References


32. Engel JL, Fairley CK, Greaves KE, Vodstrcil LA, Ong JJ, Bradshaw CS, et al. Patterns of sexual practices, sexually transmitted infections and other genital infections in women who have sex with women only (WSWO), women who have sex with men only (WSM), and women who have sex with men and women (WSMW): findings from a sexual health clinic in Melbourne, Australia, 2011-2019. Arch Sex Behav 2022 Jul;51(5):2651-2665. [doi: 10.1007/s10508-022-02311-w] [Medline: 35776396]


**Abbreviations**

- AMR: antimicrobial resistance
- HERS: Health Research in Sex Workers
- MSHC: Melbourne Sexual Health Centre
- NSW: New South Wales
- OR: odds ratio
- RhED: Resourcing Health & Education
- SSHC: Sydney Sexual Health Centre
- STI: sexually transmitted infection
- SWOP: Sex Workers Outreach Project
Protocol


Xinchun Gu1*, PhD; Conall Watson2*, PhD; Utkarsh Agrawal1*, PhD; Heather Whitaker3*, PhD; William H. Elson1*, MBBS, MSc; Sneha Anand1, PhD; Ray Borrow4, PhD; Anna Buckingham5, BMedSci, BMBS, MRCP; Elizabeth Button1, MSc; Lottie Curtis6, MSc; Dominic Dunn1; Alex J. Elliot7, PhD; Filipa Ferreira1, PhD; Rosalind Goudie1, MSc; Uy Hoang1, BSc, MBBS, MPH, MD; Katja Hoschler8, PhD; Gavin Jamie1, MSc; Debasish Kar1, MSc; Beatrix Kele8, PhD; Meredith Leston1, MSc; Ezra Linley4, PhD; Jack Macartney1; Gemma L Marsden6, BSc, MSc, PhD; Cecilia Okusi1, MRES; Omid Parvizi1,8, PhD; Catherine Quinot2, PhD; Praveen Sebastianpillai9, BSc; Vanashree Sexton1, PhD; Gillian Smith5, MBBS; Timea Suli1, PhD; Nicholas P B Thomas6, BSc, PhD; Catherine Thompson8, PhD; Daniel Todkill1, MBChB; Rashmi Wimalaratna1, BSc, MBBS; Matthew Inada-Kim5, MBBS; Nick Andrews3, PhD; Victoria Tzortziou-Brown6, PhD; Rachel Byford1, BA; Maria Zambon9, PhD; Jamie Lopez-Bernal2, PhD; Simon de Lusignan1, MSc, MD, CITP

1Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
2Immunisation and Vaccine-Preventable Diseases Division, UK Health Security Agency, London, United Kingdom
3Statistics, Modelling and Economics Department, UK Health Security Agency, London, United Kingdom
4Vaccine Evaluation Unit, UK Health Security Agency, Manchester, United Kingdom
5NHS England, London, United Kingdom
6Royal College of General Practitioners, London, United Kingdom
7Real-time Syndromic Surveillance Team, UK Health Security Agency, Birmingham, United Kingdom
8Respiratory Virus Unit, UK Health Security Agency, London, United Kingdom
9Virus Reference Department, UK Health Security Agency, London, United Kingdom
*these authors contributed equally

Corresponding Author:
Simon de Lusignan, MSc, MD, CITP
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Primary Care Building, Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 01865 617 283 ext 17 283
Email: simon.delusignan@phc.ox.ac.uk

Abstract

Background: Prepandemic sentinel surveillance focused on improved management of winter pressures, with influenza-like illness (ILI) being the key clinical indicator. The World Health Organization (WHO) global standards for influenza surveillance include monitoring acute respiratory infection (ARI) and ILI. The WHO’s mosaic framework recommends that the surveillance strategies of countries include the virological monitoring of respiratory viruses with pandemic potential such as influenza. The Oxford-Royal College of General Practitioner Research and Surveillance Centre (RSC) in collaboration with the UK Health Security Agency (UKHSA) has provided sentinel surveillance since 1967, including virology since 1993.

Objective: We aim to describe the RSC’s plans for sentinel surveillance in the 2023-2024 season and evaluate these plans against the WHO mosaic framework.
Methods: Our approach, which includes patient and public involvement, contributes to surveillance objectives across all 3 domains of the mosaic framework. We will generate an ARI phenotype to enable reporting of this indicator in addition to ILI. These data will support UKHSA’s sentinel surveillance, including vaccine effectiveness and burden of disease studies. The panel of virology tests analyzed in UKHSA’s reference laboratory will remain unchanged, with additional plans for point-of-care testing, pneumococcus testing, and asymptomatic screening. Our sampling framework for serological surveillance will provide greater representativeness and more samples from younger people. We will create a biomedical resource that enables linkage between clinical data held in the RSC and virology data, including sequencing data, held by the UKHSA. We describe the governance framework for the RSC.

Results: We are co-designing our communication about data sharing and sampling, contextualized by the mosaic framework, with national and general practice patient and public involvement groups. We present our ARI digital phenotype and the key data RSC network members are requested to include in computerized medical records. We will share data with the UKHSA to report vaccine effectiveness for COVID-19 and influenza, assess the disease burden of respiratory syncytial virus, and perform syndromic surveillance. Virological surveillance will include COVID-19, influenza, respiratory syncytial virus, and other common respiratory viruses. We plan to pilot point-of-care testing for group A streptococcus, urine tests for pneumococcus, and asymptomatic testing. We will integrate test requests and results with the laboratory-computerized medical record system. A biomedical resource will enable research linking clinical data to virology data. The legal basis for the RSC’s pseudonymized data extract is The Health Service (Control of Patient Information) Regulations 2002, and all nonsurveillance uses require research ethics approval.

Conclusions: The RSC extended its surveillance activities to meet more but not all of the mosaic framework’s objectives. We have introduced an ARI indicator. We seek to expand our surveillance scope and could do more around transmissibility and the benefits and risks of nonvaccine therapies.

(KJMIR Public Health Surveill 2024;10:e52047) doi:10.2196/52047

KEYWORDS
sentinel surveillance; pandemic; COVID-19; human influenza; influenza vaccines; respiratory tract infections; vaccination; World Health Organization; respiratory syncytial virus; phenotype; computerized medical record system

Introduction

Prior to the COVID-19 pandemic, sentinel surveillance was orientated toward influenza and its associated winter pressures [1-6]. It has subsequently evolved to include a systematic collection of acute respiratory infections (ARIs) and a wider range of indicators. The World Health Organization (WHO) Global Influenza Surveillance and Response System (GISRS) was launched in 1952 to provide a global response to influenza and other respiratory infections [7,8]. The focus of the GISRS and other national surveillance networks is seasonal influenza monitoring and associated vaccination effectiveness studies [9,10], as well as pandemic preparedness [11,12]. Viral testing became an essential component, with serosurveillance introduced into some systems [13]. Prepandemic virological testing was largely carried out in the winter season with influenza-like illness (ILI) as the key clinical indicator of community influenza infection [14]. In addition to ILI, ARI started to be used by the European Centre for Disease Prevention and Control as a surveillance indicator [15,16]. The WHO 2013 global epidemiological standards for influenza surveillance proposed the use of severe ARI (SARI) as an indicator. SARI is defined as an incident ARI in a person admitted to a hospital [17,18]. The WHO mosaic framework suggests sentinel ILI, ARI, and SARI surveillance as the core approach for monitoring the epidemiological characteristics of respiratory viruses in interpandemic periods [16].

Subsequent to the COVID-19 pandemic, the WHO published its mosaic framework for respiratory disease surveillance [16,19]. It recommended that the surveillance strategies of countries include the virological monitoring of influenza, SARS-CoV-2, respiratory syncytial virus (RSV), and other viruses with pandemic potential. The mosaic has a broad framework and includes 14 surveillance objectives set out across three domains: (1) detection and assessment of respiratory viruses; (2) monitoring their epidemiological characteristics; and (3) informing on the use of health interventions [16,19].

The Royal College of General Practitioners (RCGP) has been collecting data about respiratory and other infections in England in its epidemic research center since 1957 [20]. This research center became rebranded as the RCGP Research and Surveillance Centre (RSC) and has been conducting sentinel surveillance since 1967 [21], in collaboration with the UK Health Security Agency (UKHSA) and its predecessor bodies. The RSC has included reference laboratory virology since the 1993-1994 season [22]. The network has grown to almost 2000 practices in England and Wales (31.6% of the active practices) with a contemporary extract of over 19 million patients (31.9% of the England and Wales population) in 2023 [23].

This protocol describes the Oxford-RCGP RSC’s plans for the 2023-2024 sentinel surveillance season and evaluates them against the WHO mosaic framework. The RSC will be offering all-year-round virology and sentinel surveillance of respiratory infections in collaboration with the UKHSA. Our extended surveillance includes the adoption of SARI as an important severity indicator, alongside ILI and other components of the WHO mosaic framework.

The objectives are as follows: (1) describe the planned patient and public involvement (PPI) with the RSC, with the aims of improving public understanding of the RSC’s program and co-designing changes to our sentinel surveillance; (2) develop
an ARI digital phenotype and contemporaneously report the incidence of ARI and SARI, including new severity indicators, using primary care data; (3) collect and share high-quality data to support vaccine effectiveness (VE) studies for COVID-19 and influenza vaccines in the coming season and enable the reporting of RSV’s disease burden; (4) ensure that the volume of virology and serology sampling from member practices following our sampling framework is sufficient to determine VE by vaccine type and has the minimum required clinical data recorded; (5) introduce technological developments by using general practitioner (GP) and laboratory links to support virology and serology sampling, establishing a messaging system to enable more representative sampling and targeted sampling when required, increasing point-of-care testing (POCT) capability, piloting virology sampling from asymptomatic individuals, and testing for pneumococcus infection; (6) create a biomedical resource that provides a unique longitudinal clinical resource and enables genomic surveillance by linking individual-level human phenotypes to the genomic sequences of viruses detected in those individuals; and (7) describe the legal basis and governance framework for conducting sentinel surveillance.

Methods

Comparison With the WHO Mosaic Framework

We describe our approach to sentinel surveillance in functional components. Many of these, such as PPI, span across all of the mosaic’s sentinel objectives. Others sit outside or beyond its scope, for example, bacterial causes of infection and information governance requirements. We conclude the results section with a table summarizing the surveillance objectives achieved and those to be delivered beyond its scope.

PPI Group

We will use 2 channels of PPI and engagement within the RSC. The first channel is with national PPI groups, the Health Data Research-UK and the RCGP Patient and Carer Participation Groups. Additional national patient representatives will be recruited from the People in Research portal of the National Institute for Health and Care Research [24] to ensure geographical representation across England. The second channel includes local patient participation groups of general practice members of the RSC network. Opportunities for patients to provide feedback will be communicated to all network practices in England.

We will invite patient representatives to participate in 2 meetings per year held with the UKHSA and Oxford-RCGP RSC. The meetings aim to raise awareness of surveillance, including its use of patient data, and to gain input on the RSC’s work, particularly on optimizing the acceptability of sampling and the feedback of results. PPI feedback will be used mainly to improve communication with patients, but may inform other areas of the surveillance. We will send a monthly newsletter covering topics and findings related to surveillance to increase transparency and patient engagement (Multimedia Appendix 1). We will use the cube framework to plan and evaluate PPI [25]. To improve transparency in reporting PPI and its impact, we will use a standard international guideline (GRIPP2 checklist) [26].

ARI Phenotype

A digital phenotype is a set of rules that allows the identification of cases, such as ILI or ARI, from the computerized medical record (CMR) [27]. At the RSC, we have a well-established digital phenotype for ILI. To bring our strategy in line with the most recent WHO recommendations, we are developing a new ARI phenotype for the coming season.

The phenotype is being built around Systematized Nomenclature of Medicine (SNOMED) Clinical Terms (CT) as this is the medical terminology mandated by the National Health Service (NHS) and used by all primary care providers [28]. The phenotype consists of code lists representing common conditions that make up the ARI concept. This allows the identification of ARI events coded by member practices in the CMR.

We plan to build on the ARI phenotype by characterizing the nature and severity of ARI events. We will do this by exploring primary care codes recorded in association with ARI and by linking the primary care data to secondary care records [29,30]. We differentiate ARI from SARI by checking whether ARI leads to hospitalization, as recorded in the CMR.

Our phenotype development uses the Phenotype Execution Modeling Architecture (PhEMA) toolkit [31]. The Health Level-7 Fast Healthcare Interoperability Resources (FHIR) is a global standard for passing health care data between systems [31] and is used by the NHS. The PhEMA comprises developing an ontological layer using clinical query language and then presenting a “code list” using the Health Level-7 FHIR value set format or as a SNOMED CT refset. For example, the SNOMED clinical term “Lower respiratory tract infection (disorder)” (SCTID: 5041700) has 10 child codes (also known as subtypes), which are all automatically included. If extra child codes are included, they will automatically be included in our definition, unlike the extensional process where “code lists” need frequent review [32]. We have a “Helper Tool” developed in-house to facilitate the selection of SNOMED CT (Figure 1).
Figure 1. Tool used to facilitate the intentional identification of Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) refsets. The lower figure demonstrates how SNOMED CT supertypes and subtypes are included or excluded using the Helper Tool.

High-Quality Data to Support Influenza and COVID-19 VE and Other Studies

We will collect and share high-quality data, such as primary care CMR and vaccine exposure data, to support VE studies for COVID-19 and influenza vaccines and enable the reporting of RSV’s disease burden in the coming season. We will focus on the data collection areas that need better data quality as follows: (1) Identifying and coding cases of ARI, SARI, and ILI as a problem title in CMRs, including whether these are first or new episodes [33]; (2) Consistent capturing of symptoms, signs, and any markers of severity upon the presentation of people with ARI; (3) Complete recording of vaccine exposure data, particularly for influenza and COVID-19 vaccines, but other vaccines, such as pneumococcus, may be relevant; (4) Complete recording of sociodemographic and comorbidity data; and (5) Recording of outcomes in both primary care and linked secondary care data sets. This is particularly important for people who have had virology sampling (see the next section) and practices that are incentivized through data quality payments.

We also need sufficient data (ie, sample size) to support UKHSA’s syndromic surveillance [34,35], our Weekly Return [36,37], and our Annual Report [38]. The Weekly Return and Annual Report will be in their 56th year of production. This will set out to make the network as nationally and regionally representative as possible by recruiting additional practices, particularly to improve virology sampling.

Virology and Serology Sampling and Testing

We collect serology samples from volunteer patients who are attending their practice for a routine blood sample appointment by asking them to contribute one more blood sample to the serology surveillance. Patients are invited through messages before their routine blood tests and are provided a link for further information about the serology surveillance. Verbal consent is taken from the patients before the sampling. We aim to collect 500 serology samples per month from each of the 3 age groups (younger than 18 years, 18 to 64 years, and 65 years or older) with representative sampling across the network. We plan to divide the younger than 18 years age group based on immunization age groups in the future to allow more granular information to be collected. Serology samples are batch processed in different labs depending on operational needs, for example, understanding vaccine waning in immunocompromised people.

We collect virology samples (nasal and pharyngeal swabs) all year round from patients who present ILI or ARI symptoms to general practices and meanwhile capture their symptom onset day. We aim to reach 1000 samples per week every week in the coming year. Virological samples collected in RSC surveillance practices are analyzed in UKHSA’s respiratory virus reference laboratory in Colindale, London. Practices code the results returned from the UKHSA according to SNOMED CT or Read Code (Multimedia Appendix 2), along with the date that the swab was taken. The test results are used for virological surveillance as well as test-negative case-control studies [39] that evaluate VE for influenza and COVID-19 vaccines.

The UKHSA tests for a panel of 8 viruses (Figure 2), including (1) SARS-CoV-2; (2) influenza (influenza A subtypes are differentiated based on their hemagglutinin [H] and neuraminidase [N] surface proteins; the 2 subtypes that are commonly in circulation are H1N1 and H3N2, though some influenza A cases are only reported as influenza A; influenza B is reported collectively, with close monitoring of circulating lineages); (3) RSV A and B; (4) human metapneumovirus (hMPV); (5) other seasonal coronaviruses (NL63, 229E, OC43, and HKU1), in addition to SARS-CoV-2; (6) adenovirus; (7) human rhinovirus; and (8) enterovirus.

To reach a sufficient volume of tests, the behavioral change we want primary care clinicians to achieve is conducting virology and serology sampling where possible and making CMR system entries of high data quality. We will use the behavioral change wheel, otherwise known as the COM-B model, to theorize behavioral change expected in clinicians. This suggests that for a behavioral change (B) to take place, an individual requires capability (C), opportunity (O), and motivation (M) [40]. We are using the COM-B model as a high-level theoretical frame to describe our interaction with practices rather than a formal experimental work.
New Technologies and Capabilities

We have an ambitious program of introducing new technologies and extending the scope of our surveillance. We are looking to introduce 3 new technologies in the coming year: (1) LabLinks, integrating our surveillance sampling with online pathology test requests carried out via primary care CMR systems; (2) creating a POCT-ready nested cohort within the RSC; and (3) introducing a messaging system that enables us to effectively target specific groups for sampling.

The 2 surveillance capabilities we are looking to introduce are (1) urinary antigen (UAG) tests that test for pneumococcus infection, and (2) collecting virology swabs from asymptomatic infections.

We will conduct a feasibility study of testing for *Streptococcus pneumoniae* infection using UAG. We would like to perform this study over 1 to 2 years, recruiting 2 practices from each NHS region, with ARI patients having both respiratory swabs and urine tests. We will also pilot asymptomatic testing for our panel of viral tests [41]. Running this within the RSC will enable linkage of virology results to medical records and the monitoring of household transmission as we can identify people in the same household [42-44]. The protocols for the feasibility and pilot studies will be published separately.

Creation of a Biomedical Resource

We will complete the work to create a unique biomedical resource that offers 2 unique opportunities for research. First, we will create a longitudinal database that runs back to the start of the RSC’s involvement in sentinel surveillance in 1967. Second, we will link clinical records and the virology results for all tests performed at UKHSA’s reference laboratory. This will create a resource to support the emerging discipline of genomic surveillance [45,46] by linking clinical phenotype, as defined in an individual’s CMR, at the individual level with details of the infecting virus, including its genome sequence. Summary results of the unique longitudinal data are presented.

Ethical Considerations

Pseudonymized primary care data and samples from general practices collected for surveillance purposes are processed under Regulation 3 of the Health Service (Control of Patient Information Regulations 2002) and annually renewed under Regulation 7 by UKHSA’s Caldicott Guardian [47]. Any further research or studies require their own ethical approval and approval of the Joint RSC Committee of the University of Oxford and RCGP.

There are data-sharing agreements in place with every GP practice within the network, where the purpose of data collection and the processing of activities are stated. We link primary care data collected via sentinel surveillance to secondary care data provided by NHS England via a bespoke data sharing agreement that is renewed annually. All the data controllers (UKHSA, RCGP, and University of Oxford) complete NHS England’s data security and protection toolkit to meet the performance standards set by the National Data Guardian [48].

Results

PPI Group

We presented the protocol to patient representatives in July 2023. Patient representatives proposed promoting self-testing for virology to help increase the uptake of sampling. They also suggested considering swabbing in different settings, including a pharmacy, as patients with respiratory illnesses may be more likely to visit a pharmacy rather than a GP.

The patient groups have suggested that the following themes and content should form part of our PPI communications: (1) information about how patient data are used in surveillance, (2) information about how sentinel surveillance fits in with pandemic preparedness, and (3) further personal areas of interest, for example, to better understand patterns in patients with post–COVID-19 condition. The PPI representatives expressed that some patients may have concerns about patient confidentiality and data use, which can be overcome by written dissemination and regular engagement.
Based on initial feedback, we have tailored our communication to focus on key areas of interest to the public. For example, we have delivered a presentation and discussion group to patients at a Midlands practice, which focused on how data are used for surveillance and how transparency can be improved. We have developed a poster designed for practice waiting rooms to promote awareness of the RSC. We consider patient feedback in training our surveillance practices (eg, we inform new practices that patients like to receive their swab results, which may enhance patients’ experience of care).

**ARI Phenotype**

To ensure that we identify and classify all relevant coded events in the CMR, the ARI phenotype will be hierarchical. The top level will represent the overall ARI concept. The middle level will include codes for upper respiratory tract infection (URTI), lower respiratory tract infection (LRTI), exacerbations of chronic lung disease, and ILI. The bottom level will include codes for lower-level syndromes, such as sinusitis, that are descendants of middle-level syndromes (Figure 3). We will recommend that RSC network member practices consider coding ARI-associated diseases and symptoms as the diagnosis when patients present with ARI symptoms.

We have also developed a recommended coding sequence for patients presenting with ARI to primary care. The overall ARI and ILI signals are important for respiratory surveillance; hence, we recommend considering ILI first, using the RSC’s ILI definition: ARI with measured or clinically plausible temperature ≥38 °C, cough, systemic upset, such as headache or myalgia, sudden onset, and absence of a more plausible diagnosis. Our current surveillance categories do not reliably capture exacerbations of chronic lung diseases, which are vulnerable to severe outcomes from infective exacerbations [49,50]. We are therefore asking practices to consider this next. We then ask the consulting clinician to consider LRTI and subsequently URTI (Figure 4).

We are also standardizing the recording of symptoms, signs, and any emergency management decisions made for ARI events, so we request RSC practices to record these in the CMR. We fully understand the pressures on consulting time, so we will be producing data entry forms that practices can use in the major brands of CMR systems to facilitate high-quality data entry. The key symptom data to be coded are as follows: (1) date of onset of symptoms; (2) presence of “absence of fever” because infections, particularly in older people, may not be associated with fever [51]; (3) sore throat symptoms; (4) cough or no cough, and if coughing, is it productive; (5) coryza and nasal symptoms; and (6) presence or absence of shortness of breath and wheezing. The signs we would like to see coded are as follows: (1) measured tympanic temperature; (2) peripheral oxygen saturation, where available; (3) pulse rate; (4) respiratory rate; (5) upper respiratory signs where present, including cervical and anterior cervical lymphadenopathy, and any tonsillar exudate or enlargement; and (6) lower respiratory signs, including wheezing or other physical signs.

**Figure 3.** Research and Surveillance Centre recommendations for classifying the etiology of respiratory symptoms or signs in people presenting to primary care. COPD: chronic obstructive pulmonary disease; ILI: influenza-like illness; LRTI: lower respiratory tract infection; SoB: shortness of breath; URTI: upper respiratory tract infection.
**High-Quality Data to Support VE and Other Studies**

To enhance data quality, RSC member practices will receive feedback about their recording of key sociodemographic variables, vaccine exposure, and risk groups. We also provide a dashboard to enable RSC practices to compare their rate of vaccination with the rest of the network (Figure 5).

Most of the relevant data in VE studies are recorded as part of standard care, but some are not recorded to a satisfactory standard. For example, some important sociodemographic data are not automatically recorded, including ethnicity, smoking status, and obesity. Vaccine exposure data should include brand and batch wherever possible, but these recordings are problematic for vaccinations outside general practices.

Risk groups and patient outcomes are also important for VE and other studies, and most of these data are recorded well as part of chronic disease management. We used these data to derive the Cambridge Multimorbidity Score, a single measure of multimorbidity for all adults, and the electronic frailty index (eFI). While the eFI can be used from the age of 50 years, we will use it in the coming season for people aged 65 years or older [52,53].

Our data are linked to national collection of hospital and death data, so we can report severe outcomes [54]. Our data are also linkable with the National Immunization Management System, which provides vaccination data for COVID-19 and influenza in England. These data will be used to estimate VE, with mid-season and end-of-season studies for influenza, an autumn VE study for COVID-19 vaccines, and a burden of disease study for RSV. These results will contribute to the Joint Committee on Vaccination and Immunization impact of vaccine policy, contribute to WHO reviews, and be published in peer-reviewed journals.

The protocol for our VE studies is included in Multimedia Appendix 3.
Virology and Serology Sampling and Testing

Ongoing recruitment and checking the representativeness of the network by region will ensure the RSC is representative. Maps of the distribution of virology and serology sampling practices and the entire network are included in Multimedia Appendix 4. We plan to increase the number of virology samples taken each week. The total so far for the 2022-2023 season is 11,001, with 878 as the highest number of virology swabs collected in week 51 of the year 2022. The median weekly total is 259 samples, with an interquartile range of 173 to 304 samples. These data have been and will be used to estimate VE for influenza and COVID-19 vaccines with a test-negative design, and can be used in the future to study the VE of RSV.

Applying COM-B, our practice liaison team will be working with practices to achieve higher rates of virology samples. We will be continuing our regular visits and weekly and monthly reports to practices, and the scope will be driven by our learning from visits about how best to change (increase) sampling behavior and from our PPI input (Multimedia Appendix 5). Our virology dashboard provides practices with a comparison of what viruses are circulating in their practices compared with nationally (Figure 6). This can support practices to review their antibiotic prescriptions and antiviral medication uptake, and support better antibiotic stewardship.

We will need to control numbers in subgroups to achieve better representativeness of our serology sampling in the coming year, and produce a dashboard that practices can use to monitor activity (Figure 7). We will implement a technology-driven or manual approach to ensure national representativeness by age band and region. We will also offer pediatric phlebotomy training to encourage sampling in young people and children (Multimedia Appendix 6). The scale of serology sampling will be determined according to season.
Figure 6. Circulating virology data taken from our observatory showing the percentage of positive samples by viral strain for Research and Surveillance Centre (RCS) general practitioners (GPs) combined and an example of an individual practice (2022-2023 season). hMPV: human metapneumovirus; RSV: respiratory syncytial virus.

Figure 7. Serology data taken from our observatory and dashboard showing the number of samples collected by age band for Research and Surveillance Centre general practitioners combined (2022-2023 season).

New Technologies and Capabilities
We plan to integrate LabLinks into the RSC. Currently, customized kits are provided to RSC practices for virology and serology sampling, and these kits are returned to UKHSA laboratories through the post. We plan instead to integrate RSC sampling with the electronic pathology test–requesting system currently integrated into primary care CMR systems. The details of the LabLinks program are described in Multimedia Appendix 7. Figure 8 provides an overview of the process.

We will create a POCT nested cohort of practices willing to participate in feasibility studies. The key area of interest is POCT for group A streptococcal infection because there was a higher peak in the incidence of group A streptococcal infection in late 2022 than in the previous 8 years (Figure 9). A protocol for using molecular POCT is included in Multimedia Appendix 8 [55-57].

We will test EMIS Recruit [58] as a messaging system to invite targeted risk groups (immunocompromised) or younger people who had a booked blood test to consider volunteering to provide an extra blood sample for serology. The messaging system runs through EMIS Recruit to detect patients in target groups that have a recent blood test request. A message is sent to the patients to invite them to participate in our serology sampling. We will also be asking practices whether there would be interest in pediatric phlebotomy training to increase the sample number in younger children.

We will recruit one or two virology sampling practices per region that volunteer to collect urinary samples for UAG testing to infer pneumococcal infection.

We will start with children under 5 years of age coming for vaccination. We may then move on to include people aged 40 to 74 years attending NHS health checks and people aged 40 years or older attending hypertension clinics in primary care.
**Figure 8.** Virological surveillance data workflow. Samples are taken in practice. They pass through the local pathology laboratory to the UK Health Security Agency (UKHSA) viral reference laboratory at Colindale. Virology results go back to the general practitioner (GP) and patient and also into the Research and Surveillance Centre (RSC) database. ICE: Integrated Clinical Environment; LIMS: Laboratory Information Management System.

**Figure 9.** Weekly incidence of scarlet fever or streptococcal sore throat presenting to the English primary care sentinel network practices from 2010 to 2023. The incidence in the last quarter of 2022 was higher than that seen in the previous 10 years.

**Creation of a Biomedical Resource**

The RSC has unique longitudinal data stretching back to 1967. Currently, we are progressing with the assembly of these data into a single quinquagenarian resource (**Figure 10**). We have curated ILI incidences between 1967 and 2022 as an example (**Figure 11**).
Legal Basis and Governance Framework for Conducting Sentinel Surveillance

Surveillance is authorized each year through a commissioning letter, authorized by the UKHSA, and sent out to all RSC sentinel network general practices. Privacy notices for individuals registered at a GP within the network are publicly available [59]. We request all member practices to share these with their registered patients.

Evaluation of RSC Surveillance Compared With the WHO Mosaic Framework

The RSC meets many of the surveillance objectives of the WHO mosaic framework but within the scope of its virology plus primary and secondary care data (Table 1).

In Domain I (detection and assessment of respiratory viruses), we have a comprehensive virology panel and a nationally representative primary care network. Our data are strong but could be stronger with respect to having detailed information about clinical presentation. Our household key and information about residential care provide only limited information about transmission [43], and piloting whether there is asymptomatic spread will provide additional evidence about the spread of the disease.

In Domain II (epidemiological characteristics of respiratory viruses), our data are strong. Our linked data set allows us to monitor severe outcomes and mortality. We do not have sufficient coverage of all high-risk settings and do not collect data from hospitals where nosocomial infection is common. We can readily identify community-recorded vulnerable populations, and we term these “risk groups” [60]. We do not directly measure whether health care systems are overwhelmed, but we do record community rates of illnesses compared with other years [61].

In Domain III (informing about health interventions), we are able to infer the impact of interventions, such as lockdowns and shielding, during the COVID-19 pandemic, and we can see from our data the impact of school closures [62,63]. We are strong...
in measuring vaccine uptake and effectiveness, and our data have been used for vaccine adverse events [60,64]. We only have limited abilities to assess the effectiveness of some antivirals and other therapeutics owing to their central administration, and our capabilities are greater where these are recorded in the GP CMR. We have the capability to assess diagnostic tests and will compare POCT with reference virology laboratory results [55,65]. NHS England ARI hubs are planned to join the network, and we plan to evaluate the impact of these networks. We do not provide candidate vaccine viruses.

The RSC has additional surveillance objectives (Table 2). PPI and bacterial surveillance are essential. We could have a role at the system level of exploring how POCT might have an impact on treatment selection and health outcomes. Additionally, we see compliance with information governance standards as essential and linkage between clinical and viral sequencing data as enabling genomic surveillance.
# Table 1. Research and Surveillance Centre implementation of the World Health Organization mosaic framework.

<table>
<thead>
<tr>
<th>WHO domain and WHO surveillance objective</th>
<th>RSC delivery</th>
<th>Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain I: Detection and assessment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapidly detect outbreaks and other events</td>
<td>Our sentinel network covers over 32% (N=19 million) of the population and includes virology and serology sampling. A new ARI phenotype and POCT enhance the capability.</td>
<td>Full delivery: RSC and UKHSA</td>
</tr>
<tr>
<td>Assess transmissibility and its risk factors, and extent of infection</td>
<td>RSC data include a household key to identify any household spread, and we can identify people in residential homes. We are starting asymptomatic testing.</td>
<td>Partial delivery: RSC and UKHSA</td>
</tr>
<tr>
<td>Describe clinical presentation and risk factors for severe outcomes</td>
<td>Our ARI phenotype includes, as child concepts, most clinical presentations. We have specified key clinical data to collect. Links to hospital data provide severe outcomes.</td>
<td>Full delivery: RSC and UKHSA</td>
</tr>
<tr>
<td><strong>Domain II: Monitoring epidemiological characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitor characteristics of illnesses over time</td>
<td>Our surveillance of ILI, ARI, and SARI, applying the ARI phenotype, enables the ongoing monitoring of respiratory illnesses over time.</td>
<td>Full delivery: RSC and UKHSA</td>
</tr>
<tr>
<td>Monitor characteristics of circulating viruses</td>
<td>Our collaboration with the UKHSA in virology and serology sampling (including asymptomatic individuals) supports the monitoring of circulating viruses.</td>
<td>Full delivery: RSC and UKHSA</td>
</tr>
<tr>
<td>Monitor high-risk settings and vulnerable populations</td>
<td>Long-term investment in UK health computing and pay-for-performance means that primary care records capture risk groups. Other settings may be excluded.</td>
<td>Partial delivery: UKHSA from other settings</td>
</tr>
<tr>
<td>Monitor the impact on and coping abilities of health care systems</td>
<td>We can make year-on-year comparisons of data, running back over many years. However, there are no specific &quot;coping abilities.&quot;</td>
<td>Partial delivery: RSC and UKHSA</td>
</tr>
<tr>
<td><strong>Domain III: Informing use of interventions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitor the impact of nonmedical interventions</td>
<td>We have conducted epidemiological studies to explore the impact of nonmedical interventions during COVID-19 (eg, shielding).</td>
<td>Exemplar studies: RSC and UKHSA</td>
</tr>
<tr>
<td>Provide candidate vaccine viruses</td>
<td>We do not provide candidate vaccine viruses as part of surveillance.</td>
<td>Out of scope</td>
</tr>
<tr>
<td>Vaccine coverage, effectiveness, impact, and cost-effectiveness</td>
<td>Standardized national data indicate excellent coverage and impact. We have the capacity to supply data for vaccine effectiveness and cost-effectiveness studies.</td>
<td>Partial delivery: RSC and UKHSA</td>
</tr>
<tr>
<td>Monitor the effectiveness of antivirals and other therapeutics</td>
<td>We have conducted studies on the effectiveness of antivirals but have limited ability to assess new therapies owing to their central administration and data access issues.</td>
<td>Exemplar studies: RSC and UKHSA</td>
</tr>
<tr>
<td>Monitor the effectiveness of diagnostic tests</td>
<td>We have provided a comparison of results from POCT and UKHSA reference virology laboratories. This work could be scaled; see our additional objectives.</td>
<td>Exemplar studies: RSC and UKHSA</td>
</tr>
<tr>
<td>Monitor the effectiveness of clinical care pathways</td>
<td>We can monitor care pathways where we have access to data. Gaps include out-of-hours, NHS 111, and care homes. UKHSA syndromic surveillance fills these gaps.</td>
<td>Partial delivery: RSC and UKHSA</td>
</tr>
<tr>
<td>Monitor adverse events to vaccines and therapeutics</td>
<td>Ad hoc studies monitor adverse events of interest, either through data (passively) or by providing additional questionnaires. This is not a systematic part of surveillance.</td>
<td>Exemplar studies: RSC and UKHSA</td>
</tr>
</tbody>
</table>

WHO: World Health Organization.  
RSC: Research and Surveillance Centre.  
Each row is cumulative. Only new features are added in each row.  
ARI: acute respiratory infection.  
POCT: point-of-care testing.  
UKHSA: UK Health Security Agency.  
ILI: influenza-like illness.  
SARI: severe acute respiratory infection.
Table 2. Research and Surveillance Centre additional surveillance objectives that might be added to the World Health Organization mosaic framework.

<table>
<thead>
<tr>
<th>Additional RSC(^d) objectives</th>
<th>RSC delivery(^b)</th>
<th>Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPI(^c)</td>
<td>PPI is fundamentally important. We need the support of patients and the public for what we do. It is hard for us to fulfill our information governance responsibilities regarding informing patients and the public without strong PPI work.</td>
<td>Partial delivery: RSC and UKHSA(^d)</td>
</tr>
<tr>
<td>Monitor bacterial infections</td>
<td>Primary and secondary bacterial infections are significant contributors to the burden of respiratory disease. Informing hospitals about these pressures is challenging without considering bacterial diseases. Our innovations include (1) POCT(^e) for group A streptococcal infection and (2) urinary antigen tests to test for pneumococcus infection. We plan to improve the use of routinely collected bacteriology data.</td>
<td>Piloting by the RSC in collaboration with the UKHSA</td>
</tr>
<tr>
<td>POCT system-level evaluation</td>
<td>While we have evaluated POCT at the individual test level (mainly the sensitivity and specificity of specific virological tests), the RSC has the capability to evaluate the effectiveness of POCT at the system level. Outcomes could include: (1) antimicrobial stewardship, (2) use and effectiveness of antivirals, and (3) early intervention in high-risk individuals or populations.</td>
<td>Piloting by the RSC</td>
</tr>
<tr>
<td>Secured data governance</td>
<td>Our data governance portfolio, including the commissioning letter as described in the legal basis and governance framework for conducting sentinel surveillance, facilitates secure data use within the RSC.</td>
<td>Legal and national standards met by the RSC</td>
</tr>
<tr>
<td>Genomic surveillance resource</td>
<td>Our biomedical resource will link clinical data to virology sequencing data held in a publicly accessible resource. This will enable the linkage of clinical disease features and outcomes to the viral genomic sequence.</td>
<td>Partial delivery: RSC and UKHSA to complete this Wellcome Trust–funded project in 2024</td>
</tr>
</tbody>
</table>

\(^a\)RSC: Research and Surveillance Centre.  
\(^b\)Each row is cumulative. Only new features are added in each row.  
\(^c\)PPI: patient and public involvement.  
\(^d\)UKHSA: UK Health Security Agency.  
\(^e\)POCT: point-of-care testing.

Discussion

Principal Findings

The RSC and UKHSA are providing the most comprehensive primary care respiratory infection surveillance in the United Kingdom. We have extended the number of surveillance approaches for the coming season, meeting more of the areas proposed in the WHO mosaic framework (Table 1) [19].

In Domain I (detection and assessment of respiratory viruses), we have a robust system that has run over decades. However, we have scope to improve the rapidity and reliability of our results by increasing our sampling numbers and data quality so we can better detect changes.

In Domain II (epidemiological characteristics), we also have a robust system. There is potential to integrate work about asymptomatic infection, do more to include high-risk settings in our system, and develop indicators of our health care system’s ability to cope.

Domain III (informing the use of health interventions) is an area where we provide some key data, but this could be strengthened. We are able to monitor the impact of nonmedical interventions like social distancing [62,63], but other data are also needed to extend our capability of informing health interventions. We do report vaccine coverage, though data about vaccine exposure where vaccines are not given in primary care are more limited. We have primary care data and can link to other data to monitor the effectiveness of antivirals and other therapies, but access to centrally held data sets can be slow. We are well placed to report VE and adverse events of interest following vaccination [37,54] rapidly when using primary care data but with a greater lead time when we need to link to secondary care data. We can measure the costs of medically attended conditions. However, including measures of health-related quality of life would extend our ability to assess the disease burden.

Finally, we propose additional objectives that might be added (see Table 2). Among these, we consider that PPI and addressing the burden of primary and secondary bacterial infections are the most important.

Comparison With Prior Work

The strength of the RSC’s sentinel surveillance has long been established, but it has now been greatly extended, with the network more than doubling in size during the pandemic [66]. Other countries adapted their surveillance during the COVID-19 pandemic, including setting patient sampling routes outside of primary care (Sweden, Netherlands, and Scotland), decentralizing reference testing laboratories (France, Portugal, Scotland, and Spain), and optimizing digital data collection (Sweden, Netherlands, England, Scotland, France, Portugal, and Spain) [67]. Most of these changes were temporary, but the changes in the RSC network, such as the expanded sentinel practices and the introduction of electronic links to laboratories, have remained and will be further developed.

Australia, Belgium, the Netherlands, and the United States have shown that GPs would like to use more POCT to help them diagnose acute conditions [68]. Currently, the Welsh government has introduced a pharmacy-led service to undertake a structured
clinical assessment using clinical prediction scores and POCT for cases of suspected strep A infection [69].

**Strengths and Limitations**

The strengths of our network are its size (just under a third of the English national population), the level of sampling, and the commitment to improving data quality. The network has shown adaptability through the COVID-19 pandemic and a strong partnership working with the UKHSAs. We also have collaborations with other European sentinel networks and international collaborations [67]. The United Kingdom has a registration-based system that is free at the point of care, which allows good population coverage and facilitates the presentation of population-based infection rates. A unique national ID, the NHS number, ensures that primary care data can be linked to hospital and death data, allowing severe outcomes to be reported.

The limitations of routine data are that they are recorded by busy clinicians often working under pressure, and thus, they can be incomplete. Despite our best efforts, there can be gaps in data quality [70]. It is inevitable that our data will not capture all cases, and disease etiology might not be precise. While primary care data can be reliably reported within 3 days in arrears, it is much slower to gain access to secondary care data and other data sets. The national policy is to move toward a smaller number of secure data environments, and the RSC may need to migrate into one of these [21]. NHS primary care is increasingly working at scale, with the NHS setting up ARI hubs to work across geographical areas. We are exploring recruiting ARI hubs into our network.

**Conclusions**

The RSC has grown and adapted through the pandemic. Our biggest areas of change will be the introduction of an ARI phenotype, using technology to reduce the barriers to virology sampling and hopefully increase the scale and representativeness of virology sampling. The challenges in implementing change and the requests for more consistent data recording risk discouraging practices from remaining part of the RSC. Overall, our plans for the coming season will deliver more of the WHO surveillance mosaic.

**Acknowledgments**

We thank participating Research and Surveillance Centre (RSC) member practices and their patients for sharing pseudonymized data for disease surveillance, quality improvement, research, and education. We also thank computer system suppliers in English primary care: EMIS, TPP, In Practice Systems, and Wellbeing (Magentus). The RSC’s respiratory disease surveillance work is funded by the UK Health Security Agency. The Wellcome Biomedical resource is funded by the Wellcome Trust (212763/Z/18/Z).

**Conflicts of Interest**

SdL has received funding through his university from Astra-Zeneca, Eli-Lilly, GSK, MSD, Novo Nordisk, Sanofi, Seqirus, and Takeda, and has been a member of advisory boards for Astra-Zeneca, Sanofi, and Seqirus. He is the Director of the Oxford-Royal College of General Practitioner Research and Surveillance Centre. MZ is the chair of the charitable organization International Society for Influenza and other Respiratory Viruses (ISIRV) (not remunerated) and a member of the UK Government Scientific Advisory Groups Scientific Advisory Group for Emergencies (SAGE), New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG), and Joint Committee on Vaccination and Immunization (JCVI) (not remunerated). UH has received funding from Sanofi for vaccine-related workshops and has been a member of the advisory board for Janssen. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Patient and public involvement: Images from the first monthly newsletter.
[DOCX File, 488 KB - publichealth_v10i1e52047_app1.docx ]

Multimedia Appendix 2
Coding virology swab results.
[DOCX File, 15 KB - publichealth_v10i1e52047_app2.docx ]

Multimedia Appendix 3
Proposals for vaccine effectiveness studies for the coming season.
[DOCX File, 155 KB - publichealth_v10i1e52047_app3.docx ]

Multimedia Appendix 4
Distribution of virology and serology sampling practices and the entire network.
[DOCX File, 1323 KB - publichealth_v10i1e52047_app4.docx ]

Multimedia Appendix 5
Practice liaison team communications.
Multimedia Appendix 6
Summary of a pilot study to introduce pediatric phlebotomy training in the Research and Surveillance Centre network by region.

Multimedia Appendix 7
An introduction to LabLinks.

Multimedia Appendix 8
A protocol for a study using point-of-care testing.

References


24. People in Research. URL: https://www.peopleinresearch.org/ [accessed 2023-07-17]


59. Orchid Subject Privacy Notice. ORCHID. URL: https://orchid.phc.ox.ac.uk/privacy-policy [accessed 2023-07-20]


Abbreviations

- ARI: acute respiratory infection
- CMR: computerized medical record
- eFI: electronic frailty index
- FHIR: Fast Healthcare Interoperability Resources
- GISRS: Global Influenza Surveillance and Response System
- GP: general practitioner
- hMPV: human metapneumovirus
- ILI: influenza-like illness
- LRTI: lower respiratory tract infection
- NHS: National Health Service
- PhEMA: Phenotype Execution Modeling Architecture
- POCT: point-of-care testing
- PPI: patient and public involvement
- RCGP: Royal College of General Practitioners
- RSC: Research and Surveillance Centre
- RSV: respiratory syncytial virus
- SARI: severe acute respiratory infection
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms
UAG: urinary antigen
UKHSA: UK Health Security Agency
URTI: upper respiratory tract infection
VE: vaccine effectiveness
WHO: World Health Organization

Edited by A Mavragani, O Leal Neto; submitted 30.08.23; peer-reviewed by A Azcoaga-Lorenzo, L Kent; comments to author 19.12.23; revised version received 02.01.24; accepted 17.01.24; published 03.04.24.

Please cite as:
JMIR Public Health Surveill 2024;10:e52047
URL: https://publichealth.jmir.org/2024/1/e52047
doi:10.2196/52047
PMID: 38569175

©Xinchun Gu, Conall Watson, Utkarsh Agrawal, Heather Whitaker, William H. Elson, Sneha Anand, Ray Borrow, Anna Buckingham, Elizabeth Button, Lottie Curtis, Dominic Dunn, Alex J. Elliot, Filipa Ferreira, Rosalind Goudie, Uy Hoang, Katja Hoschler, Gavin Jamie, Debasish Kar, Beatrix Kele, Meredith Leston, Ezra Linley, Jack Macartney, Gemma L Marsden, Cecilia Okusi, Omid Parviz, Catherine Quinot, Praveen Sebastianpillai, Vanashree Sexton, Gillian Smith, Timea Suli, Nicholas P B Thomas, Catherine Thompson, Daniel Todkill, Rashmi Wimalaratna, Matthew Inada-Kim, Nick Andrews, Victoria Tzortziou-Brown, Rachel Byford, Maria Zambon, Jamie Lopez-Bernal, Simon de Lusignan. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 03.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Assessing Outcomes in HIV Prevention and Treatment Programs With Female Sex Workers and Men Who Have Sex With Men: Expanded Polling Booth Survey Protocol

Parinita Bhattacharjee¹,², MA; Leigh M McClarty¹, MSc, PhD; Joshua Kimani², MBChB, MPH; Shajy Isaac¹,³, PhD; Rhoda Wanjiru Kabuti², BSc; Antony Kinyua³, BA; Jaffred Karakaja Okoyana⁴; Virjinia Njeri Ndukuyu⁴; Helgar Musyoki⁵, MBChB, MPH; Anthony Kiplagat⁶, BSc; Peter Arimi⁵, MBChB, MSc; Souradet Shaw¹, MSc, PhD; Faran Emmanuel¹, MBBS, MSc; Monica Gandhi⁷, MPH, MD; Marissa Becker¹, MSc, MD; James Blanchard¹, MD, PhD

¹Institute for Global Public Health, University of Manitoba, Winnipeg, MB, Canada
²Partners for Health and Development in Africa, Nairobi, Kenya
³India Health Action Trust, Delhi, India
⁴Sex Workers Outreach Program Clinic, Nairobi, Kenya
⁵The Global Fund to Fight AIDS, Tuberculosis and Malaria, Nairobi, Kenya
⁶Nairobi County Health Department, Nairobi, Kenya
⁷School of Medicine, University of California at San Francisco, San Francisco, CA, United States

Corresponding Author:
Leigh M McClarty, MSc, PhD
Institute for Global Public Health
University of Manitoba
R065 Medical Rehabilitation Building
771 McDermot Ave
Winnipeg, MB, R3E 0T6
Canada
Phone: 1 2047893245
Email: leigh.mcclarty@umanitoba.ca

Abstract

Background: Assessing HIV outcomes in key population prevention programs is a crucial component of the program cycle, as it facilitates improved planning and monitoring of anticipated results. The Joint United Nations Programme on HIV and AIDS recommends using simple, rapid methods to routinely measure granular and differentiated program outcomes for key populations. Following a program science approach, Partners for Health and Development in Africa, in partnership with the Nairobi County Government and the University of Manitoba, aims to conduct an outcome assessment using a novel, expanded polling booth survey (ePBS) method with female sex workers and men who have sex with men in Nairobi County, Kenya.

Objective: This study aims to (1) estimate the incidence and prevalence of HIV; (2) assess biomedical, behavioral, and structural outcomes; and (3) understand barriers contributing to gaps in access and use of available prevention and treatment services among female sex workers and men who have sex with men in Nairobi.

Methods: The novel ePBS approach employs complementary data collection methods, expanding upon the traditional polling booth survey (PBS) method by incorporating additional quantitative, qualitative, and biological data collection components and an improved sampling methodology. Quantitative methods will include (1) PBS, a group interview method in which individuals provide responses through a ballot box in an unlinked and anonymous way, and (2) a behavioral and biological survey (BBS), including a face-to-face individual interview and collection of linked biological samples. Qualitative methods will include focus group discussions. The ePBS study uses a 2-stage, population- and location-based random sampling approach involving the random selection of locations from which random participants are selected at a predetermined time on a randomly selected day. PBS data will be analyzed at the group level, and BBS data will be analyzed at an individual level. Qualitative data will be analyzed thematically.
Results: Data were collected from April to May 2023. The study has enrolled 759 female sex workers (response rate: 759/769, 98.6%) and 398 men who have sex with men (response rate: 398/420, 94.7%). Data cleaning and analyses are ongoing, with a focus on assessing gaps in program coverage and inequities in program outcomes.

Conclusions: The study will generate valuable HIV outcome data to inform program improvement and policy development for Nairobi County’s key population HIV prevention program. This study served as a pilot for the novel ePBS method, which combines PBS, BBS, and focus group discussions to enhance its programmatic utility. The ePBS method holds the potential to fill an acknowledged gap for a rapid, low-cost, and simple method to routinely measure HIV outcomes within programs and inform incremental program improvements through embedded learning processes.

KEYWORDS
female sex workers; men who have sex with men; Kenya; polling booth survey; program science; HIV prevention; outcome assessment

Introduction

Background

Kenya has made significant strides in its HIV response, evident by the continuous decline in HIV prevalence among adults aged 15 to 49 years within the general population. HIV prevalence has dropped from a peak of approximately 10% in the mid-1990s to 4.3% in 2021 [1]. An analysis of the HIV epidemic within Kenya [2] highlights substantial heterogeneity. HIV prevalence varies geographically, ranging from 20.1% in Homa Bay County to 0.2% in Mandera and Wajir counties. Notably, HIV prevalence in Kenya is also disproportionally high among female individuals (5.5%) than among male individuals (2.9%) [3]. Furthermore, compared to the general Kenyan population, HIV prevalence is significantly higher among key populations, namely, female sex workers at 29%, men who have sex with men at 18.9%, people who inject drugs at 18%, and transgender people at 40% [4,5]. This underscores the critical importance of properly scaled and tailored interventions to address the HIV epidemic within key populations in the country [4].

In 2020, a nationwide mapping and population size estimation exercise estimated that there were 197,096 female sex workers and 61,650 men who have sex with men living in Kenya (National AIDSSTI Control Programme, unpublished data, December 2020). Approximately 20% of key populations are concentrated in Nairobi County, the largest of the 47 counties in Kenya [6]. Findings from a 2017 national study underscore the high prevalence of HIV among female sex workers and men who have sex with men in Nairobi, with self-reported rates of 22% and 27%, respectively [7]. In the same study, 91% of female sex workers in Nairobi reported using condoms with their last client, while 81% of men who have sex with men reported condom use during last anal sex [7]. Furthermore, peer educators played a significant role in outreach efforts, with 88% of female sex workers and 76% of men who have sex with men reporting contact with a peer educator within the last 3 months. The majority (82%) of both groups had undergone HIV testing within the last 3 months, while 22% of female sex workers and 27% of men who have sex with men reported that they were living with HIV. However, only 65% of female sex workers and 68% of men who have sex with men living with HIV reported being enrolled in antiretroviral therapy programs. Notably, 56% of female sex workers and 18% of men who have sex with men reported experiencing incidents of police violence within the last 6 months, respectively [7].

HIV prevention programs for key populations in Nairobi County are implemented by several partner organizations, all of whom report to the county government and the National Key Population Program, led by the Kenyan Ministry of Health, on a monthly basis (National AIDS and STI Control Programme, unpublished data, September 2023). Having up-to-date, representative, and robust data on key populations is of paramount importance for effective planning, management, and monitoring of programs and interventions aimed at addressing their needs [8]. The most recent population-based behavioral survey conducted among key populations in Nairobi dates back to 2017 [7] and did not include biological data. The absence of recent biobehavioral data among key population groups limits the ability of both county and national programs to understand the current epidemiologic situation, assess the impact of investments in HIV prevention and treatment programs for key populations, and identify gaps in program coverage. The high cost and resource requirements for implementing nationally representative, population-based integrated biobehavioral surveillance surveys factor into the relative infrequency of their deployment [8]. The availability of methods that can efficiently collect timely data among key population groups at appropriate scale and relatively low cost across diverse contexts is extremely limited [9]. Striking a balance between the need for methodological rigor and practical, adaptable approaches that are accessible to countries with diverse income levels is a pressing requirement [10].

Objectives

Partners for Health and Development in Africa (PHDA), in collaboration with the University of Manitoba and the Nairobi County Government, aims to conduct an outcome assessment study using a novel, expanded polling booth survey (ePBS) method for data collection. This outcome assessment focuses on 2 key populations with whom PHDA works in Nairobi County, Kenya: female sex workers and men who have sex with men. The objectives of the study are to (1) estimate the incidence and prevalence of HIV; (2) assess biomedical (HIV testing, condom use, pre-exposure prophylaxis [PrEP] uptake, and HIV treatment), behavioral (knowledge and risk behavior), and structural outcomes (experiences of violence, stigma, and discrimination); and (3) gain a deeper understanding of barriers
that contribute to disparities in accessing and using HIV services among female sex workers and men who have sex with men in Nairobi County, Kenya. Importantly, this study will pilot the feasibility of the novel ePBS method to measure outcomes in key population programs, with the intention of disseminating the approach for scaling up in other key population programs within and outside Kenya.

Methods

Study Setting
Nairobi County is the largest of Kenya’s 47 counties by population and has a robust key population program. According to the latest mapping and population size estimation exercise, Nairobi County has the highest concentration of female sex workers (n=39,227) and men who have sex with men (n=15,271) living or working in the country, or both [6]. The county has identified 2032 locations, across its 12 subcounties, where female sex workers solicit clients and 369 locations where men who have sex with men meet their sexual partners [6]. Bars (without lodging) have the highest proportion of female sex workers and men who have sex with men among these locations [6]. Importantly, same-sex sexual practices are criminalized throughout Kenya, and in 2017, Nairobi County implemented a ban on all forms of sex work [11,12]. This has contributed to the elevated levels of violence, stigma, and discrimination experienced by female sex workers and men who have sex with men [7] within the county. There are currently 4 large implementing partners responsible for prevention programs for female sex workers and men who have sex with men in Nairobi, which are funded by the Global Fund to Fight AIDS, Tuberculosis, and Malaria and the US President’s Emergency Plan for AIDS Relief. Among these partners, PHDA implements the largest prevention program in Nairobi and Kenya, serving approximately 34,000 female sex workers and 14,000 men who have sex with men. PHDA reaches these groups across all 17 counties in Kenya through an extensive network of peer educators and 9 Sex Workers Outreach Program (SWOP) clinics. To implement this program, PHDA receives funding from the US President’s Emergency Plan for AIDS Relief and technical support from the University of Manitoba and the University of Maryland, Baltimore. Furthermore, this collaboration involves Nairobi County and the National AIDS and STI Control Programme (NASCOP) [13].

Program Science Approach
This study adopts a program science approach, which systematically applies scientific theory and empirical knowledge to answer critical programmatic questions that are generated by the program itself [14]. The approach seamlessly integrates the program cycle with research questions, strategically embedding iterative cycles of research and learning within a program, thus facilitating rapid uptake of knowledge and evidence into program operations. Furthermore, the program science approach generates embedded research agendas to address ongoing programmatic challenges [15]. In the context of this study, the research questions and study objectives have evolved from 2 primary programmatic concerns within the Nairobi County’s key population program. The first concern revolves around the absence of recent population-based data to assess program outcomes (biomedical, behavioral, and structural) pertinent to key populations in Nairobi County. The second concern is the lack of a low-cost, reliable, and easily reproducible method capable of routinely evaluating granular and differentiated program outcomes for key populations, assessing and identifying gaps in the program’s coverage, and making programmatic adjustments to address emerging challenges.

Program science emphasizes the importance of routinely identifying, quantifying, examining, and addressing gaps in program coverage, through tools, such as the effective program coverage framework, to reduce inequities in health outcomes and achieve population-level impact [16]. A program science approach underscores the importance of engaging affected communities [17] and actively works toward developing program and policy solutions to public health challenges [18]. As such, this study is conducted by PHDA in partnership with the community advisory board of the SWOP clinics and the Department of Public Health, Nairobi County.

Research Questions
The study sets out to answer four specific research questions.

1. What is the HIV incidence and prevalence among female sex workers and men who have sex with men in Nairobi, Kenya?
2. Has the program successfully met its targets related to biomedical, behavioral, and structural outcomes among female sex workers and men who have sex with men in Nairobi, Kenya?
3. What is the estimated program coverage of the prevention services among female sex workers and men who have sex with men in Nairobi, Kenya?
4. What are the key barriers experienced by female sex workers and men who have sex with men when accessing and using prevention services in Nairobi, Kenya?

Outcomes for Measurement
In total, 12 outcomes will be measured to address the study’s research questions (Textbox 1). Specifically, to address research questions 1 and 2, five biomedical outcomes, 2 behavioral outcomes, and 2 structural outcomes will be measured. For research questions 3 and 4, program coverage, availability coverage, contact coverage, and utilization coverage will be measured among both key population groups, following the definitions outlined in the program science–informed effective program coverage framework [16]. HIV prevalence and incidence at the population level will also be measured.
Measured HIV outcomes among female sex workers and men who have sex with men using an expanded polling booth survey in Nairobi County, Kenya, from April to May 2023.

### Themes and outcomes

#### Biomedical
- Prevalence of sexually transmitted infections (STIs) symptoms, diagnosis, and treatment
- Consistent condom use
- Routine HIV testing (and receipt of results)
- Linkage to and retention in HIV treatment and care
- HIV viral load suppression

#### Behavioral
- Self-reported HIV risk behaviors
- Knowledge of HIV

#### Structural
- Prevalence of gender-based violence
- Prevalence of stigma and discrimination

#### Program coverage
- Availability coverage
- Contact coverage
- Utilization coverage

### Study Timeline

The ePBS study was designed with a quick turnaround; all activities are expected to be completed within 9 months. The data collection phase was planned to span 30 consecutive days, followed by data cleaning and analyses. Results will be finalized for dissemination at the end of the 9-month period.

### Study Team

#### Composition

A total of 4 study teams were established to manage the implementation of ePBS, and each team concurrently collected data in 4 different locations daily. Each team consisted of 6 members including a prevention officer, a key population outreach worker (a member of either key population community), a clinical officer, an HIV and testing service (HTS) counselor, and 2 qualitative researchers (1 focus group discussion [FGD] facilitator and 1 notetaker). Peer educators working in the SWOP clinics supported the study teams in participant enrollment. All ePBS team members, except the qualitative researchers, were enrolled from PHDA-SWOP clinic staff. All 4 teams were supervised by senior members (site supervisors) of the PHDA-SWOP clinic team and representatives from the Nairobi County government.

#### Study Team Training

The study team underwent a comprehensive 5-day training. The first 3 days of training included classroom instruction, during which the study’s investigators introduced and facilitated training on (1) the conceptual and theoretical underpinnings of the study (ie, the program science approach and the effective program coverage framework) [16]; (2) the ePBS methodology; (3) data collection methods and tools; (4) logistics, implementation flow, and study procedures; and (5) data entry processes. Subsequently, 2 days were dedicated to field practice during which mock data collection exercises were conducted to enhance the skills of the study team and to provide opportunities for troubleshooting any anticipated challenges in the field. Importantly, all 5 days of ePBS training prioritized practical exercises to ensure that the study team felt confident and proficient, thereby ensuring consistency and quality in data collection.

### ePBS Method

The ePBS method is a rapid, cross-sectional study design using complementary quantitative and qualitative methods for data collection. The innovative approach expands upon the traditional polling booth survey (PBS) method [19] by incorporating additional qualitative, quantitative, and biological data collection methods and an improved population-based approach to random sampling. Within the ePBS method, quantitative methods include (1) PBS, a group interview method in which individuals provide anonymous and unlinked responses to survey questions through a ballot, which has been shown to reduce reporting and social desirability biases for sensitive questions about conventionally stigmatizing behaviors, experiences, and social circumstances [19,20] and (2) behavioral and biological survey (BBS), which includes a short face-to-face, individual interview linked to the collection of biological samples (blood and urine). Qualitative methods include FGD with a subset of ePBS participants to inquire about factors that contribute to gaps in
program coverage. Hence, the expansion of the traditional PBS by adding a BBS and FGD is referred to as ePBS.

Cross-sectional designs are commonly used for population-based surveys and to assess the prevalence of diseases in a population [21]. They are also useful for informing planning, monitoring, and evaluation of public health programs and interventions. In this study, we use quantitative methods to measure prevention program outputs and relevant outcomes among female sex workers and men who have sex with men participants, as outlined in research questions 1 to 3. We use qualitative methods to gain insight into the barriers affecting the accessibility and use of HIV services, as specified in research question 4.

**PBS Component**

**Overview**

The PBS method is a well-established group survey method comprising only binary questions, in which individuals provide responses to survey questions by placing a token corresponding with their answer into a ballot box [22]. Individual responses remain anonymous and unlinked, ensuring confidentiality. This anonymity has been shown to enhance participants’ trust in the confidentiality of the data collection process, thereby minimizing social desirability biases and lending itself to more accurate reporting on sensitive behavioral information [23]. Data generated through PBS have been found to be comparable to the findings from other studies that used different methodologies with the same populations [20].

**Sampling Procedure**

Potential participants are selected using a 2-stage, randomized, probability-based sampling procedure and organized into small homogenous groups by key population groups (ie, female sex workers and men who have sex with men) and typology of the location at which sex work and cruising take place (eg, bar, bar with lodges, street, and clubs). Each PBS session consists of 10 to 12 participants of the same key population group who congregate and solicit clients or cruise in the same locations. Given the group interview format, the number of survey questions is intentionally limited, and each question is short and simple to facilitate easy response [22]. The key population program in Kenya has previously used PBS, and participants from different key population groups have found it to be an acceptable approach [7].

**Implementation Procedure**

The PBS component of the ePBS study was led by prevention officers and outreach workers. Conducting PBS requires materials that are not typically considered for other survey methods. First, each participant in a PBS session must have 3 different colored ballot boxes associated with yes, no, and not applicable responses (typically green, red, and white, respectively) and a set of cards to use as tokens to be dropped into the ballot boxes. Cards should be clearly numbered in alignment with the number of survey questions. Each participant’s station also requires a polling booth carton to create a private space that is not visible to other participants in the PBS session (often a cardboard box with 1 open face pointed at the participant). The facilitator at each PBS session should be equipped with copies of the standardized data collection tool (survey) and reporting forms to tally and record responses to each question.

Peer educators from SWOP mobilized 10 to 12 potential participants and invited them to participate in a PBS session, each of which was conducted as follows:

- Participants invited to the PBS were provided with an individual polling booth carton at the study venue. The polling booths were separated by at least 1 m to ensure privacy and confidentiality for each participant and their responses.
- The facilitator welcomed the participants and explained the objective of the study, read out the consent format, provided time and space for questions, and sought informed consent from every participant.
- Each participant received 3 ballot boxes (green, red, and white) and a pack of cards, with each card numbered and stacked in a sequential order to correspond with the number of questions in the survey.
- The facilitator confirmed that each participant had the correct number of cards arranged in the proper order before proceeding with the questions.
- The facilitator provided a full explanation of the PBS procedure, including an example question and a practice response, to assure participants that their responses will remain anonymous and unlinked.
- In terms of responses, the facilitator explained the following:
  - If the response to the question is yes, the participant should put the card with the corresponding question number into the green box.
  - If the response to the question is no, the participant should put the card with the corresponding question number into the red box.
  - If the question does not apply to the participant, the participant should put the card with the corresponding question number into the white box.
  - If the person does not want to answer the question, the card with the corresponding question number should be kept outside the provided boxes.

- The facilitator read the survey questions, one at a time, making the exercise engaging and lively. They ensured that each question was clearly understood by all participants by (1) reading each question clearly, slowly, and loudly to ensure that each participant heard the question clearly; (2) reading out the questions in a language that is easily understood by participants; (3) repeating the question, if necessary; (4) using local terms and giving sufficient pause, taking care not to rush through the questions.
- After all the PBS questions had been administered, the facilitator (1) collected the cards separately for each of the boxes: green, red, and white and (2) tallied and recorded the number of responses from each card in each of the colored boxes into a standard reporting form.
- All data generated through the PBS process were entered into a tablet by the study team using a standardized database created specifically for the ePBS. The data collection was facilitated using SurveyCTO (Dobility, Inc), a secure and
scalable mobile data collection platform designed for researchers working in offline settings.

**BBS Component**

**Overview**
The BBS is an expansion to the standard PBS method and was led by the study teams’ clinical officers and HTS counselors.

**Table 1.** Biological samples collected from female sex workers and men who have sex with men during the biobehavioral survey in Nairobi, Kenya, from April to May 2023.

<table>
<thead>
<tr>
<th>Biomarkers</th>
<th>Sample type</th>
<th>Type of test</th>
<th>Test location</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV diagnosis</td>
<td>Capillary blood</td>
<td>Rapid test</td>
<td>Study site</td>
</tr>
<tr>
<td>Recency of HIV infection</td>
<td>Plasma</td>
<td>Laboratory-based test</td>
<td>PHDA laboratory</td>
</tr>
<tr>
<td>HIV-1 RNA viral load</td>
<td>Plasma</td>
<td>Laboratory-based test</td>
<td>PHDA laboratory</td>
</tr>
<tr>
<td>Actively taking PrEP&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Urine</td>
<td>Rapid test</td>
<td>Study site or in the PHDA laboratory</td>
</tr>
</tbody>
</table>

<sup>a</sup>PHDA: Partners for Health and Development in Africa.

<sup>b</sup>PrEP: pre-exposure prophylaxis.

**Individual, Face-to-Face Structured Interviews**
Following the PBS, all participants who provided informed consent participated in a brief, face-to-face, structured interview with a clinical officer to gather limited demographic information, basic behavioral information related to HIV risk, and program access and service use data. These data will be linked, at the individual level, to biological samples (blood and urine), which were collected following the face-to-face interviews.

**Biological Samples**

**Urine Sample**
To protect the confidentiality of participants’ HIV status, urine samples were collected from all participants who provided informed consent to providing biological samples. However, only samples from those participants who had reported currently taking PrEP in the preceding face-to-face interview underwent testing using a rapid tenofovir assay developed and validated by the University of California, San Francisco in conjunction with Abbott Laboratories [24]. Depending on timing and the facilities available on-site, the samples were either tested for tenofovir at the study site or sent to the PHDA laboratory for processing the following day (Table 1).

**Blood Samples**
HTS counselors collected fingerprick blood samples from all participants who provided informed consent for on-site rapid HIV testing. Following Kenyan rapid testing algorithms, Determine HIV-1 or HIV-2 (Abbott Diagnostic Medical Co Ltd) was used as the initial screening assay. All reactive samples underwent confirmatory testing using First Response (Premier Medical Corp Ltd) rapid tests. Pre- and posttest counseling sessions were administered to all participants undergoing HIV rapid testing in accordance with the national guidelines [25].

In addition, clinical officers drew 5 mL venous blood samples from consenting participants who tested positive for HIV in the rapid test. Whole blood samples were transferred to the PHDA laboratory for further analyses, including HIV viral load and recency testing (Table 1). The HIV viral load analyses will be conducted using GeneXpert HIV-1 Viral Load Kits (Cepheid). HIV recency will be determined using the Asante HIV-1 Rapid Recency Assay (Sedia Biosciences Corporation). All laboratory-based tests will be carried out in accordance with the PHDA laboratory’s standard operating procedures and in adherence to the national guidelines [26].

**Biological Sample Management**
All specimens, both urine and blood, were affixed with preprinted standard codes along with the corresponding survey identification at the time of collection. Following collection, all biological samples were carefully placed in boxes at room temperature. These samples were transported to the PHDA laboratory using PHDA vehicles for processing and subsequent testing. The transfer of samples from study sites to the PHDA laboratory followed a protocol outlined by the PHDA laboratory’s standard operating procedures. Both urine and blood samples will be maintained at a temperature of −80°C for 5 years within the PHDA laboratory per standard operating procedures. This extended storage period ensures the availability of samples for the confirmation of any discrepant laboratory results that may arise over time. Stored specimens will not be linked to any personal identifiers and cannot be traced back to the individuals who provided them, preserving the confidentiality and privacy of the participants.

**FGD Component**
The FGD component of the ePBS study was led by the study teams’ qualitative researchers. For each key population group, participants were selected from every fifth PBS session and were engaged in the FGD after providing informed consent. Each FGD consisted of the same 10 to 12 participants who participated in the preceding PBS session. The FGDs were facilitated by trained qualitative researchers using an FGD guide. The discussions were conducted in either Kiswahili or Sheng, which is a local slang language predominantly spoken in urban areas of Kenya, particularly Nairobi. All FGDs were audio...
recorded and will be translated and transcribed verbatim for analysis.

**Participant Inclusion Criteria**

To be eligible for participation in this ePBS study, female sex workers had to be assigned female at birth, aged at least 18 years, and acknowledge having received money or gifts in exchange for sexual intercourse with an individual assigned male at birth at least once in the past 3 months. Eligible men who have sex with men had to be assigned male at birth, aged at least 18 years, and report at least 1 anal sex act (insertive or receptive) with another individual assigned male at birth in the past 3 months. To be eligible, all participants had to be capable and willing to provide written or verbal informed consent to participate in the 3 components of the ePBS study, self-identify as a sex worker or a man who has sex with other men, and report actively practicing sex work or cruising within Nairobi County.

**Target Sample Size Calculation**

To be representative of each key population group, the study’s sampling strategy was based on probability sampling techniques that provided an equal chance to all members of the study populations, meeting defined inclusion criteria, to be included in the study. The determination of an appropriate sample size for a single study domain is usually based on the following considerations: (1) the number of measurement units in the population, (2) the parameter of interest, and (3) the degree of confidence by which the parameter is estimated and the desired level of precision.

One objective of our study was to estimate the HIV prevalence and incidence among female sex workers and men who have sex with men in Nairobi County. As such, sampling parameters were set with the prevalence and sample size having 95% statistical confidence, with a desired precision of ±5% or –5%.

The sample size for PBS and BBS study was calculated using the following formula: 

\[ n = \frac{1.96^2 p(1-p)(DEFF))}{d^2} \]

Where \( p \) is HIV prevalence, \( DEFF \) is the design effect of the sampling approach, and \( d \) is the desired precision.

The finite population correction factor is then applied to the derived sample size using the following formula: 

\[ nf = \frac{n \times N}{(n + (n - 1))} \]

Where \( nf \) is the sample size adjusted for finite population correction, \( n \) is the sample size required, and \( N \) is the population size.

To calculate the sample size for this ePBS study, we used the most recent estimate of HIV prevalence among female sex workers (26%) and men who have sex with men (12%) in Nairobi, as reported in a 2017 study conducted by NASCOP [7]. The sample size calculation was adjusted to accommodate an anticipated nonresponse rate of 15%. The final target sample size, adjusted for population size and nonresponse, for female sex workers was \( n = 769 \) and for men who have sex with men was \( n = 420 \). Ideally, each PBS session comprises 10 to 12 participants; the total target sample sizes for female sex workers and men who have sex with men were divided by 12 to determine the total number of PBS sessions required for each key population group, that is, 64 PBS sessions with female sex workers and 35 sessions with men who have sex with men.

For each key population group, participants in every fifth PBS session were invited to participate in an FGD. In total, 20 FGDs were conducted with ePBS participants (female sex workers: \( n = 13 \), 65%; men who have sex with men: \( n = 7 \), 35%).

**Developing a Sampling Frame**

**Overview**

The study used a 2-stage, random sampling approach. In the first stage, random locations (ie, places where female sex workers solicit clients or engage in sex and where men who have sex with men cruise or engage in sex) were selected from a previously validated list. In the second stage, participants were randomly selected from the locations on a randomly selected day of the week.

**Stage 1 Sampling**

As a first step, the ePBS study team developed the sampling frame by validating lists of mapped locations in Nairobi County where female sex workers and men who have sex with men solicit clients, cruise, meet sexual partners, and engage in sexual activities. The sample frame was derived from a programmatic mapping exercise conducted in 2018 [6]. The lists were updated—that is, closed locations were removed and newly emerged locations were added—through a validation process facilitated by the study team members who are in regular, direct contact with key population members. Through this validation process, population size estimates for each location, the days and times when these locations are operational, and the days when the locations have the highest number of female sex workers or men who have sex with men were verified and revised, as needed, via a standardized approach.

The lists of validated locations (1 each for female sex workers and men who have sex with men) formed the sampling frames representing the universe of locations across Nairobi County. Key population members spending time within a given location tend to be more homogeneous, in terms of typology and risk behavior, than those in different locations [27]. As such, potential participants for each PBS session were randomly selected from a single location. During the first stage of sampling, 99 locations (equal to the number of PBS sessions planned for each group) were randomly selected such that 1 PBS session was conducted per selected location. The random selection of key population group–specific locations from which potential participants were identified was stratified by subcounty and location typology.

**Stage 2 Sampling**

In stage 2 of sampling, each of the 99 randomly selected locations was randomly allocated to an assigned week and day on which potential participants were enrolled from the location. For example, data collection for this ePBS study was projected to take 30 days; for each location, 1 week (ie, from weeks 1 to 4 of the study period) was randomly selected first, followed by the random selection of 1 specific day within that chosen week. This process determined the precise week and day that each selected location was designated for the study. Ultimately, 10 to 12 potential participants were randomly approached by...
community researchers at the location on the specified day and week and invited to take part in the study.

This 2-stage sampling approach ensured that participants were representative of the total population of the key population group by geography (subcounty) and location typology across days of the week, yielding a comprehensive and diverse sample of participants from communities of female sex workers and men who have sex with men in Nairobi County.

Study Procedure

The overall flow of the ePBS study procedure from participant identification and enrollment to PBS, BBS, and FGD (every fifth PBS session) is outlined in Figure 1. The duration to complete PBS and BBS was 120 minutes, including 10 minutes for initial screening and consent, 60 minutes for the PBS, 20 minutes for pretest counseling, and 30 minutes for BBS. For every fifth PBS group participating in an FGD, an additional 45 to 60 minutes was added to their total estimated participation time.

Figure 1. Expanded PBS study procedure flowchart used to collect data from female sex workers and men who have sex with men in Nairobi, Kenya, from April to May 2023. BBS: behavioral and biological survey; FGD: focus group discussion; PBS: polling booth survey; PrEP: pre-exposure prophylaxis; SWOP: Sex Workers Outreach Program.

Participant Enrollment and Data Collection Site Preparation

Potential participant enrollment occurred at selected locations on the allocated days of the week, according to the study’s sampling frame. Enrollment primarily occurred in the late afternoon or evening on the day of data collection to increase the likelihood of meeting a higher number of female sex workers or men who have sex with men in the location. Peer educators and outreach workers from the SWOP clinics facilitated the identification of potential participants within their own key population group. They were responsible for introducing potential participants to the study and assessing their eligibility. Any eligible individuals were provided with detailed information about ePBS. If eligible participants were interested in participating in the study, they were brought to the data collection site by peer educators and outreach workers. As much as possible, data collection occurred within the selected locations. Ahead of participant enrollment, the study team members scouted selected sites to identify suitable spaces for safe and confidential data collection. Team members were responsible for setting up stations for PBS and private, quiet locations for face-to-face structured interviews and biological sample collection. If the selected sites posed challenges (e.g., logistical or safety challenges) for data collection, prescouted alternate locations nearby were used instead.

Ethical Considerations

Informed Consent

At the data collection site, before ePBS data collection commenced, prevention officers, with the support of community researchers, guided participants through the informed consent process, ensuring that all participants understood and were comfortable with the information included in the consent form. Any questions from potential participants were addressed before commencing the study. Nonconsenting or unwilling participants were thanked for their time and excluded from the study. All eligible participants who provided written or verbal informed consent received a unique identifying code that was used for the BBS portion of the study.

Participant Compensation

Participants received compensation for their time and travel at a standard local rate of Ksh 500 (US $5). Those participating
in FGDs received an additional Ksh 500 (US $5) for their extra time.

**Safety**

Conducting research studies with key populations who are criminalized in Kenya requires careful consideration to ensure community safety. Before initiating data collection, extensive consultations with key population communities were conducted. The primary objectives of the study and the methodology were shared and explained during these consultations. Feedback obtained from the community was instrumental in shaping the study procedures to ensure the communities’ safety and comfort. Furthermore, a collaborative partnership with Nairobi County authorities was established to enhance the safety and protection of both the study teams and the key populations and participants. The study teams carried official identity cards and study approval letters to show that they were conducting an official study. The data collection took place in locations where female sex workers and men who have sex with men generally congregate and that were considered to be safe. Location managers were sensitized and informed about the study ahead of implementation. Study team members were trained in crisis management and were provided a protocol to follow in case of an emergency or a breach in safety for the study team or participants. A team of supervisors, including senior team members of PHDA and Nairobi County officials, visited the sites each day for supervision and support.

**Ethics Approvals**

The research study has received ethics approvals from Amref Health Africa Ethics and Scientific Review Committee, Kenya (Amref ESRC P1365/2022); Health Research Ethics Board, University of Manitoba, Canada (HS25883); National Commission for Science, Technology, and Innovation, Kenya (NACOSTI/P/23/24009); and Nairobi City County, County Health Research Ethics Committee (NCCG/HWN/REC/349).

**Privacy and Confidentiality**

All investigators have completed a web-based course on the protection of human participants in research and have obtained certificates. Furthermore, investigators and all study team members have signed data confidentiality agreements with PHDA, pledging to uphold the highest standards of data privacy and security.

During the informed consent process, female sex workers and men who have sex with men participating in the study were clearly informed of their right to withdraw from the study at any point if they so desired. This withdrawal did not affect their access to the services provided by SWOP clinics. The potential benefits and risks associated with participating in the study were thoroughly explained to the participants during the informed consent process.

**Study Instruments**

Standardized instruments adapted for female sex workers and men who have sex with men were used for the PBS and BBS. The FGD guide was standardized for all populations. All tools were available in English and Kiswahili.

**Data Entry and Management**

Each study team had 2 tablets that were preloaded with customized web-based survey software (SurveyCTO) and were used for data entry and management. Data collected during PBS sessions were first compiled in hard copy format using standardized reporting forms that were filled out by the prevention officer and community researchers after PBS responses were tallied. Reporting form tallies were then entered into tablets by the prevention officers on each study team. BBS data were directly entered into tablets using SurveyCTO. Data quality checks were automated within the web-based system. Data were uploaded securely to a central server at the end of each field day via a virtual private network (VPN).

Select biological samples collected were sent to the PHDA laboratory (Table 1). These laboratory data were managed using the Laboratory Information Management Software (LMIS), uploaded to a repository in CSV format, and merged with data collected through SurveyCTO by the study’s data manager.

The study team’s qualitative researchers were responsible for digitally audio recording FGDs and transcribing and translating (into English) the FGD data.

**Data Security**

Data were transmitted over an encrypted channel (HTTP secure), and personal identifiers were automatically encrypted at data entry. Access to central servers is restricted. Data will be stored in encrypted format for 5 years, following national data security protocols.

**Data Cleaning**

Daily data quality checks will be performed on PBS and BBS data collected through SurveyCTO, addressing discrepancies, missing data, and inconsistencies as they are identified. The study teams’ prevention officers and site supervisors will oversee the data cleaning and quality assurance procedures.

**Data Analysis**

Data from PBS and BBS will be analyzed at the group and individual levels, respectively. Descriptive statistics will be calculated, and univariate analyses will be conducted to measure previously specified outcomes (Textbox 1) and assess risk factors associated with HIV. Guided by the coverage cascade element of the effective program coverage framework [16], coverage gap analyses will be performed to identify and quantify gaps in availability coverage, contact coverage, and use coverage in relation to required coverage targets. Thematic analysis will be done for FGD data to capture recurring concepts and ideas related to barriers and facilitators to accessing, contacting, and using prevention services that emerged during discussions.

**Knowledge Translation and Dissemination**

The study’s findings and lessons on the feasibility and utility of ePBS as a data collection method will be shared through participation in the Nairobi County Key Population Technical Working Group and the National Key Population Technical Working Group led by the NASCOP. This working group comprises members representing implementing partners, key...
population community representatives, donors, and the Kenyan Ministry of Health officials. These stakeholders will have the opportunity to use the study’s findings to inform and strengthen programs and policies for key populations in Nairobi. In addition, the study results will be shared with the PHDA-SWOP Community Advisory Board and peer educators. These insights will support the development of strategies to enhance program effectiveness and improvement.

Research findings and methodological insights from implementing the novel ePBS study will be shared with academic communities in Kenya and globally through international conferences and peer-reviewed journal articles. PHDA is a part of an extensive, existing network of researchers in public health sciences, health policy, and basic and natural sciences who focus on HIV and sexually transmitted infections. Other scientific forums within the University of Manitoba will be leveraged to share the findings.

Results

The study enrolled 759 female sex workers (response rate: 759/769, 98.6%) and 398 men who have sex with men (response rate: 398/420, 94.7%) across 64 and 35 PBS sessions, respectively (Table 2).

In total, 758 (99.9%) female sex workers (1 participant refused participation in BBS) and 398 (100%) men who have sex with men participated in the face-to-face interview for the BBS, and all (n=1156) provided blood and urine samples for rapid tests. In total, 14.79% (171/1156) of participants (female sex workers: n=101, 59.1%; men who have sex with men: n=70, 40.9%) tested positive for HIV via rapid test. Among those testing positive, 93.6% (160/171) of participants (female sex workers: n=100, 62.5%; men who have sex with men: n=60, 37.5%) provided consent for venous blood sample collection. Three whole blood samples from female sex worker participants were discarded due to missing bar codes on the sample tubes. A total of 73 participants reported currently receiving PrEP in the BBS face-to-face interview (Table 2), and rapid urine tests were conducted on the samples from these participants (female sex workers: n=56, 77%; men who have sex with men: n=17, 23%).

Table 2. Expanded polling booth survey (ePBS) participant enrollment and data collection among female sex workers and men who have sex with men in Nairobi County, Kenya, from April to May 2023 (n=1157).

<table>
<thead>
<tr>
<th></th>
<th>Female sex workers (n=759), n (%)</th>
<th>Men who have sex with men (n=398), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBS participants</td>
<td>759 (100)</td>
<td>398 (100)</td>
</tr>
<tr>
<td>BBS? participants</td>
<td>758 (99.9)</td>
<td>398 (100)</td>
</tr>
<tr>
<td>HIV rapid tests conducted</td>
<td>758 (99.9)</td>
<td>398 (100)</td>
</tr>
<tr>
<td>Reactive HIV rapid test results</td>
<td>101 (13.3)</td>
<td>70 (17.6)</td>
</tr>
<tr>
<td>Venous blood samples collected</td>
<td>100 (13.2)</td>
<td>60 (15.1)</td>
</tr>
<tr>
<td>Urine tests conducted</td>
<td>56 (7.4)</td>
<td>17 (4.3)</td>
</tr>
</tbody>
</table>

*aBBS: behavioral and biological survey.

A total of 20 FGDs were completed with female sex workers (n=14, 70%) and men who have sex with men (n=6, 30%).

Data cleaning and analysis are ongoing. Aggregate responses from each PBS session and individual data from BBS respondents have been entered into separate data modules in the data entry software. The data will then be exported to SPSS (version 28.0; IBM Corp) for analysis. Weights will be applied to data during analyses to account for the sampling design and to adjust for unequal selection probabilities. CIs will be calculated, and descriptive analyses of the PBS and BBS data will be performed separately to address the study objectives.

The FGD data will be simultaneously translated and transcribed into English by the qualitative study team. Qualitative notes taken during the FGDs will be incorporated into a final transcript to provide appropriate contexts and details where necessary. Thematic analysis will be used to systematically capture recurring concepts, ideas, and topics that emerged during the discussions.

Furthermore, we will use the effective program coverage framework [16] to assess gaps in availability coverage, contact coverage, and use coverage within the condom, PrEP, and antiretroviral therapy program components. Equity analyses will be conducted with program coverage data to assess inequalities in the coverage of each component intervention based on age, geography, and location typology.

Discussion

Overview

Joint United Nations Programme on HIV and AIDS encourages countries to annually measure outcomes using nimble, rapid, and low-cost methods [28]. Through this study, the novel ePBS method was implemented for the first time, which expands upon the traditional PBS method by integrating BBS and FGD modules with minimal requirements for additional resources. Furthermore, this innovative approach adopted robust sampling methods to ensure the representativeness required for a population-based survey, unlike the traditional PBS method, which uses program-based sampling approaches [7,23]. This approach included data gathered at the group level (PBS and FGD) and individual level (BBS), encompassing behavioral, biological, and structural aspects of a combination prevention approach, from a total sample of 1157 participants within a
30-day timeframe. This method significantly enhances the range of tools currently available to collect routine outcome data from HIV prevention programs. Importantly, the ePBS method empowered the research team to efficiently gather a rich data set comprising both quantitative and qualitative data for program optimization and expanded the understanding of local HIV epidemics among key populations. The inclusion of community researchers in the study team and the involvement of peer educators at the sampled locations fostered trust and instilled confidence among study participants, enhancing the study’s credibility, which is reflected in the study’s high response rate.

Program and Policy Implications

Notably, this study draws upon a program science approach using the effective program coverage framework to assess HIV program coverage [16,18] and outcomes among female sex workers and men who have sex with men in Nairobi County, Kenya. This embedded approach facilitates the emergence of research questions from programmatic needs and challenges, and findings contribute to defining program and policy priorities. The study questions were formulated by PHDA-SWOP in collaboration with the community advisory board and the Ministry of Health, Nairobi County, who are the primary beneficiaries of the study findings. Engaging stakeholders involved in implementing key population programs, advocating for their rights, developing policies, and being accountable for creating an enabling environment enhance their commitment to use the study’s findings [29].

The study will provide comprehensive data on biological, behavioral, and structural outcomes, as well as program coverage, within an HIV prevention program tailored for female sex workers and men who have sex with men in Nairobi County. In alignment with the Global AIDS Strategy 2021 to 2026, which includes a high-level target of ensuring that 95% of people at risk of HIV infection have access to and use effective combination prevention options, our study findings will serve as a critical measure of the outcomes of such a combination prevention program [30]. In resource-limited settings, the use of ePBS methods can serve as a pragmatic and affordable approach to assessing program outcomes. It allows for rapid data collection and accommodates the integration of qualitative data collection methods and individualized surveys with standard PBS. Critical to a program science approach, because this research is embedded within Nairobi County’s key population program, the results of this ePBS study can be seamlessly incorporated to refine the program strategy to improve coverage and outcomes. In addition, this embeddedness allowed for the ePBS study to be carried out by the program team, thereby minimizing expenses in comparison to traditional integrated biobehavioral surveillance surveys.

Limitations

It is important to acknowledge that this study has inherent limitations, primarily due to its cross-sectional design. The results of this study will be descriptive in nature and will not contribute to the understanding of causal relationships. In addition, the study relies upon several self-reported variables, which are always subject to recall and social desirability biases. Finally, this study, and the ePBS method generally, greatly benefit, in terms of scalability and implementation, from being embedded within an established HIV prevention program for key populations, which might not be feasible in all settings.

Conclusions

The ePBS method has the potential to meet the demand for a rapid, cost-effective, and practical approach to regularly measuring HIV outcomes in key population programs and generating insights for ongoing program improvement. The study gathered crucial data on HIV outcomes and piloted a new method called ePBS, which combines PBS, BBS, and FGDs to enhance its practicability. Our findings will help improve the HIV prevention program for key populations and develop policies in Nairobi County.

Acknowledgments

The authors would like to thank the Nairobi County HIV team including Jesca Omai, Vanessa Chelimo, Shillah Mwavua, Douglas Kibanda, Ann Wahome, Edwin Mwangi, Nicasio Ngaji, and Lilian Tuara, who supervised data collection and ensured that an enabling environment was created for conducting the study. The authors especially thank Ephantus Wachira, Jackline Achieng, Rachel Lishindu, and Hadija Abdullah from the Partners for Health and Development in Africa (PHDA)-Sex Workers Outreach Program (SWOP) outreach team for their contribution to developing the tools and data collection. Furthermore, the authors would like to appreciate the contributions to data collection from the SWOP team members: Reuben Nyuru, Dan Khaemba, Martin Mbuthia, David Ngata, Tali Lumbasi, Kennedy Ojwang, Chrispo Nyamweya, Mary Kung’u, Anderson Mafusa, Kenneth Munene, Caroline Muriuki, Bernard Juma, Irene Wanjiru, Rachael Lishindu, Celilia Mwangi, Monica Okumu, and Pollet Ong’ayo. The authors would like to thank Erastis Irungu and Peggy Obam for setting up the PHDA laboratory for the study. They would also like to thank Santosh Singh for setting up the database for the study. In addition, the authors would like to appreciate the contribution of Dr Collins Owek, Apollo Kamau, Tony Kariri, Betty Rugendo, Wycliff Kioko, Festus Muriuki, and Laisa Lumumba from PHDA for providing supervisory support to the data collection team. The authors would like to thank Siobhan Malone and Geoff Garnet from the Bill & Melinda Gates Foundation (BMGF) for their support in the expanded polling booth survey method. This publication is made possible by the support of BMGF (investment ID INV050927). The views expressed herein are those of the authors and do not necessarily reflect the official policy or position of BMGF.
Data Availability
The data sets generated and analyzed during this study will not be publicly available due to the sensitivity of the data collected—sex work and same-sex relationships are criminalized in Kenya—and a prevailing antagonistic environment against LGBTQI+ (lesbian, gay, bisexual, transgender, queer, intersex, and others) populations exists across the country. However, data can be made available from the corresponding author on reasonable request.

Authors' Contributions
PB, LMM, SI, JK, RWK, A Kinyua, HM, A Kiplagat, PA, SS, FE, MLB, and JB participated in the design of the study. PB, SI, LMM, MLB, and SS developed the quantitative tools. LMM developed the qualitative tool. SI conducted sample size calculations. JK, PB, SI, LMM, HM, RWK, A Kinyua, and A Kiplagat trained the study teams and supervised data collection. JKO and VNN supported in the training and were involved in data collection. JK supervised biological sample collection and laboratory tests. PB wrote the first draft of the manuscript with support from LMM and SI. All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

References


Abbreviations

BBS: behavioral and biological survey
ePBS: expanded polling booth survey
FGD: focus group discussion
HTS: HIV and testing service
LGBTQI+: lesbian, gay, bisexual, transgender, queer, intersex, and others
NASCOP: National AIDS and STI Control Programme
PBS: polling booth survey

https://publichealth.jmir.org/2024/1/e54313

JMIR Public Health Surveill 2024 | vol. 10 | e54313 | p.1103
(page number not for citation purposes)
Assessing Outcomes in HIV Prevention and Treatment Programs With Female Sex Workers and Men Who Have Sex With Men: Expanded Polling Booth Survey Protocol


Refereed: 05.11.23; peer-reviewed by U Roy Paladhi, D Rao; comments to author 27.02.24; revised version received 16.03.24; accepted 20.03.24; published 19.06.24.

Please cite as:
Assessing Outcomes in HIV Prevention and Treatment Programs With Female Sex Workers and Men Who Have Sex With Men: Expanded Polling Booth Survey Protocol
JMIR Public Health Surveill 2024;10:e54313
URL: https://publichealth.jmir.org/2024/1/e54313
doi: 10.2196/54313
PMID:

©Parinita Bhattacharjee, Leigh M McClarty, Joshua Kimani, Shajy Isac, Rhoda Wanjiru Kabuti, Antony Kinyua, Jaffred Karakaja Okoyana, Virjinia Njeri Ndakuyu, Helgar Musyoki, Anthony Kiplagat, Peter Arimi, Souradet Shaw, Faran Emmanuel, Monica Gandhi, Marissa Becker, James Blanchard. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 19.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Assessing the Labeling Information on Drugs Associated With Suicide Risk: Systematic Review

Soo Min Jeon1*, PharmD, PhD; HyunJoo Lim2*, Hyo-bin Cheon2, PharmD; Juhee Ryu3, PharmD, PhD; Jin-Won Kwon4, PhD

1Jeju Research Institute of Pharmaceutical Sciences, College of Pharmacy, Jeju National University, Jeju, Republic of Korea
2College of Pharmacy, Kyungpook National University, Daegu, Republic of Korea
3Research Institute of Pharmaceutical Sciences, College of Pharmacy, Kyungpook National University, Daegu, Republic of Korea
4BK21 FOUR Community-Based Intelligent Novel Drug Discovery Education Unit, College of Pharmacy and Research Institute of Pharmaceutical Sciences, Daegu, Republic of Korea
*these authors contributed equally

Corresponding Author:
Jin-Won Kwon, PhD
BK21 FOUR Community-Based Intelligent Novel Drug Discovery Education Unit
College of Pharmacy and Research Institute of Pharmaceutical Sciences
80 Daehak-ro
Daegu, 41566
Republic of Korea
Phone: 82 539508580
Fax: 82 539508557
Email: jwkwon@knu.ac.kr

Abstract

Background: Drug-induced suicide (DIS) is a severe adverse drug reaction (ADR). Although clinical trials have provided evidence on DIS, limited investigations have been performed on rare ADRs, such as suicide.

Objective: We aimed to systematically review case reports on DIS to provide evidence-based drug information.

Methods: We searched PubMed to obtain case reports regarding DIS published until July 2021. Cases resulting from drugs that are no longer used or are nonapproved, substance use, and suicidal intentions were excluded. The quality of each case report was assessed using the CASE (Case Reports) checklist. We extracted data regarding demographics, medication history, suicide symptoms, and symptom improvement and evaluated the causality of DIS using the Naranjo score. Furthermore, to identify the potential suicidal risk of the unknown drugs, we compared the results of the causality assessment with those of the approved drug labels.

Results: In 83 articles, we identified 152 cases involving 61 drugs. Antidepressants were reported as the most frequent causative drugs of DIS followed by immunostimulants. The causality assessment revealed 61 cases having possible, 89 cases having probable, and 2 cases having definite relationships with DIS. For approximately 85% of suspected drugs, the risk of suicidal ADRs was indicated on the approved label; however, the approved labels for 9 drugs, including lumacaftor/ivacaftor, doxycycline, clozapine, dextromethorphan, adalimumab, infliximab, piroxicam, paclitaxel, and formoterol, did not provide information about these risks.

Conclusions: We found several case reports involving drugs without suicide risk information on the drug label. Our findings might provide valuable insights into drugs that may cause suicidal ADRs.

(JMIR Public Health Surveill 2024;10:e49755) doi:10.2196/49755

KEYWORDS
suicide; adverse drug events; review; drug; mental health; systematic review; drug induced suicide; drug reaction; substance use; suicidal; medication; suicide symptoms; suicidal risk; drugs; adverse drug event
Introduction

Suicide is a serious global public health problem. According to a 2019 report from the World Health Organization, more than 700,000 people die annually by suicide, implying that 1 person dies by suicide every 40 seconds [1]. Suicide results in fatal outcomes, enormous emotional distress for families, and a significant economic burden on society [2]. Therefore, much attention is required to reduce the burden of suicide. Suicide is associated with multiple factors, such as sex, age, history of suicide attempts, familial history of suicide, alcohol use, psychiatric disorders, and physical disorders. Among the identified risk factors, psychiatric and physical health disorders are the main controllable factors that prevent suicidal adverse drug reactions (ADRs). However, therapeutic drugs for these disorders have also been reported as a cause of suicide.

Psychiatric drugs, such as antidepressants, anxiolytic drugs, and antipsychotics, are representative drugs that have potential adverse effects, such as suicidal behavior. For instance, Stone et al [3] have conducted a meta-analysis involving 372 randomized controlled trials of antidepressants and reported that the odds ratio for suicidal behavior and ideation among patients aged younger than 25 years was 2.30 (95% CI 1.04-5.09). Moreover, several studies have demonstrated that some commonly prescribed drugs, such as hormonal contraceptives and beta-blockers, may induce suicidal ADRs [4,5]. According to a study by Qato et al [6], labels of 103 drugs indicated suicidal symptoms as side effects and the rate of administration of these drugs increased by approximately 1.09-1.35 folds from 2005-2006 to 2013-2014. Moreover, they revealed that patients who took these medicines had approximately 5.3% greater concurrent symptoms of these adverse events than those who did not.

Owing to the growing awareness regarding drug-induced suicide (DIS), several studies have investigated the association between specified drugs and suicide [7-15]. Most of these studies involved case-control and cohort study designs and disproportionality analysis of pharmacovigilance databases. However, investigating the overall suicidal risk of all available drugs is challenging. Furthermore, the databases used in observational studies may not fully account for detailed clinical presentations, symptoms, and all other relevant factors contributing to suicide risk, such as preexisting mental health conditions, social support, and family history. Therefore, more comprehensive research is necessary to fully understand the potential association between all available drugs and DIS, considering individual patient factors and drug interactions.

In pharmacovigilance, case reports are crucial in identifying new safety signals and alerting physicians regarding the potential rare ADRs [16]. Case reports provide detailed clinical presentations and increase awareness regarding drug safety, although they are limited in establishing causality evidence. Thus, this study aimed to conduct a systematic review of published case reports on DIS associated with all available drugs. We further assessed the causality of DIS in each case and compared it with drug labeling information to establish new safety signals of DIS.

Methods

Overview

We conducted a systematic review of case reports on DIS published until July 2021 according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines (Multimedia Appendix 1) [17]. The literature search was performed on PubMed based on a combination of the following keywords—“suicide” and “drugs”—using Medical Subject Headings and text words. Two independent authors validated the search strategy, using the detailed search terms presented in Textbox S1 in Multimedia Appendix 2. For drug keywords, we used the active ingredient and generic name of the drug to obtain as many case reports as possible. The active ingredients or generic names of drugs were searched on “DrugBank.” We limited the article type and language to case reports and English, respectively.

Study Selection

In this systematic review, we included case reports on suicidal behavior after administering the prescribed medications. We defined suicidal ideation and behavior as self-harm, suicidal ideation, and completed suicide. The exclusion criteria were (1) case reports on drugs that were no longer used or were nonapproved, (2) case reports on patients who took drugs for suicidal purposes, and (3) case reports without any mention of suicidal ADRs (eg, suicide genes). There were no restrictions on patient conditions, such as age, sex, psychiatric status, and outpatient or inpatient treatment.

Two authors with a background in pharmacology were independently involved in the process of article selection. They screened the titles and abstracts of all searched articles that met our eligibility criteria. If no consensus was reached between the 2 authors, a third independent author was solicited to make the final decision.

Data Extraction

We extracted data from relevant case reports regarding patient demographics (ie, age, sex, psychiatric status, and socioeconomic status), medication history (ie, ADR, drug dosing, time to event, concomitant medication, and laboratory test), symptoms, improvement in symptoms after dose reduction or drug discontinuation, and causality assessment results.

Data Synthesis and Analysis

For all cases of DIS, we summarized patients’ demographics, including age, sex, and underlying psychiatric status. The demographics of cases were presented as the frequencies and percentages. Age was categorized into 3 groups (<19, 19-59, and 60+ years). Data were presented according to the classes of suspected drugs involved in DIS cases. The drug classes were classified according to the anatomical therapeutic chemical (ATC) code [18]. The third level of the ATC (ie, the chemical, pharmacological, or therapeutic subgroup) was mainly used to categorize drugs; some drugs were classified based on the second level of the ATC (ie, the pharmacological or therapeutic subgroup). Detailed code classification is presented in Table S1 in Multimedia Appendix 2.
Furthermore, we assessed the probability of drug administration that leads to DIS using the Naranjo ADR score [19], which has been widely used as a tool for causality assessment in case reports. This scale included 10 simple and clarified questions, and the answer to each question is categorized into 3 options “Yes,” “No,” and “Unknown.” This scale can consider the missing values when evaluating the causality. The calculated score was classified into 4 levels, as follows: doubtful (≤0), possible (1-4), probable (5-8), and definite (≥9). As some questions in the original version of the Naranjo algorithm were not detailed enough to assess causality, we defined and used detailed criteria (see Multimedia Appendix 3). Although few articles had already included information regarding the Naranjo score as a causality assessment, we reevaluated this score for all articles. The quality of the selected reports was evaluated using the CARE (Case Reports) guidelines [20].

Investigation of Potential Drugs for DIS Risk

We compared the causality assessment results with corresponding label information to identify drugs with the potential risk of suicide. In this analysis, we used the average Naranjo score for each drug as a representative value of our causality assessment. We reviewed the label information for all suspected drugs using the Food and Drug Administration (FDA)-approved label and Micromedex and classified these drugs into 3 groups, as shown in Textbox 1 (high-, intermediate-, and low-level evidence). Drugs were considered “high-level evidence” if the label information indicated a direct expression of “suicide” (ie, suicide, suicide attempts, suicidal ideation, suicidal thoughts, suicidal behavior, suicidality, or suicide risk). Drugs were considered “intermediate-level evidence” if the label information did not directly indicate suicide but specified psychiatric disorders often associated with suicide-related symptoms (ie, depression, anxiety, delirium, hallucinations, and psychotic behavior). The remaining drugs were considered “low-level evidence.” In our case reports, if the suspected drugs associated with DIS were classified as low-level evidence according to the label information, we considered them as potentially unrecognized drugs for DIS.

Textbox 1. Current evidence levels for reported drugs associated with DIS based on the approved labeling information.

<table>
<thead>
<tr>
<th>Evidence Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Drugs with information that includes a direct expression of “suicide,” such as suicide, suicide attempts, suicidal ideation, suicidal thoughts, suicidal behavior, suicidality, and suicide risk.</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Drugs with information regarding psychiatric disorders associated with suicide-related symptoms, such as depression, anxiety, delirium, hallucinations, and psychotic behavior.</td>
</tr>
<tr>
<td>Low</td>
<td>Drugs without any adverse drug reaction information for suicide and suicide-related symptoms.</td>
</tr>
</tbody>
</table>

Results

We identified 2877 articles from the PubMed database and removed 7 duplicated records. After reviewing the titles and abstracts of these articles for eligibility, we excluded 2778 articles that met the exclusion criteria. Overall, 9 of the remaining 92 articles were excluded because they did not contain full text (n=7) or had ineligible study designs (n=2). Finally, we selected 83 articles that involved 152 individual drug-suicide relationships (Figure 1). In total, there were 70, 65, and 17 cases of suicidal ideation, suicide attempts, and completed suicide, respectively. The studies had total scores ranging from 17/30 to 24/30 according to the CARE checklist (Multimedia Appendix 4).
In these case reports, 61 individual drugs were identified as drugs associated with DIS and were grouped into 31 classes. Table 1 summarizes the demographics and characteristics of DIS cases according to the class of the suspected drugs. There were 26 cases of DIS related to antidepressants, which was the most frequently reported drug class. For antidepressants, most patients were aged between 19 and 59 years (n=16) followed by the age group of 2-18 years (n=8). A psychiatric diagnosis was made in all patients, except for 1; and 6 patients had a history of suicide attempts. In 25 patients (with a mean age of 37.8 years), immunostimulants were identified as the causative drug for DIS. A total of 7 patients showed a personal or family history of psychiatric disorders or suicide attempts. Antiepileptics were identified as the causative drugs for DIS in 15 patients, of whom 8 had a history of psychiatric disorders. Furthermore, 8 patients associated with anabolic steroids were all male and aged between 19 and 59 years (mean age of 26.5 years). All 8 patients associated with psychostimulants were aged 2-18 years (mean age 10.3 years; range 9-13 years), which was the youngest among all the drug classes.
Table 1. Demographics of suspected drug for suicide by drug classes.

<table>
<thead>
<tr>
<th>Drug Class</th>
<th>N(^a) (N/152), n (%)</th>
<th>Sex(^b)</th>
<th>Female, n (%)</th>
<th>Age (years)(^b)</th>
<th>Mean (range)</th>
<th>&lt;19, n (%)</th>
<th>19-60, n (%)</th>
<th>60+, n (%)</th>
<th>History of psychiatric status(^b)</th>
<th>Psychiatric diagnosis, n (%)</th>
<th>Substance abuse, n (%)</th>
<th>Suicide attempts, n (%)</th>
<th>Family history, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antidepressants</td>
<td>26 (17.0)</td>
<td>Male, n (%)</td>
<td>9 (34.6)</td>
<td>17 (65.4)</td>
<td>29.1 (10-62)</td>
<td>8 (30.8)</td>
<td>16 (61.5)</td>
<td>2 (7.7)</td>
<td>25 (96.2)</td>
<td>22 (84.6)</td>
<td>6 (23.1)</td>
<td>3 (11.5)</td>
<td></td>
</tr>
<tr>
<td>Immunostimulants</td>
<td>25 (16.3)</td>
<td>12 (48.0)</td>
<td>13 (52.0)</td>
<td>37.8 (19-62)</td>
<td>0 (0.0)</td>
<td>24 (96.0)</td>
<td>1 (4.0)</td>
<td>4 (16.0)</td>
<td>6 (24.0)</td>
<td>3 (12.0)</td>
<td>3 (12.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antiepileptics</td>
<td>15 (9.8)</td>
<td>10 (66.7)</td>
<td>5 (33.3)</td>
<td>35.1 (10-66)</td>
<td>2 (13.3)</td>
<td>12 (80.0)</td>
<td>1 (6.7)</td>
<td>8 (53.3)</td>
<td>9 (60.0)</td>
<td>4 (26.7)</td>
<td>1 (6.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anabolic steroids</td>
<td>8 (5.2)</td>
<td>8 (100.0)</td>
<td>0 (0.0)</td>
<td>26.5 (21-34)</td>
<td>0 (0.0)</td>
<td>8 (100.0)</td>
<td>0 (0.0)</td>
<td>7 (87.5)</td>
<td>7 (87.5)</td>
<td>0 (0.0)</td>
<td>4 (50.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychostimulants</td>
<td>8 (5.2)</td>
<td>6 (75.0)</td>
<td>2 (25.0)</td>
<td>10.3 (9-13)</td>
<td>8 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>8 (100.0)</td>
<td>8 (100.0)</td>
<td>0 (0.0)</td>
<td>4 (50.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs used in addictive disorders</td>
<td>6 (3.9)</td>
<td>2 (33.3)</td>
<td>4 (66.7)</td>
<td>35.5 (19-67)</td>
<td>0 (0.0)</td>
<td>6 (100.0)</td>
<td>0 (0.0)</td>
<td>6 (100.0)</td>
<td>6 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antiacne preparations</td>
<td>6 (3.9)</td>
<td>5 (83.3)</td>
<td>1 (16.7)</td>
<td>44.8 (38-58)</td>
<td>1 (16.7)</td>
<td>4 (66.7)</td>
<td>0 (0.0)</td>
<td>4 (66.7)</td>
<td>5 (83.3)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF(^c) and AR(^d) products, nonsteroids</td>
<td>5 (3.3)</td>
<td>3 (60.0)</td>
<td>2 (40.0)</td>
<td>19.2 (17-20)</td>
<td>0 (0.0)</td>
<td>5 (100.0)</td>
<td>0 (0.0)</td>
<td>4 (80.0)</td>
<td>4 (80.0)</td>
<td>0 (0.0)</td>
<td>1 (20.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antipsychotics</td>
<td>5 (3.3)</td>
<td>4 (80.0)</td>
<td>1 (20.0)</td>
<td>47.6 (37-57)</td>
<td>0 (0.0)</td>
<td>4 (80.0)</td>
<td>1 (20.0)</td>
<td>5 (100.0)</td>
<td>5 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory system products</td>
<td>5 (3.3)</td>
<td>0 (0.0)</td>
<td>5 (100.0)</td>
<td>15.0 (12-17)</td>
<td>5 (100.0)</td>
<td>0 (0.0)</td>
<td>3 (60.0)</td>
<td>3 (60.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibiotics</td>
<td>4 (2.6)</td>
<td>3 (75.0)</td>
<td>1 (25.0)</td>
<td>30.8 (18-53)</td>
<td>1 (25.0)</td>
<td>3 (75.0)</td>
<td>0 (0.0)</td>
<td>1 (25.0)</td>
<td>1 (25.0)</td>
<td>0 (0.0)</td>
<td>2 (50.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antimycobacterial</td>
<td>4 (2.6)</td>
<td>4 (100.0)</td>
<td>0 (0.0)</td>
<td>35.5 (20-55)</td>
<td>0 (0.0)</td>
<td>4 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>2 (50.0)</td>
<td>1 (25.0)</td>
<td>1 (25.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antivirals</td>
<td>4 (2.6)</td>
<td>3 (75.0)</td>
<td>1 (25.0)</td>
<td>30.0 (13-47)</td>
<td>1 (25.0)</td>
<td>3 (75.0)</td>
<td>0 (0.0)</td>
<td>2 (50.0)</td>
<td>3 (75.0)</td>
<td>1 (25.0)</td>
<td>2 (50.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immunosuppressants</td>
<td>4 (2.6)</td>
<td>2 (50.0)</td>
<td>2 (50.0)</td>
<td>46.0 (32-56)</td>
<td>0 (0.0)</td>
<td>4 (100.0)</td>
<td>0 (0.0)</td>
<td>1 (25.0)</td>
<td>1 (25.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analgesics and antipyretics</td>
<td>3 (2.0)</td>
<td>1 (33.3)</td>
<td>2 (66.7)</td>
<td>53.3 (39-66)</td>
<td>0 (0.0)</td>
<td>2 (66.7)</td>
<td>1 (33.3)</td>
<td>3 (100.0)</td>
<td>3 (100.0)</td>
<td>1 (33.3)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anesthetics</td>
<td>3 (2.0)</td>
<td>1 (33.3)</td>
<td>2 (66.7)</td>
<td>39.3 (25-64)</td>
<td>0 (0.0)</td>
<td>2 (66.7)</td>
<td>1 (33.3)</td>
<td>2 (66.7)</td>
<td>3 (100.0)</td>
<td>1 (33.3)</td>
<td>1 (33.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antimalarials</td>
<td>3 (2.0)</td>
<td>3 (100.0)</td>
<td>0 (0.0)</td>
<td>31.3 (27-40)</td>
<td>0 (0.0)</td>
<td>3 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypnotics and sedatives</td>
<td>3 (2.0)</td>
<td>1 (33.3)</td>
<td>2 (66.7)</td>
<td>48.3 (38-59)</td>
<td>0 (0.0)</td>
<td>3 (100.0)</td>
<td>0 (0.0)</td>
<td>3 (100.0)</td>
<td>3 (100.0)</td>
<td>0 (0.0)</td>
<td>1 (33.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-Parkinson drugs</td>
<td>2 (1.3)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>48.5 (44-53)</td>
<td>0 (0.0)</td>
<td>2 (100.0)</td>
<td>0 (0.0)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corticosteroids</td>
<td>2 (1.3)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>55.5 (43-68)</td>
<td>0 (0.0)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cough suppressants</td>
<td>2 (1.3)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>56.0 (46-66)</td>
<td>0 (0.0)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs for obstructive airway diseases</td>
<td>2 (1.3)</td>
<td>1 (50.0)</td>
<td>1 (50.0)</td>
<td>40.0 (9-71)</td>
<td>1 (50.0)</td>
<td>0 (0.0)</td>
<td>1 (50.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Sex</td>
<td>Age (years)</td>
<td>History of psychiatric status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>---------</td>
<td>-------------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(N/152), n (%)</td>
<td>Male, n (%)</td>
<td>Female, n (%)</td>
<td>Mean (range)</td>
<td>&lt;19, n (%)</td>
<td>19-60, n (%)</td>
<td>60≤, n (%)</td>
<td>Psychiatric diagnosis, n (%)</td>
<td>Substance abuse, n (%)</td>
<td>Suicide attempts, n (%)</td>
<td>Family history, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antimigraine preparations</td>
<td>1 (0.7)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>43.0 (43-43)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>1 (100.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antimycotics</td>
<td>1 (0.7)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>67.0 (67-67)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antineoplastic agents</td>
<td>1 (0.7)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>52.0 (52-52)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiolytics</td>
<td>1 (0.7)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>62.0 (62-62)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>1 (100.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-blocking agents</td>
<td>1 (0.7)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>21.0 (21-21)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs for constipation</td>
<td>1 (0.7)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>61.0 (61-61)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hormone antagonists and related agents</td>
<td>1 (0.7)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>45.0 (45-45)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propulsive</td>
<td>1 (0.7)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>30.0 (30-30)</td>
<td>0 (0.0)</td>
<td>1 (100.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Number of patients who were duplicated because the cases were associated with more than 2 individual drugs.
- The proportion was estimated as the proportion of reported cases within each drug classes.
- AI: anti-inflammatory.
- AR: antirheumatic.

Figure 2 and Table S2 in Multimedia Appendix 2 show the results of the causality assessment of all cases based on the Naranjo algorithm. The results revealed a definite causal association with DIS in 2 cases, probable association in 90 cases, and possible association in 61 cases. The 2 cases with definite causation of DIS were identified in cases associated with doxycycline (antibacterial) and perampanel (antiepileptics), respectively.
Table 2 shows the results of the comparative analysis between our results and the drug labels. The labels of 51 (85%) of 60 drugs contained information regarding the risk of suicidal ADRs. The high- and intermediate-level evidence groups included 33 and 18 drugs, respectively. However, the risk of suicidal ADRs was not indicated on the label of the 9 suspected drugs. Of these, 6 drugs had a “probable” causal relationship (lumacaftor/ivacaftor, doxycycline, piroxicam, clozapine, adalimumab, and paclitaxel), whereas the remaining 3 drugs had “possible” causality (infliximab, dextromethorphan, and formoterol).

Table 2. Comparative assessment of suspected drugs in suicidal adverse drug reactions between reviewed case reports and the approved labeling information.

<table>
<thead>
<tr>
<th>Causality assessment based on the Naranjo algorithm</th>
<th>High (N=33), n</th>
<th>Intermediate (N=18), n</th>
<th>Low (N=9), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probable</td>
<td>18</td>
<td>12</td>
<td>6(^c)</td>
</tr>
<tr>
<td>Possible</td>
<td>15</td>
<td>6</td>
<td>3(^d)</td>
</tr>
</tbody>
</table>

\(^a\) In this analysis, we excluded anabolic androgenic steroids because they did not present the information of the involved agent.

\(^b\) Causality was evaluated for each drug based on the average Naranjo scores in the case reports for each drug.

\(^c\) Lumacaftor/ivacaftor, doxycycline, piroxicam, clozapine, adalimumab, and paclitaxel.

\(^d\) Infliximab, dextromethorphan, and formoterol.
The number of cases and the criteria and average score of Naranjo for 17 case reports implicating suspected drugs with low-level evidence on the approved labels were as follows, respectively: lumacaftor/ivacaftor (5 cases, probable 5); doxycycline (3 cases, probable 7.3); piroxicam (1 case, probable 5); clozapine (2 cases, probable 6); adalimumab (2 cases, probable 6.0); paclitaxel (1 case, probable 5.0); infliximab (1 case, possible 4); dextromethorphan (2 cases, possible 3.5); and formoterol (1 case, possible 3). Multimedia Appendix 5 [21-55] provides detailed information on the evidence of the Naranjo score of each case.

Discussion

Principal Findings

To the best of our knowledge, this is the first comprehensive systematic review of DIS based on published case reports. While several studies have reported on the suicide risk of drugs [7-15], most of them have concentrated on limited drugs for which risks have previously been reported. Therefore, our study aimed to bridge this gap by conducting a systematic review of case reports for all available drugs with suicidal risk, including those not previously known to have such a risk. We identified 61 drugs from 29 classes associated with DIS based on published case reports, of which 9 drugs showed a potential risk of suicide despite not being mentioned in their labels.

Most documented DIS cases involved drugs that already had a potential suicidal risk indicated on their labels. This tendency might result from an increased focus on drugs recognized for their potential risk of suicidal ADRs. For instance, antidepressants, the most frequently reported drugs related to suicidal ADRs, carry black box warnings of suicidal risk issued by the FDA. While some drug labels did not directly mention the risk of suicide, they indicated the potential psychiatric ADRs that can lead to suicidal behaviors. Propranolol labels present depression and hallucinations among its psychiatric side effects, which could contribute to triggering suicidal events. Recent studies using pharmacovigilance databases have also strengthened evidence regarding the suicidal risks associated with these drugs [11-15].

This systematic review has found 9 drugs—lumacaftor/ivacaftor, doxycycline, piroxicam, clozapine, paclitaxel, adalimumab, infliximab, dextromethorphan, and formoterol—that potentially carry suicidal risks not mentioned on the labels. Despite the absence of available data on labels, previous studies have investigated the potential for suicide or suicide-related behaviors associated with these drugs. However, the precise mechanism related to these suicidal risks remains still unclear. For instance, doxycycline, which exhibited the highest causality for suicidal ADRs in our study, might be related to heightened psychiatric adverse effects due to increased retinoid levels by CYP450 inhibitory effect [21]. In addition, other studies suggested changes in the gut microbiome composition after antibiotic use, leading to the incidence of psychiatric disorders, such as depression and schizophrenia [22-24]. A previous systematic review reported that several clinical trials reported the improvement of depressive, anxiety, and mood symptoms after probiotic treatment [25].

The suicide risk of piroxicam may be related to inhibiting cyclooxygenase activity involved in synthesizing prostaglandins [26,27]. However, there is a lack of direct information about changes in prostaglandin levels [28]. Paclitaxel may be attributed to glutamine depletion and neurotransmitter [29], even though a clear association with suicide is challenging due to the limited information on patients' posttreatment status [30]. Other drugs, tumor necrosis factor-α inhibitors [31-35], paclitaxel [29], dextromethorphan [36-38], formoterol [39], clozapine [40-43], and lumacaftor/ivacaftor [44-46] are not easy to explain the clear mechanism for the suicidal events. Despite these complexities, most cases involving these 9 drugs reported symptom improvement after drug discontinuation. Hence, the potential risk of suicide associated with these drugs cannot be disregarded, and further studies are a crucial for clear understanding of the potential suicidal risk of these drugs.

Limitations

This study also has certain limitations. First, as case reports are not experimental studies conducted with a placebo, the causal relationship between drugs and ADRs cannot be determined. Although we used the Naranjo algorithm to evaluate causality, other factors, such as the underlying psychiatric disorders and other drugs, could affect suicidal events. Therefore, our findings only provide some safety signals regarding which drugs may have a higher risk of suicide not identified in clinical trials [56]. Second, the number of cases for each drug is insufficient for generalizing the findings to the population level. Third, we used PubMed only and did not manually search other journals or sources of gray literature and restricted the inclusion of studies to those published in English. Therefore, our findings might not represent all published case reports. Nevertheless, we constructed a detailed search query using the term of all marketed drugs, which may minimize the impact of limitation. Furthermore, studies have demonstrated that excluding non-English studies might have little effect on the results [57,58].

Conclusions

Our systematic review of case reports identified a potential risk of suicide associated with several drugs. Most of the cases involved drugs with a documented suicide risk on the label, but 9 drugs were identified without such reporting. Although it is difficult to establish evidence of ADRs based on case reports alone, our findings highlighted the need to raise awareness of the potential for DIS beyond what is documented in drug labels. Therefore, further research exploring suicidal ADRs across various drugs is essential.
Acknowledgments
This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022R1I1A1A01065589). This research was also supported by “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE; 2023RIS-009).

Authors’ Contributions
SMJ contributed to lead in conceptualization. SMJ conducted the study selection, data extraction, interpret the analyses, and wrote and edited the original paper. HJL conducted the study selection, data extraction, and wrote and edited the original paper. HbC conducted the study selection and data extraction. JR contributed to interpret the analyses and edited the original paper. JWK contributed to interpret the analyses, edited the paper, and coordinated the project. All authors approved the final version to be published.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[PDF File (Adobe PDF File), 66 KB - publichealth_v10i1e49755_app1.pdf]

Multimedia Appendix 2
Supporting methods and results.
[DOCX File, 112 KB - publichealth_v10i1e49755_app2.docx]

Multimedia Appendix 3
Causality assessment using the Naranjo algorithm.
[DOCX File, 26 KB - publichealth_v10i1e49755_app3.docx]

Multimedia Appendix 4
CARE check list.
[XLSX File (Microsoft Excel File), 19 KB - publichealth_v10i1e49755_app4.xlsx]

Multimedia Appendix 5
Summary of cases of suspected drugs without the required information on the approved label.
[DOCX File, 69 KB - publichealth_v10i1e49755_app5.docx]

References


38. Highlights of prescribing information. AUVELITY (dextromethorphan hydrobromide and bupropion hydrochloride) extended-release tablets, for oral use. 2022. URL: https://www.accessdata.fda.gov/drugsatfda_docs/label/2022/215430s000lblCorrect2.pdf [accessed 2022-09-20]


Abbreviations

ADR: adverse drug reaction
ATC: anatomical therapeutic chemical
CARE: Case Reports
DIS: drug-induced suicide
FDA: Food and Drug Administration
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

©Soo Min Jeon, HyunJoo Lim, Hyo-bin Cheon, Juhee Ryu, Jin-Won Kwon. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 30.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
The Impact of Rapid Handpump Repairs on Diarrhea Morbidity in Children: Cross-Sectional Study in Kwale County, Kenya

Patrick Thomson1,2, MEng, PhD; Justin Stoler3,4, MPH, PhD; Michelle Byford5, MPhil, PhD; David J Bradley1,6, MA, DM, DSc

1School of Geography and the Environment, University of Oxford, Oxford, United Kingdom
2Department of Engineering Science, University of Oxford, Oxford, United Kingdom
3Department of Geography and Sustainable Development, University of Miami, Coral Gables, FL, United States
4Department of Public Health Sciences, University of Miami, Miami, FL, United States
5Barts Health NHS Trust, London, United Kingdom
6London School of Hygiene & Tropical Medicine, London, United Kingdom

Corresponding Author:
Patrick Thomson, MEng, PhD
Department of Engineering Science
University of Oxford
Old Road Campus Building
Headington
Oxford, OX3 7DQ
United Kingdom
Phone: 44 1865 617675
Email: patrick.thomson@ouce.ox.ac.uk

Abstract

Background: Handpumps are used by millions of people as their main source of water. Although handpumps represent only a basic form of water provision, there have been continuous efforts to improve the performance of these systems as they are likely to remain in use for many years to come. The introduction of a professional maintenance service in southern Kenya has shown an order of magnitude improvement in operational performance over community-based management, with 90% of handpump faults repaired within 3 days of being reported. One driver behind these efforts is the assumption that a more reliable water supply will lead to a reduction in water-related disease. However, it is not clear if operational improvements lead to health gains. Despite limited empirical evidence, some modeling studies suggest that even short periods of drinking contaminated water can lead to disproportionate negative health impacts.

Objective: The aim of this study was to assess whether the improvements in operational performance from the rapid professional maintenance of rural handpumps lead to improved household health outcomes.

Methods: From a sample of households using handpumps as their primary water source in Kwale County, Kenya, we measured the 2-week prevalence of World Health Organization–defined diarrhea in children, reported by the adult respondent for each household. We compared the rates before and after a period during which the households’ handpumps were being professionally maintained. We then conducted a cross-sectional analysis, fitting logistic regression models with reported diarrhea as the dependent variable and speed of repair as the independent exposure of interest, adjusting for household socioeconomic characteristics; dwelling construction; and Water, Sanitation, and Hygiene (WASH)-related factors. We fitted an additional model to examine select interactions between covariates.

Results: Reported diarrhea in children was lower in households whose pumps had been repaired within 24 hours (adjusted odds ratio 0.35, 95% CI 0.24-0.51). This effect was robust to the inclusion of multiple categories of covariates. No reduction was seen in households whose pump repairs took more than 24 hours. Analysis of interaction terms showed that certain interventions associated with improved WASH outcomes were only associated with reductions in diarrhea in conjunction with socioeconomic improvements.

Conclusions: Only pump repairs consistently made within 24 hours of failure led to a reduction in diarrhea in the children of families using handpumps. While the efficacy of reduction in diarrhea is substantial, the operational challenges of guaranteeing same-day repairs limits the effectiveness of even best-in-class pump maintenance. Maintenance regimes that cannot bring...
handpump downtimes close to zero will struggle to generate health benefits. Other factors that reduce diarrhea prevalence have limited effect in isolation, suggesting that WASH interventions will be more effective when undertaken as part of more holistic poverty-reduction efforts.

(Keywords: water; Kenya; WASH; water, sanitation, and hygiene; maintenance; diarrhea; diarrhoea; SDG 6; service provision; handpump; child health)

Introduction

In rural areas with no grid electricity or limited funds to run diesel generators, many communities depend upon handpumps to access shallow groundwater for their daily water needs. Groundwater, in contrast to surface water, is more likely to be available year-round and less likely to require treatment to be potable [1]. Despite efforts throughout the period of the Millennium Development Goals and now in response to Sustainable Development Goal (SDG) 6.1, “safely managed water for all,” millions of people will still be relying on water from handpumps beyond 2030. While “safely managed” water—as defined by the World Health Organization (WHO)/United Nations Children’s Fund (UNICEF) Joint Monitoring Programme (JMP) as the use of “an improved drinking water source which is located on premises, available when needed, and free of fecal and priority chemical contamination”—remains the goal for water services for all, the size of the task and difficulty in achieving this goal has been acknowledged through the JMP Service Ladder’s inclusion of an interim service level of “basic” water, the category that includes most handpumps [2].

There are many reasons for aspiring to the goal of safely managed water for all. One of the main drivers of this goal—and Water, Sanitation, and Hygiene (WASH) interventions more generally—has been the understanding that disease morbidity can be reduced through better water service provision [3]. Piped water to the home, the most usual conception of safely managed water, is associated with lower morbidity, in particular of diarrheal disease [4]. However, with 26% of the world’s population—and 73% in sub-Saharan Africa—still lacking safely managed water and the achievement of SDG 6.1 unlikely at current rates of progress [5], it is important to understand the health implications of having only a basic water supply, as these are at risk of contamination [6], which can lead to diarrheal disease [7].

The contribution of diarrheal disease to the worldwide disease burden is falling but remains substantial [8]. Diarrhea itself contributes to the burden of disease for individuals and consequently to the global burden of disease estimates. Diarrhea is the cause of 1.31 million deaths worldwide, with 303,045 of those being of children under 5 years of age in sub-Saharan Africa; for that same demographic, diarrhea also causes 27 million disability-adjusted life years (DALYs) [9], which corresponds to approximately one in every seven DALYs. Diarrhea interacts with nutrition, reducing nutrient absorption [10,11]; therefore, this figure may significantly underestimate the disease burden from diarrhea by up to 39% [12]. Again, this is especially egregious in children, as undernutrition can produce long-term, nonrecoverable negative impacts on physical and cognitive development [13-16], with associated long-term economic and welfare impacts and related policy implications [17,18].

Community-based management, the model by which local communities are responsible for the repair and maintenance of their pumps, became the default management model for rural water supply in the late 1970s, with mixed success [19-24], and an estimated one in four handpumps in sub-Saharan Africa are out of action today [25]. Since the 2000s, there has been increasing interest in professional management models more akin to how urban piped water systems are run, whether directly by governments or subcontracted to a water service provider [26-29]. More recent evidence indicates that communities do have a preference for private sector involvement in service provision over community management or direct government provision, but only if service performance is high [30-32].

We here present results obtained from a cross-sectional survey following the introduction of a professional handpump maintenance service in Kwale County in southern Kenya. The service was initially funded by a research program that was investigating whether the intervention of a rapid, professional repair service, supported by data measuring handpump use using sensors attached to the pumps [33], could result in repairing handpumps faster than had been the case under community-based management. The rapid repair service reduced average pump downtimes from approximately 1 month to a few days, an improvement that is valued by households and communities [34-37]. This study examined whether this improvement in service performance also led to health benefits. Although there is limited empirical evidence to inform service decisions, some studies suggest that even the shortest periods of drinking contaminated drinking water can lead to disproportionate negative health impacts [38,39], implying a threshold effect with a nonlinear relationship between quality of service and the health benefits of that service. Obtaining empirical evidence about this relationship would inform both local operational decisions and wider WASH policy, and in doing so shed light on the role of handpumps in meeting SDG 6.1.

We hypothesized that shorter pump downtimes resulting from the professional maintenance service would be associated with lower rates of child diarrhea. We examined this by combining data on pump failures and subsequent repairs, generated by the pump mechanics, with self-reported child diarrhea data from households that were surveyed before the repair service was introduced and after 18 months of the service. To overcome the
challenges associated with an observational study, we analyzed the data using two complementary techniques. The first technique compared child diarrhea prevalence (during a 2-week period) in the same households before and after the repair service was introduced to control for time-invariant household characteristics. The second approach involved a cross-sectional analysis comparing households reporting diarrhea with those not reporting diarrhea, adjusting for household characteristics gathered at the same time as the self-reported diarrhea data.

**Methods**

**Study Area**

This study took place in Kwale County, located on the southern coast of Kenya between Mombasa and the border with Tanzania. This region covers an area of around 8300 km$^2$ and has a population of 720,000 that is 82% rural, with the seventh-highest poverty rate among Kenya’s 47 counties. The study area was limited to the Msambweni and Matuga subcounties closer to the coast, covering an area of around 2200 km$^2$. The coastal climate has a bimodal rainfall pattern with an average annual precipitation of 1400 mm with significant interannual variability. Within the study area there is heterogeneity both in socioeconomic characteristics and physical geography. We simplified the study area into three geographical zones for the purposes of analysis (see Figure 1) [36]. In the area in and around the urban center of Ukunda, the population density is higher and there are more alternative sources of water available during handpump breakdowns. There is also easier access to medical facilities in this area. Inland, where shallow wells are supplemented by deeper boreholes drawing water from sandstone, the population is less dense and livestock ownership is more common. Being more sparsely populated makes access to other sources of water more challenging; however, the deeper boreholes into the sandstone enable better source protection relative to shallow wells and improve the likelihood of a given handpump producing uncontaminated water. This is in contrast to the coastal strip where the population is still rural but the density is higher. Alternative nonsurface water sources in the form of other handpumps and shallow wells are relatively close by. However, this proximity of other sources, combined with the high population density and the fact that the wells in this area are mostly shallow and were hand-dug into the karstic coral, make source protection extremely difficult in this area.

![Figure 1. Map of the study area in Kwale County, Kenya, with the locations of handpumps marked in zones defined in Katuva et al [36]. Red indicates coastal, green indicates inland, and blue indicates Ukunda.](image-url)
Handpump Repair Service

The professional repair service was set up in Kwale as part of a project examining whether professional handpump maintenance, aided by better data on handpumps, could lead to reduced pump downtimes [33,40]. A waterpoint mapping exercise was undertaken in August 2013, recording the location and functionality of the handpumps in the study area, along with basic information about pump repair and management arrangements. A total of 626 pumps were found, with 351 identified as having been functional during the previous 12 months. We then set up an office in Msambweni, near the center of the study area, and recruited two experienced pump fundis (mechanics) and a manager. Two hundred handpumps were offered a free maintenance service, which corresponded to all of the functioning handpumps in a smaller geographical area across Msambweni and Matuga subcounties. The service started in February 2014 and continued until the end of 2015.

We provided the mechanics with motorbikes, tools, and a stock of spare parts to be able to quickly respond to handpump failures and complete most repairs during a single visit to the pump.

Figure 2. Timeline of activities that generated data used in this study.

The survey questionnaire included questions on household makeup and demographics; socioeconomic status, including key consumption and wealth indicators; basic self-reported health indicators for each household member; water use, collection, and storage; and waterpoint institutional arrangements, payment policies, and behavior. These questions were asked of one respondent for each household. For this study, we used the presence or absence of diarrhea reported for children in the household in the 2 weeks prior to the survey as the primary outcome measure; the other items were considered as covariates in the analysis. The only variable used from the 2013 survey was self-reported diarrhea in the 2 weeks prior to the survey.

A second household survey was conducted in March, April, and May of 2015. This was a repeat of the first survey, the aim being to see what had changed for the communities in the study area during the time of the project, specifically in relation to the pump repair service. The two surveys were matched: 59% were matched by household and respondent, 30% were matched by household but not respondent, and the remaining 11% represent cases where a different family/household was living in the dwelling. All the socioeconomic data used in this study were from the 2015 survey as these data represented the households’ situation during the time they were receiving the free repair service. At the time of this survey, the maintenance service had been operational for at least 12 months.

When repairs required more labor, such as removing rising mains, the communities provided pro-bono support to the mechanics. Toward the end of 2015, we spun the repair service out from the research program, establishing it as an independent locally owned entity, named FundiFix. At this point, communities were given the opportunity to continue receiving the maintenance service through paying a heavily subsidized monthly fee [41].

Data Sources

The study timeline is schematically summarized in Figure 2. Building on the mapping exercise completed in August 2013, a more extensive household survey was conducted in October and November 2013. A total of 2508 households who had access to a functioning handpump were surveyed, randomly selected from those using each of the 351 functional handpumps identified during the earlier waterpoint mapping exercise. Household selection was made by the enumerator team leader, selecting a direction and distance from the pump for each enumerator based on a throw of a six-sided die. A mean of 7.1 households were interviewed per handpump.

The two mechanics and their manager who were providing the handpump maintenance service kept detailed records of repairs undertaken, the most significant details being when the breakdown was first identified, when they attended, and when the repair was completed. These records yielded the time and duration of pump outages. Over the whole period, 599 repairs were completed on 169 pumps. The time between a breakdown being reported and repaired varied due to factors such as the complexity of the required repair, the geographical remoteness of the pump, and other logistical or operational constraints. The majority of repairs (90%) were completed within 3 days of a breakdown—the stated target of the service—and 44% were repaired within 24 hours of the fault being reported.

Blinding and Concealment

Efforts were made to ensure blinding and concealment when collecting the data. The enumerators were blinded to the intervention as, other than being aware that the mechanics were operating in the area as part of another University of Oxford–funded research project, they had no knowledge of which pumps had been repaired. While the households were certain not blinded to whether or not they had had a repair, it would not have been readily apparent to participants that the questions about their household members’ health were to be linked to the repairs, because the health questions were part of a large questionnaire and questions about the pump were intentionally asked in a later section of the survey than the health questions.
Mirroring the enumerators, the mechanics had very little knowledge of the household surveys. They were aware that the first household survey had taken place and that the second was occurring, and that it was broadly connected to their work, but they were not in any way involved in the survey. Treatment allocation—in this case, pump repair by the mechanics—could not be entirely randomized given that chance and other factors, measurable or otherwise, will have led to pump breakdown; however, the mechanics’ responses to reported breakdowns were unpredictable and not influenced by the enumerators or research team.

Inclusion Criteria
A total of 1451 households used pumps that were offered the free repair service. Out of these, 294 were excluded as our mechanics had not made a repair to their handpumps during the trial period. A further 182 were excluded as it was reported during the household survey that the reference handpump was not their primary water source. Finally, 135 were excluded as there were no children reported in the household, leaving a sample of 840 households served by 139 pumps. These pumps received 412 repairs, of which 41% were repaired within 1 day and 39% more within 2 days. The longest repair time was 11 days. In the self-controlled analysis, an additional 89 households were excluded because the families living in the dwelling had changed between the two surveys.

Statistical Analysis
Using the data from the second household survey, we first performed bivariate analysis to test the relationship between speed of pump repairs and self-reported diarrhea. We then used the household-matched data from the first survey to compare the period prevalence of diarrhea reported in 2013 against repair performance measured in 2015. This was to determine whether any relationship between the outcome and exposure was due to time-invariant characteristics of the household that were somehow correlated to the speed of repairs and not due to the repair service itself.

We then performed a cross-sectional analysis, fitting multivariable logistic regression models of self-reported diarrhea using the data from the 2015 household survey to adjust for covariates that have been shown to affect diarrhea period prevalence in the WASH-related literature. We fitted a series of multivariable regression models that iteratively adjust for an additional thematic block of covariates: (1) socioeconomic factors, (2) dwelling factors, and (3) WASH factors (see the Results for detailed factors). Additionally, we fitted a model with interaction terms to further explore results of the first three models, which was also guided by the WASH-related literature. Standard errors were adjusted for clustering using the three geographical zones to control for potentially unobserved/unquantified factors that are hypothesized to be correlated among households within each zone [42]. Because multiple households typically share a handpump, the handpump was another potential analytical unit for the clustering of standard errors between households. As a robustness check, we performed an additional analysis that clustered errors by pump instead of geographical zone.

All statistical analyses were undertaken with Stata version 14.2 (StataCorp, College Station, TX, USA).

Ethical Considerations
This research was conducted with permission from Kenya’s National Council for Science and Technology and with ethics approval from the University of Oxford’s Central University Research Ethics Committee (CUREC reference: SOGE C1 13-125). Informed consent was given by all research participants, all of whom were aged 18 years or over.

Results

Unadjusted Analysis
Bivariate logistic regression models did not provide evidence of any association between households reporting diarrhea and the number of repairs performed on their respective handpumps or the time elapsed since the most recent repair. However, this analysis did reveal an association between the speed of repairs (or pump downtime) and reported diarrhea (Table 1). Pumps consistently repaired within 24 hours upon a breakdown, irrespective of the number of repairs, were associated with a lower reported diarrhea rate. This effect disappeared when analyzed against longer repair times (Table 2).

The analysis comparing the same exposure of rapid repairs in 2015 but with the outcome of self-reported disease from the same household in 2013 (ie, when the repair service could not have had any influence) showed no effect. This suggests that the disease reduction was not due to unobserved time-invariant household characteristics that were somehow correlated with the speed of repairs. Rates of reported diarrhea were comparable between the two surveys: 8.2% in 2013 (dry season) and 9.8% in 2015 (wet season).

Table 1. Bivariate logistic regression analysis of child diarrhea with number and speed of repairs.

<table>
<thead>
<tr>
<th>Intervention/exposure</th>
<th>Unadjusted odds ratio (95% CI)</th>
<th>Analysis type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of repairs (2015)</td>
<td>0.95 (0.85-1.07)</td>
<td>Cross-sectional</td>
</tr>
<tr>
<td>Days since last repair (2015)</td>
<td>1.00 (1.00-1.00)</td>
<td>Cross-sectional</td>
</tr>
<tr>
<td>Pump repairs within 24 hours</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>0.40 (0.27-0.58)</td>
<td>Cross-sectional</td>
</tr>
<tr>
<td>2013</td>
<td>1.01 (0.44-2.29)</td>
<td>Self-controlled</td>
</tr>
</tbody>
</table>
Table 2. Relative risk of households reporting child diarrhea for different repair speeds.

<table>
<thead>
<tr>
<th>Intervention/exposure</th>
<th>Number of households</th>
<th>Households reporting diarrhea, n (%)</th>
<th>Rate ratio</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All pump repairs completed within 24 hours</td>
<td>123</td>
<td>7 (5.7)</td>
<td>reference</td>
<td>N/A</td>
</tr>
<tr>
<td>All repairs completed within 48 hours (with some over 24 hours)</td>
<td>171</td>
<td>23 (13.5)</td>
<td>2.36</td>
<td>.02</td>
</tr>
<tr>
<td>Repairs taking over 48 hours to complete</td>
<td>546</td>
<td>69 (12.6)</td>
<td>2.22</td>
<td>.02</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

Cross-Sectional Analysis

The cross-sectional analysis revealed possible predictors of reporting diarrhea. Table 3 shows the odds ratios for diarrhea against other household information gathered during the 2015 survey, along with summary statistics for those characteristics. The quality of dwelling construction was associated with lower reported diarrhea in children, as were having a female head of household and cultivating crops. Apparent risk factors were those related to higher household occupancy and having no regular income. The WASH-related factors of higher per-person water collection and having an improved toilet were apparently protective. Of note, 1.5 jerrycans of water were collected for each household member per day, corresponding to 30-35 liters per person, with few households treating water either by adding chlorine (eg, Waterguard) or by boiling.

Table 3. Bivariate logistic regression analysis of associations of household characteristics with reporting child diarrhea (12% of 840 households) from the survey conducted in March-May 2015.

<table>
<thead>
<tr>
<th>Household characteristic</th>
<th>Unadjusted odds ratio (95% CI)</th>
<th>P value</th>
<th>Percentage or mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of repairs</td>
<td>0.95 (0.85-1.07)</td>
<td>.42</td>
<td>2.7</td>
</tr>
<tr>
<td>Repairs within 24 hours</td>
<td>0.40 (0.27-0.58)</td>
<td>&lt;.001</td>
<td>15%</td>
</tr>
<tr>
<td>Improved floor</td>
<td>0.37 (0.22-0.61)</td>
<td>&lt;.001</td>
<td>35%</td>
</tr>
<tr>
<td>Improved roof</td>
<td>0.53 (0.31-0.89)</td>
<td>.02</td>
<td>34%</td>
</tr>
<tr>
<td>Improved walls</td>
<td>0.52 (0.39-0.71)</td>
<td>&lt;.001</td>
<td>46%</td>
</tr>
<tr>
<td>Improved toilet</td>
<td>0.56 (0.38-0.81)</td>
<td>.002</td>
<td>55%</td>
</tr>
<tr>
<td>Female head</td>
<td>0.65 (0.46-0.92)</td>
<td>.01</td>
<td>39%</td>
</tr>
<tr>
<td>Primary religion Islam</td>
<td>1.03 (0.71-1.49)</td>
<td>.14</td>
<td>85%</td>
</tr>
<tr>
<td>Household head completed secondary school</td>
<td>0.96 (0.56-1.63)</td>
<td>.87</td>
<td>30%</td>
</tr>
<tr>
<td>Number of people per household</td>
<td>1.11 (1.09-1.13)</td>
<td>&lt;.001</td>
<td>5.4</td>
</tr>
<tr>
<td>Sleeping rooms in dwelling</td>
<td>0.91 (0.87-0.95)</td>
<td>&lt;.001</td>
<td>3.1</td>
</tr>
<tr>
<td>Number of people per bedroom</td>
<td>1.27 (1.26-1.29)</td>
<td>&lt;.001</td>
<td>2.0</td>
</tr>
<tr>
<td>No regular income</td>
<td>1.41 (1.18-1.69)</td>
<td>&lt;.001</td>
<td>43%</td>
</tr>
<tr>
<td>Own animals</td>
<td>1.08 (0.57-2.06)</td>
<td>.81</td>
<td>35%</td>
</tr>
<tr>
<td>Grows crops</td>
<td>0.89 (0.67-1.18)</td>
<td>.42</td>
<td>74%</td>
</tr>
<tr>
<td>Distance from handpump (m)</td>
<td>1.00 (1.00-1.00)</td>
<td>.64</td>
<td>136</td>
</tr>
<tr>
<td>Distance from health dispensary (km)</td>
<td>0.96 (0.84-1.10)</td>
<td>.57</td>
<td>1.8</td>
</tr>
<tr>
<td>Distance from nearest market (km)</td>
<td>1.05 (1.03-1.06)</td>
<td>&lt;.001</td>
<td>5.0</td>
</tr>
<tr>
<td>Soap observed</td>
<td>0.65 (0.36-1.19)</td>
<td>.17</td>
<td>75%</td>
</tr>
<tr>
<td>Water treated</td>
<td>0.87 (0.38-1.98)</td>
<td>.73</td>
<td>10%</td>
</tr>
<tr>
<td>Water storage in the dry season (days)</td>
<td>1.10 (0.96-1.27)</td>
<td>.17</td>
<td>1.8</td>
</tr>
<tr>
<td>Water storage in the wet season (days)</td>
<td>1.02 (0.91-1.15)</td>
<td>.69</td>
<td>2.7</td>
</tr>
<tr>
<td>Number of jerrycans per person per day</td>
<td>0.68 (0.64-0.72)</td>
<td>.001</td>
<td>1.53</td>
</tr>
</tbody>
</table>

Multivariable Analysis of Self-Reported Diarrhea

Table 4 presents the adjusted odd ratios (AORs) and 95% CIs of iterative mixed-effects multivariable logistic regression models of child diarrhea: model 1 begins with socioeconomic factors, model 2 introduces a block of dwelling-related factors, and then model 3 introduces a block of WASH-related factors. The protective effect of handpump repairs consistently made within 24 hours remained robust across all multivariable models.
Of the socioeconomic factors, having a female head of household (AOR 0.54, 95% CI 0.36-0.82; \( P =.004 \)) and growing crops (AOR 0.67, 95% CI 0.48-0.95; \( P =.02 \)) were associated with lower reported diarrhea across all models. Having no income (AOR 1.36, 95% CI 1.11-1.69; \( P =.003 \)) and being further from the market (AOR 1.10, 95% CI 1.07-1.13; \( P <.001 \)) were associated with higher odds of child diarrhea. Having more people in the household was also associated with reporting diarrhea (AOR 1.12, 95% CI 1.04-1.21; \( P =.003 \)), although this was possibly an artifact of aggregating all cases of child diarrhea in the household into a binary variable; nevertheless, an alternate analysis with the number of children in the household produced a similar result. Among the WASH-related factors, having an improved toilet was associated with lower diarrhea (AOR 0.61, 95% CI 0.44-0.84; \( P =.002 \)). The effect of the dwelling characteristics was less consistent across the three models. As a robustness check, we also refitted these models with errors clustered by pump, rather than by geographical zone, but we observed no substantive differences from the results presented in Table 4.

We built on the full model (model 3) by introducing interaction terms between the following pairs of covariates: having no income and growing crops, a female household head and the presence of soap, and having an improved floor and improved toilet. These interactions are presented in Table 5. Having no regular income and not growing crops was the largest risk factor for reporting diarrhea. Strikingly, if a household was growing crops, it had no additional odds of reporting diarrhea if it had no regular income. Conversely, the benefit of having soap in a household seemed only to accrue to households headed by a woman, in which case the odds of reporting diarrhea reduced substantially. Finally, having an improved floor substantially reduced the odds of reporting diarrhea in households with an improved toilet, but provided no added benefit to those without.

### Table 4. Mixed-effects logistic regression models of self-reported diarrhea, adjusted for blocks of covariates, using 2015 household survey data and handpump repair data from 2014 and 2015.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Adjusted odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
</tr>
<tr>
<td>Repairs within 24 hours</td>
<td>0.38 (0.24-0.62)</td>
</tr>
<tr>
<td><strong>Socioeconomic factors</strong></td>
<td></td>
</tr>
<tr>
<td>No regular income</td>
<td>1.34 (1.12-1.60)</td>
</tr>
<tr>
<td>Grows crops</td>
<td>0.73 (0.50-1.07)</td>
</tr>
<tr>
<td>People per household</td>
<td>1.14 (1.12-1.26)</td>
</tr>
<tr>
<td>Female head</td>
<td>0.63 (0.45-0.89)</td>
</tr>
<tr>
<td>Distance from market (km)</td>
<td>1.06 (1.04-1.09)</td>
</tr>
<tr>
<td><strong>Dwelling factors</strong></td>
<td></td>
</tr>
<tr>
<td>Improved roof</td>
<td>—</td>
</tr>
<tr>
<td>Improved walls</td>
<td>—</td>
</tr>
<tr>
<td>Improved floor</td>
<td>—</td>
</tr>
<tr>
<td>Number of bedrooms</td>
<td>—</td>
</tr>
<tr>
<td><strong>WASH factors</strong></td>
<td></td>
</tr>
<tr>
<td>Improved toilet</td>
<td>—</td>
</tr>
<tr>
<td>Soap observed</td>
<td>—</td>
</tr>
<tr>
<td>Jerrycans per person per day</td>
<td>—</td>
</tr>
<tr>
<td>Water treated</td>
<td>—</td>
</tr>
</tbody>
</table>

\( ^a \) Not included in model.  
\( ^b \) WASH: Water, Sanitation, and Hygiene.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Odds ratio (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repairs within 24 hours</td>
<td>0.36 (0.24-0.55)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Distance from market (km)</td>
<td>1.09 (1.06-1.13)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Improved roof</td>
<td>0.73 (0.36-1.48)</td>
<td>.32</td>
</tr>
<tr>
<td>Improved walls</td>
<td>0.96 (0.69-1.33)</td>
<td>.92</td>
</tr>
<tr>
<td>People per household</td>
<td>1.11 (1.03-1.20)</td>
<td>.02</td>
</tr>
<tr>
<td>Number of bedrooms</td>
<td>1.06 (1.03-1.10)</td>
<td>.02</td>
</tr>
<tr>
<td>Jerry cans per person per day</td>
<td>0.85 (0.69-1.05)</td>
<td>.12</td>
</tr>
<tr>
<td>Water treated</td>
<td>1.01 (0.42-2.46)</td>
<td>.98</td>
</tr>
<tr>
<td><strong>Interactions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income and crops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular income and grows crops</td>
<td>1.00 (0.69-1.45)</td>
<td>.86</td>
</tr>
<tr>
<td>No regular income and grows crops</td>
<td>1.10 (0.68-1.79)</td>
<td>.63</td>
</tr>
<tr>
<td>No regular income and does NOT grow crops</td>
<td>2.30 (1.85-2.86)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Household head and soap</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male head and soap observed</td>
<td>0.79 (0.43-1.44)</td>
<td>.40</td>
</tr>
<tr>
<td>Female head and soap NOT observed</td>
<td>0.89 (0.62-1.30)</td>
<td>.41</td>
</tr>
<tr>
<td>Female head and soap observed</td>
<td>0.32 (0.26-0.38)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Standard of floor and toilet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved floor and NO improved toilet</td>
<td>0.80 (0.33-1.91)</td>
<td>.51</td>
</tr>
<tr>
<td>Unimproved floor and improved toilet</td>
<td>0.72 (0.53-0.98)</td>
<td>.03</td>
</tr>
<tr>
<td>Improved floor and improved toilet</td>
<td>0.29 (0.13-0.65)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

This study revealed an association between the speed of pump repairs and child diarrhea morbidity. Only households whose pumps were consistently repaired within 24 hours of a failure reported a significant reduction in child diarrhea. This finding is consistent with modeling work that suggests that even the briefest time without high-quality water can lead to disproportionate health impacts [38,39], but the microbiological pathway very likely does not fully explain the higher likelihood of self-reported illness for pumps that were repaired less rapidly. The lag time between pump failure and survey completion varied considerably, ranging from the day before the survey took place to pumps that were repaired when the maintenance service started over 1 year before the survey took place and had remained functional ever since. In these latter cases, direct microbiological causality between pump downtime and illness is unlikely.

In lieu of the presence of a consistent microbiological pathway, we would propose an additional behavioral explanation. If a household has higher confidence in the reliability of their pump, knowing that mechanical failures will be consistently fixed within 24 hours, they may be less likely to immediately fall back on using lower-quality alternative water sources as they might in the case of an extended pump downtime. If households experience any variability in the speed of pump repairs, then they are more likely to collect, store, and use water from less wholesome alternative sources. These are unlikely to be discarded once the pump is fixed, because there was a time or financial cost in acquiring them, thus raising the potential for ongoing waterborne disease exposure.

In addition to providing evidence that only rapidly repaired handpumps lead to a reduction in child diarrhea, this study highlights other factors that contribute to diarrheal disease in rural households reliant on basic water services. Striking, but unsurprising, are the socioeconomic effects, consistent with the WASH literature: the poorer you are, the more likely you are to experience water insecurity and consequently an increased burden of water-related diseases. Growing crops mitigated the impact of having no income; however, we do not have sufficient information from this study to attribute this to the nutritional benefit of growing food or because it generated occasional income or barter opportunities, either being plausible.

Individual components of traditional WASH interventions and programs did not seem to provide a beneficial health effect when considered in isolation. Our use of interaction terms (Table 5) suggested that soap being available and observed at a home was
only associated with a reduction in diarrhea in households headed by women, suggesting that hygiene behaviors in children may be better taught or enforced in households headed by women than by men. Similarly, the benefits of having an improved toilet were magnified by having an improved floor (ie, one made of cement or concrete). Such floors are easier to clean, reducing the likelihood of having human and animal feces present in homes. These empirical data are consistent with conceptual models that articulate a nonlinear relationship between the reduction in fecal contamination and consequent reduction in diarrhea [43,44].

The adaptation strategy of collecting and drinking more unsafe water in response to pump unreliability increases the risk of water-related illnesses. However, this strategy is one that households will rationally stick to unless there is compelling and enduring evidence that it is not necessary. This is consistent with studies on household preferences for rapid pump maintenance [31,34], and was supported anecdotally by conversations with people who were receiving the free repair service. However, even the well-resourced maintenance service linked to this study only fixed 15% of pumps within 24 hours every time they broke down. Acknowledging the possible selection bias in this study, community-managed handpumps are even less likely to be consistently repaired within 24 hours. This is not to discount the many other benefits from reducing pump downtimes, it is simply that health cannot be assumed to be one of them.

From an operational perspective, any pump maintenance service that reduces downtimes by an order of magnitude would be viewed as excellent in the rural water sector or even a best-in-class option. This study demonstrated an association between consistent, rapid handpump repairs and reduced child diarrhea reported in the households that use these pumps, suggesting that such a high level of service may also lead to better health outcomes. Any WASH intervention that achieves a reduction in diarrhea of over 50% would certainly be viewed as a success. If professionalizing handpump operations and maintenance to minimize pump downtimes can yield the health impacts demonstrated in this small study, can it have a substantial impact on the burden of disease?

Only the pumps with the lowest downtimes were associated with this level of health benefit. The 60% reduction in diarrhea was only the best possible outcome for a best-in-class service. The households using handpumps that had received very rapid repair corresponded to only the 15% of pumps where all repairs were completed within 24 hours. An efficacy of 60% translated into an effectiveness in reducing diarrhea morbidity of less than 10% for all the households receiving the repair service. To produce a health benefit, continuity of service must be maintained with near-zero downtime. This level of service is extremely difficult for professional service providers to maintain, and there is scant evidence that community maintenance can consistently achieve these performance levels [24,25,45,46]. FundiFix now operates a hybrid system of reactive and preventative maintenance to maintain pump uptime. Work is underway to use the sensors on the handpumps to monitor the pumps’ condition and predict failures with the aim of eliminating pump downtime altogether [40,47].

Limitations
This study’s primary outcome variable was self-reported diarrhea, with enumerators trained to ask the question according to the WHO definition. Self-reporting is known to have limitations, and diarrhea is a crude proxy for water-related diseases [48,49]. The use of self-reported diarrhea in this study does not in itself call into question the primary finding that consistently rapid repairs are paramount to the delivery of health benefits, but it does mean that caution should be used in interpreting the effect size measured here. As the cross-sectional element of the study was based around one survey, the primary outcome variable was measured at times that varied with respect to the primary exposure. This had advantages for blinding and concealment, but limits the ability to attribute any observed effect to a specific breakdown event. We can only speculate on the extent to which the mechanism for the observed effect is directly biological or behavioral.

Household matching between the two surveys was undertaken rigorously, with unmatched households excluded from analysis where appropriate. The two surveys were 18 months apart and the household characteristics were not necessarily identical during each survey wave. Household composition may have changed with births, deaths, or others joining or leaving the household; household water budgets may also have changed as members transitioned from child to adult roles, or from being breastfed to directly consuming food and water, and other factors.

The exclusion criteria used for this study necessarily restricted the survey to households confirmed to have used pumps that had received repairs from the free maintenance service. The average downtime for handpumps across the study area prior to the trial was 27 days. It is plausible that the communities that chose not to engage with the repair service were those with better repair arrangements and performance prior to the study, introducing a possible selection bias. In addition, the method of household selection was likely to have created a bias toward households living closer to the pump (the average distance from a dwelling to its reference handpump was 136 meters and the maximum was 739 meters). These issues do not affect the internal validity of the study but may reduce external validity and the wider inferences that can be made.

Most of these limitations were the result of using an observational study design, which was linked to establishing the local pump maintenance provider. While the findings may be less widely generalizable, opportunistic operational studies linked to local WASH programs—whether by governments, private operators, or development agencies—have certain benefits. Here, the findings are directly informing the service provider’s operational management of handpumps in this area. By operating within existing and ongoing local programs in this way, findings from such studies can feed directly into local decision-making, and are thus more likely to have an immediate and sustainable impact on local health outcomes [50]. In this case, these findings—already shared with the maintenance service provider—have underscored the importance of reducing pump downtimes to a minimum.
Conclusions

Implicit in the JMP’s definition of “basic water” is the assumption that improved sources of water are of higher quality than unimproved sources and would therefore be associated with the health benefits of lower diarrheal disease morbidity. Our findings suggest that handpumps may only produce health benefits if service delivery can achieve the highest levels of performance, a level of performance difficult to maintain even under ideal operational circumstances. These findings provide empirical support to prior modeling of the relationship among water service reliability, water quality, and health, which suggests that even short periods of supply intermittency may have large adverse health impacts [38,39].

The required consistent reduction in downtimes needed is a challenge for even FundiFix, a well-resourced and tightly managed service provider that achieves higher uptimes than are typical in the sector. Our findings have implications beyond handpumps and are potentially relevant to other forms of basic water service. For these, the management arrangements are often limited, back-up sources are most likely from an unimproved source, and similar household choices will have to be made in response to an unreliable supply. This underscores the need to pursue on-premises, continuous supply.

These findings are also consistent with a rich literature linking water-related diseases to poverty. We found that socioeconomic factors were more closely linked to household disease than some classic WASH factors, which taken in isolation did not reduce the incidence of child diarrhea. This implies that WASH interventions will be more effective when integrated into wider anti-poverty and service delivery efforts, rather than implemented in isolation. Both of these findings support the current direction of the WASH and Water Security research agenda, which considers the key role water plays in other fields such as education and nutrition, and emphasizes the wider systems of service delivery, governance, and rights of which WASH interventions are a part [51,52].

Acknowledgments

The authors thank Said Banje, Tim Foster, Nancy Gladstone, Susie Goodall, Rob Hope, Alan Howard, Jacob Katuva, Johanna Koehler, Idd Mwaropia, James Okoti, and Amber Pearson for their support, as well as Rural Focus Ltd and the Kwale County Government. This work was supported by the following grants: “New Mobile Citizens and Waterpoint Sustainability in Rural Africa” (ESRC, ES/JO18120/1) and “Groundwater Risk Management for Growth and Development” (NERC, NE/M008894/1), funded by the UK Economic and Social Research Council (ESRC) and the UK Natural Environment Research Council (NERC)/UK Department for International Development, respectively, along with the Household Water Insecurity Experiences (HWISE) Research Coordination Network, funded by the US National Science Foundation (BCS-1759972).

Data Availability

The full data analyzed for this study are not publicly available in order to maintain the confidentiality and anonymity of the participants. The publicly available data, which were generated during UK Research and Innovation–funded projects ES/JO18120/1 and NE/M008894/1, are available in the UK Data Service repository [53].

Authors' Contributions

PT and DJB developed the study concept and design. PT, JS, and MB analyzed the data. PT wrote the manuscript in consultation with JS and DJB. All authors discussed the results and contributed to and agreed on the final manuscript.

Conflicts of Interest

PT is an unpaid trustee of FundiFix Water Services Trust, the organization spun out from this research project of which this study is a part. The other authors have no connection to FundiFix and no conflicts to declare.

References

26. Carter R, Harvey E, Casey V. User financing of rural handpump water services. 2010 Presented at: Pumps, Pipes and Promises: Costs, Finances and Accountability for Sustainable WASH Services; November 16-18, 2010; The Hague, the


Abbreviations

AOR: adjusted odds ratio
DALYs: disability-adjusted life years
JMP: Joint Monitoring Programme
SDG: Sustainable Development Goal
UNICEF: United Nations Children's Fund
WASH: Water, Sanitation, and Hygiene
WHO: World Health Organization

Edited by A Mavragani; submitted 08.09.22; peer-reviewed by A Krishnan, X Ma, KK Yadav, P Chintalapati; comments to author 25.08.23; revised version received 23.10.23; accepted 14.12.23; published 16.01.24.

Please cite as:
Thomson P, Stoler J, Byford M, Bradley DJ
The Impact of Rapid Handpump Repairs on Diarrhea Morbidity in Children: Cross-Sectional Study in Kwale County, Kenya
JMIR Public Health Surveill 2024;10:e42462
URL: https://publichealth.jmir.org/2024/1/e42462
doi:10.2196/42462
PMID:38227359

©Patrick Thomson, Justin Stoler, Michelle Byford, David J Bradley. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 16.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Latent Heterogeneity of Online Sexual Experiences and Associations With Sexual Risk Behaviors and Behavioral Health Outcomes in Chinese Young Adults: Cross-Sectional Study

Ted C T Fong, PhD; Derek Yee Tak Cheung, PhD; Edmond Pui Hang Choi, PhD; Daniel Y T Fong, PhD; Rainbow T H Ho, PhD; Patrick Ip, MBBS; Man Chun Kung, MSc; Mona Wai Cheung Lam, MBChB; Antoinette Marie Lee, PhD; William Chi Wai Wong, MD; Tai Hing Lam, MD; Paul S F Yip, PhD

Centre on Behavioral Health, Faculty of Social Sciences, The University of Hong Kong, Hong Kong, China (Hong Kong)
School of Nursing, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
Department of Social Work & Social Administration, Faculty of Social Sciences, The University of Hong Kong, Hong Kong, China (Hong Kong)
Department of Paediatrics & Adolescent Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
Family Planning Association of Hong Kong, Hong Kong, China (Hong Kong)
Department of Psychology, Faculty of Social Sciences, The University of Hong Kong, Hong Kong, China (Hong Kong)
Department of Family Medicine & Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
Centre for Suicide Research and Prevention, Faculty of Social Sciences, The University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Pual S F Yip, PhD
Centre for Suicide Research and Prevention
Faculty of Social Sciences
The University of Hong Kong
2/F, The HKJC Building for Interdisciplinary Research
5 Sassoon Road, Pokfulam
Hong Kong, 852
China (Hong Kong)
Phone: 852 2831 5232
Fax: 852 2549 7161
Email: sfp@hku.hk

Abstract

Background: Online sexual experiences (OSEs) are becoming increasingly common in young adults, but existing papers have reported only on specific types of OSEs and have not shown the heterogeneous nature of the repertoire of OSEs. The use patterns of OSEs remain unclear, and the relationships of OSEs with sexual risk behaviors and behavioral health outcomes have not been evaluated.

Objective: This study aimed to examine the latent heterogeneity of OSEs in young adults and the associations with sexual risk behaviors and behavioral health outcomes.

Methods: The 2021 Youth Sexuality Study of the Hong Kong Family Planning Association phone interviewed a random sample of 1205 young adults in Hong Kong in 2022 (male sex: 613/1205, 50.9%; mean age 23.0 years, SD 2.86 years) on lifetime OSEs, demographic and family characteristics, Patient Health Questionnaire-4 (PHQ-4) scores, sex-related factors (sexual orientation, sex knowledge, and sexual risk behaviors), and behavioral health outcomes (sexually transmitted infections [STIs], drug use, and suicidal ideation) in the past year. Sample heterogeneity of OSEs was analyzed via latent class analysis with substantive checking of the class profiles. Structural equation modeling was used to examine the direct and indirect associations between the OSE class and behavioral health outcomes via sexual risk behaviors and PHQ-4 scores.

Results: The data supported 3 latent classes of OSEs with measurement invariance by sex. In this study, 33.1% (398/1205), 56.0% (675/1205), and 10.9% (132/1205) of the sample were in the abstinent class (minimal OSEs), normative class (occasional OSEs), and active class (substantive OSEs), respectively. Male participants showed a lower prevalence of the abstinent class (131/613, 21.4% versus 263/592, 44.4%) and a higher prevalence of the active class (104/613, 17.0% versus 28/592, 4.7%) than...
female participants. The normative class showed significantly higher sex knowledge than the other 2 classes. The active class was associated with male sex, nonheterosexual status, higher sex desire and PHQ-4 scores, and more sexual risk behaviors than the other 2 classes. Compared with the nonactive (abstent and normative) classes, the active class was indirectly associated with higher rates of STIs (absolute difference in percentage points ($\Delta$)=4.8%; $P=0.03$) and drug use ($\Delta$=7.6%; $P=0.001$) via sexual risk behaviors, and with higher rates of suicidal ideation ($\Delta$=2.5%; $P=0.007$) via PHQ-4 scores.

Conclusions: This study provided the first results on the 3 (abstent, normative, and active) latent classes of OSEs with distinct profiles in OSEs, demographic and family characteristics, PHQ-4 scores, sex-related factors, and behavioral health outcomes. The active class showed indirect associations with higher rates of STIs and drug use via sexual risk behaviors and higher rates of suicidal ideation via PHQ-4 scores than the other 2 classes. These results have implications for the formulation and evaluation of targeted interventions to help young adults.

(JMIR Public Health Surveill 2024;10:e50020) doi:10.2196/50020

KEYWORDS
Hong Kong; latent class analysis; mediation; mental health; sex knowledge; sexual risk behaviors; sexually transmitted infections; structural equation modeling; youth sexuality

Introduction
Digital technology has transformed the communication styles and lives of young generations who have grown up with the internet. During emerging adulthood, individuals experience intense developments in the physical, cognitive, and psychosocial domains [1], which include greater interactions with the social environment, self-identity development, and sexuality exploration [2]. Emerging adults typically establish their sexual identity and patterns of health and risk behaviors, which have health implications for later life. The internet and social media are essential for forming social relationships among these adults.

Popular culture in Western countries has embraced uncommitted sex, such as casual sexual relationships and experiences, among young adults [3]. Online sexual experiences (OSEs) involve a range of sex-related behaviors performed via the internet or social media, including accessing sexual information, exposure to graphic nudity, watching pornography, online dating, sexting, and naked chatting [4-6]. A study in Spain found a rising prevalence of sexting among high school students [7], and pornography use has been shown to be associated with sexually aggressive behaviors, such as sexual coercion, among American adults [8]. In the past few years, the COVID-19 pandemic has further increased OSEs as a result of decreased interpersonal physical intimacy due to social distancing measures [9].

In Chinese culture, sexuality is viewed more conservatively, and frequent parent-adolescent communication about sexuality has been shown to be associated with fewer sexual risk behaviors in young adults in mainland China [10]. However, a qualitative study found that family communication about sexuality was rare and implicit in mainland China [11]. A mixed-method study found that 87.3% of young adults in mainland China obtained sexual health information online [12]. In Hong Kong, a previous study found that one-tenth of young adults engaged in sexual risk behaviors [13]. Exposure to sexually explicit online materials was associated with subsequent psychological well-being and sexual permissive attitudes in Hong Kong adolescents [14]. Existing papers [4,7,8,12,14] only assessed specific subsets of OSEs, such as accessing sexual information, sexting, and watching pornography. The failure to acknowledge the diverse nature of OSEs implies that more research is necessary to investigate an inclusive list of OSEs.

Latent class analysis (LCA) is a statistical method that identifies latent classes based on respondents’ responses and elucidates diverse profiles of behavioral indicators [15]. Instead of using aggregated item averages, LCA can model the complexity and individual differences of OSEs. A literature search on Web of Science using the keywords of (“sexual experience*” OR “sexual behavior*” OR “sexual activit*” [title]) AND latent class [title]) on April 6, 2023, found only 2 studies that explored OSEs using LCA, with small sample sizes (N=269-322) and analyses limited to female adolescents. These studies found 4 latent classes of OSEs, which were associated with sexual health and substance use [16,17]. However, they did not include young adults or males. A recent LCA study investigated sexual behavior indicators among youth in mainland China [18], but no existing papers have examined the latent heterogeneity of OSEs in young adults.

Given the importance of online sexual behaviors in health sciences [19], it is necessary to better understand the prevalence and profiles of OSEs among young adults. Multiple sex partners and unsafe sex have been shown to be associated with alcohol and tobacco use in Chinese youth [18] and American young adults [20], with sexual risk behaviors mediating sexual health outcomes [21,22]. However, no papers have examined the direct and indirect relationships between OSEs and behavioral health outcomes, namely, sexually transmitted infections (STIs), drug use, and suicidal ideation, via sexual risk behaviors and mental health. This study had the following 3 objectives: (1) to investigate the prevalence and heterogeneity of OSEs in young adults in Hong Kong; (2) to validate the derived OSE classes and analyze their associations with external variables, including demographic, family, psychological, and sex-related characteristics; and (3) to examine the direct and indirect relationships between OSE classes and behavioral health outcomes via sexual risk behaviors and mental health.
Methods

Study Design and Research Procedures

The data for this study originated from the 2021 Youth Sexuality Study (YSS) of the Hong Kong Family Planning Association, which was conducted every 5 years to examine sexual health and behaviors, and the psychosocial health of young adults. The sampling methods are described in the 2021 YSS survey report [23]. The survey was delayed due to the COVID-19 pandemic and was conducted via mobile phone from March to June 2022. The inclusion criteria were as follows: living in Hong Kong, age 18 to 27 years, and ability to understand spoken Cantonese or Putonghua. The survey was conducted by Social Policy Research Limited, a well-known third-party survey agency in Hong Kong. This agency has conducted a number of surveys on similar sensitive topics, such as organ donation, COVID-19 health-related risks, and fertility decisions, among various stakeholders in the general public. The agency has ample experience in undertaking studies of a similar nature with considerable expertise in survey planning and implementation, and data security. A total of 22,000 phone numbers were randomly generated, and a computer-assisted telephone interviewing system was used to contact potential respondents on weekdays from 6:30 PM to 10:30 PM.

The purpose and confidentiality of the survey were clearly explained to all potential participants, and voluntary participants provided oral informed consent. After excluding 1909 invalid telephone numbers and 17,253 numbers without target respondents, 2838 participants were eligible for the study. Some of the eligible participants did not take part in the phone interviews because they were too busy and could not make another survey appointment. In total, 1205 phone interviews were completed, with a response rate of 42.5%. Trained interviewers who were fluent in Cantonese and Putonghua conducted the interviews in around 20 minutes and provided emotional support information at the end. The survey did not collect personal identifying information and used anonymous questionnaires to encourage honest answers to sensitive questions, such as OSEs, sexual orientation, and sexual risk behaviors. The phone interviews were conducted in a private and secure setting. The interviewers used nonjudgmental and nonthreatening language to help participants feel comfortable and safe. Half of the interviews underwent quality checks by independent checkers, and 93.8% (563/600) of the sampled cases were successfully verified.

Measures

The study questionnaire was designed by the 2021 YSS Task Force based on previous YSS questionnaires. It included questions on OSEs, demographic and family characteristics, mental health, sex-related factors, and behavioral health outcomes. It was field tested in the second half of 2021 and underwent minor revisions based on feedback from respondents and interviewers.

OSEs were assessed using 11 binary questions (yes/no) on lifetime experiences of online sex-related behaviors (Multimedia Appendix 1). The first 5 items were from the 2016 YSS, and the latter 6 were new items added to the 2021 YSS based on recent literature [7,24]. The 11 items referred to 3 types of OSEs [25]: nonarousal, solitary arousal, and partnered arousal. These items had good reliability (Cronbach α=.83) and satisfactory item-total correlations (r=0.40-0.62; P<.001) in the present sample.

Demographic characteristics assessed in the questionnaire included sex, age, education level, student status (versus working), and place of birth (Hong Kong or other places). Family characteristics included marital status (married versus single) and perceived family satisfaction via a single item “Are you happy with your family life?” on a 5-point Likert scale ranging from 1 (“very unhappy”) to 5 (“very happy”). Participants also reported the number of hours spent per week on social media.

The Patient Health Questionnaire-4 (PHQ-4) measured 4 anxiety and depression symptoms experienced by the participants over the past 2 weeks. The items were rated on a 4-point Likert scale ranging from 0 (“not at all”) to 3 (“nearly every day”), and the composite PHQ-4 score ranged from 0 to 12. The PHQ-4 has been validated as a brief assessment of psychological distress in Hong Kong young adults [26] and had good reliability (α=.86) in the present sample. Participants were asked about the perceived negative impact of COVID-19 on mental health on a 5-point Likert scale ranging from 1 (“much better”) to 5 (“much poorer”).

Questions on sex-related factors included sexual orientation (heterosexual=0 and nonheterosexual=1); sex desire (“How strong is your desire to have sex with others?” on a 5-point Likert scale ranging from 1 “[not at all]” to 5 “[very strong]”); and 12 knowledge statements on sex-related behaviors, HIV, and AIDS (response options of “true” or “false,” or “don’t know”). Each correct answer scored 1 point, and higher total scores (range: 0-12) indicated better sexual knowledge. The questionnaire assessed lifetime experiences of 5 sexual risk behaviors: sexually harassed others, sexual coercion (insisted on having sex while ignoring a partner’s wish), compensated dating with sex, unsafe sex without contraception, and multiple sex partners. The items on sexual knowledge and sexual risk behaviors showed good reliability (α=.74-.83) in the present sample.

The study assessed 3 behavioral health outcomes, namely, STIs, drug use, and suicidal ideation, of the participants in the past year. Drug use referred to the use of various psychotropic substances such as ketamine, ecstasy, cocaine, and heroin.

Statistical Analysis

The variables were summarized using descriptive statistics, and the reliabilities of OSEs, PHQ-4 scores, sexual knowledge, and sexual risk behaviors were assessed via Cronbach α coefficients. Sampling weights were calculated based on 2021 population census data stratified by sex, place of birth, and age group to correct for the imbalance in the sample. Weighted analyses were used for all subsequent analyses unless otherwise stated. Missing data were handled using full information maximum likelihood under the missing-at-random assumption [27]. LCA was used to identify latent classes based on the 11 OSE items among young adults. LCA models with 1 to 5 classes were estimated
using the robust maximum likelihood estimator in Mplus 8.6 [28]. Model fit was evaluated using the Bayesian information criterion (BIC), with lower values indicating better fit. The Lo-Mendell-Rubin (LMR) likelihood ratio test [29] compared the fit of the k-class LCA model to the alternative k-1 class model, with a small P-value of <.01 favoring the former. Model classification quality was assessed using entropy and average latent class probabilities, with high values (>0.90) indicating adequate classification.

The heterogeneity of the OSEs in the sample was described via prevalence and conditional item probabilities of latent classes, with probabilities above 0.40 considered substantial. LCA models were estimated separately for males and females to test the stability of the latent class structure by sex. Measurement invariance was evaluated by comparing the BIC of the LCA models with and without equality constraints on the OSE item thresholds. Latent classes were checked against demographic and family characteristics, mental health, sex-related factors, and behavioral health outcomes. Class differences were estimated using the Bolck, Croon, and Hagenaars (BCH) procedure [30], with post-hoc comparisons using Sidak correction. Sidak correction controlled the family-wise error rate with slightly higher power than the Bonferroni correction method. Multinomial logistic regression with a 3-step approach used demographic and family characteristics and sex-related factors as predictors of latent class membership, with the strengths of associations estimated by adjusted odds ratios (ORs).

We used a structural equation model (SEM) to investigate the associations among OSE latent classes, sexual risk behaviors, PHQ-4 scores, and behavioral health outcomes. The most likely OSE class membership was the primary independent variable; sexual risk behaviors and PHQ-4 scores were latent mediators; and STIs, drug use, and suicidal ideation were binary outcomes. Model fit was evaluated by the following criteria on fit indices: root mean square error of approximation (RMSEA) ≤0.06, comparative fit index (CFI) ≥0.95, and standardized root mean square residual (SRMR) ≤0.06. The SEM estimated standardized regression coefficients for direct (β) and indirect effects (αβ) of the OSE latent class on behavioral health outcomes via the mediators. The SEM included covariates, namely, sex, age, place of birth, education level, student status, marital status, family satisfaction, sexual orientation, sexual knowledge, COVID-19 mental impact, and time spent on social media. The direct and indirect effects were estimated using bootstrapping with 10,000 bootstrap draws. Statistical significance was determined by 95% CIs not including zero for the direct and indirect effects, and 95% CIs not including one for the ORs.

**Ethical Considerations**

The study received ethical approval from the Human Research Ethics Committee of the University of Hong Kong (reference number: EA200333). All participants provided oral informed consent after the survey’s purpose was explained, with anonymity, confidentiality, and voluntary participation emphasized. The privacy of personal information was protected throughout the study via anonymous data collection, and confidentiality was preserved to ask participants to provide honest answers. Eligible participation in this survey was voluntary and was not compensated. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

**Results**

**Sample Characteristics**

Table 1 shows that among 1205 participants (613 males and 592 females), the average age was 23.0 (SD 2.86) years. The majority of them were born in Hong Kong, working, single, and heterosexual. They had good sexual knowledge (mean 10.1, SD 2.19) and spent a mean of 2.94 (SD 2.04) hours per day on social media. The prevalence of sexual risk behaviors and behavioral health outcomes ranged from 3.2% (38/1205) to 12.3% (148/1205) and from 1.7% (20/1205) to 6.5% (78/1205), respectively. As shown in Multimedia Appendix 1, more than half (616/1205, 51.1% to 801/1205, 66.5%) of the participants were exposed to pornographic content, accessed sexual content, and discussed sex with others online or on social media. Naked chatting online, sending pornographic messages online, and having sex with people acquainted with online or on social media had the lowest prevalence (77/1205, 6.4% to 134/1205, 11.1%). Male participants showed higher prevalence of all 11 OSEs ($\chi^2 = 20.6-55.7; P < .001$) than female participants.
Table 1. Demographic and family characteristics, mental health data, sex-related factors, and behavioral health outcomes.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value (N=1205)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Female sex, n (%)</strong></td>
<td>592 (49.1)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Secondary school</td>
<td>389 (32.3)</td>
</tr>
<tr>
<td>Associate degree</td>
<td>266 (22.1)</td>
</tr>
<tr>
<td>Bachelor’s degree or above</td>
<td>551 (45.7)</td>
</tr>
<tr>
<td><strong>Current status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>478 (39.7)</td>
</tr>
<tr>
<td>Working</td>
<td>727 (60.3)</td>
</tr>
<tr>
<td><strong>Born in Hong Kong, n (%)</strong></td>
<td>894 (74.2)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>149 (12.4)</td>
</tr>
<tr>
<td>Single</td>
<td>1056 (87.6)</td>
</tr>
<tr>
<td><strong>Sexual status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>1117 (92.7)</td>
</tr>
<tr>
<td>Nonheterosexual</td>
<td>88 (7.3)</td>
</tr>
<tr>
<td><strong>Sexual risk behaviors, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Sexually harassed others</td>
<td>61 (5.1)</td>
</tr>
<tr>
<td>Sexual coercion</td>
<td>66 (5.5)</td>
</tr>
<tr>
<td>Compensated dating with sex</td>
<td>38 (3.2)</td>
</tr>
<tr>
<td>Unsafe sex</td>
<td>148 (12.3)</td>
</tr>
<tr>
<td>Multiple sex partners</td>
<td>82 (6.8)</td>
</tr>
<tr>
<td><strong>Behavioral health outcomes, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Sexually transmitted infections</td>
<td>20 (1.7)</td>
</tr>
<tr>
<td>Drug use</td>
<td>40 (3.3)</td>
</tr>
<tr>
<td>Suicidal ideation</td>
<td>78 (6.5)</td>
</tr>
<tr>
<td>Age (years), mean (SD; range)</td>
<td>23.0 (2.86; 18-27)</td>
</tr>
<tr>
<td>Family satisfaction(^a), mean (SD; range)</td>
<td>3.60 (0.79; 1-5)</td>
</tr>
<tr>
<td>Daily hours spent on social media, mean (SD; range)</td>
<td>2.94 (2.04; 0-8)</td>
</tr>
<tr>
<td>COVID-19 mental impact(^a), mean (SD; range)</td>
<td>3.24 (0.86; 1-5)</td>
</tr>
<tr>
<td>Psychological distress(^a) (PHQ-4(^b)), mean (SD; range)</td>
<td>2.17 (2.34; 0-12)</td>
</tr>
<tr>
<td>Sex desire(^a), mean (SD; range)</td>
<td>3.14 (0.93; 1-5)</td>
</tr>
<tr>
<td>Sexual knowledge(^a), mean (SD; range)</td>
<td>10.1 (2.19; 0-12)</td>
</tr>
</tbody>
</table>

\(^a\)Higher scores indicate better family satisfaction, more negative COVID-19 mental impact, and higher levels of psychological distress, sex desire, and sexual knowledge.

\(^b\)PHQ-4: Patient Health Questionnaire-4.

**Latent Class Models**

Multimedia Appendix 2 shows a decreasing trend in the BIC from 1-class to 5-class LCA models in the whole sample and in males and females separately, with substantial decreases in the BIC from 1-class to 3-class models. The LMR likelihood ratio tests did not show significant improvement (\(P=.10-.38\)) in model fit for the 4-class model over the 3-class model. These findings supported the 3-class LCA model with good entropy and average latent class probabilities. Multimedia Appendices 3 and 4 show the class prevalence and conditional item probabilities of the 3-class LCA model in males and females, respectively. The multiple-group LCA model with sex-invariant item thresholds showed a substantially lower BIC (12,792 versus 12,925) than the noninvariant LCA model, supporting measurement invariance of the latent class structure across sex.
As shown in Table 2, the abstinent class comprised 33.1% (398/1205) of participants and had the lowest conditional item probabilities (1/398, 0.3% to 46/398, 11.5%) of all OSEs. The normative class, comprising 56.0% (675/1205) of participants, accessed sexual content and pornographic content online or on social media, discussed sex with others on social media, and sought pornographic content online actively. This class did not show substantial prevalence (<40%) in the remaining 7 items. The active class, comprising 10.9% (132/1205) of participants, had the highest conditional item probabilities (65/132, 49.5% to 132/132, 100%) of all OSEs. The prevalences of the normative and active classes were higher among males (378/613, 61.6% and 104/613, 17.0%, respectively) than among females (301/592, 50.9% and 28/592, 4.7%, respectively).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Class 1: abstinent (N=398)</th>
<th>Class 2: normative (N=675)</th>
<th>Class 3: active (N=132)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole sample (N=1205)</td>
<td>398 (33.1)</td>
<td>675 (56.0)</td>
<td>132 (10.9)</td>
</tr>
<tr>
<td>Males (N=613)</td>
<td>131 (21.4)</td>
<td>378 (61.6)</td>
<td>104 (17.0)</td>
</tr>
<tr>
<td>Females (N=592)</td>
<td>263 (44.4)</td>
<td>301 (50.9)</td>
<td>28 (4.7)</td>
</tr>
<tr>
<td>Item(^a), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Exposed to pornographic content online or on social media</td>
<td>30 (7.6)</td>
<td>637 (94.4)(^b)</td>
<td>131 (99.2)(^b)</td>
</tr>
<tr>
<td>6. Accessed sexual content online or on social media</td>
<td>15 (3.7)</td>
<td>558 (82.7)(^b)</td>
<td>132 (100.0)(^b)</td>
</tr>
<tr>
<td>1. Discussed sex with others on social media</td>
<td>46 (11.5)</td>
<td>441 (65.3)(^b)</td>
<td>126 (95.1)(^b)</td>
</tr>
<tr>
<td>9. Actively sought pornographic content online or on social media</td>
<td>12 (2.9)</td>
<td>328 (48.6)(^b)</td>
<td>124 (94.0)(^b)</td>
</tr>
<tr>
<td>3. Dated people acquainted with online or on social media</td>
<td>28 (7.1)</td>
<td>229 (34.0)</td>
<td>119 (89.9)(^b)</td>
</tr>
<tr>
<td>11. Exposed to pornographic content in internet games</td>
<td>10 (2.6)</td>
<td>227 (33.6)</td>
<td>103 (77.7)(^b)</td>
</tr>
<tr>
<td>4. Received pornographic (text or video) messages online</td>
<td>7 (1.7)</td>
<td>186 (27.6)</td>
<td>81 (61.4)(^b)</td>
</tr>
<tr>
<td>10. Posted/shared indecent photos online or on social media</td>
<td>2 (0.4)</td>
<td>103 (15.2)</td>
<td>100 (76.1)(^b)</td>
</tr>
<tr>
<td>7. Had sex with people acquainted with online or on social media</td>
<td>2 (0.5)</td>
<td>45 (6.7)</td>
<td>86 (65.3)(^b)</td>
</tr>
<tr>
<td>5. Sent pornographic (text or video) messages online</td>
<td>1 (0.3)</td>
<td>53 (7.9)</td>
<td>66 (49.9)(^b)</td>
</tr>
<tr>
<td>8. Had naked chat online</td>
<td>2 (0.4)</td>
<td>9 (1.4)</td>
<td>65 (49.5)(^b)</td>
</tr>
</tbody>
</table>

\(^a\)Items are presented in descending order for the crude unweighted probabilities. The item numbers indicate the order in the questionnaire.

\(^b\)Substantial conditional item probabilities that are greater than 0.40 for the latent classes.

**Comparison of Profiles Across the Latent Classes**

As shown in Table 3, females had a significantly lower prevalence of the normative and active classes than the abstinent class. No significant differences ($\chi^2=2.17-4.68\; P=.10-.34$) were found in age, place of birth, education level, student status, and marital status. Compared with the abstinent class, the normative class had significantly higher sex desire and PHQ-4 scores, while the active class had significantly lower family satisfaction and spent more time on social media, and both the normative and active classes had a significantly more negative COVID-19 mental impact and higher rates of suicidal ideation. The normative class had significantly higher sexual knowledge and lower unsafe sex rates than the other 2 classes. The active class had significantly higher sex desire and PHQ-4 scores and higher rates of nonheterosexual orientation, other sexual risk behaviors, STIs, and drug use than the other 2 classes. For instance, young adults in the active class had significantly higher rates of sexually harassed others, sexual coercion, compensated dating with sex, and multiple sex partners than those in the normative class or abstinent class.
Table 3. Demographic and family characteristics, mental health, sex-related factors, and behavioral health outcomes of the 3 latent classes of online sexual experiences (N=1205).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Abstinent class (N=398, 33.1%)</th>
<th>Normative class (N=675, 56.0%)</th>
<th>Active class (N=132, 10.9%)</th>
<th>Overall 3-class difference, $\chi^2$ (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex, n (%)</td>
<td>264 (66.4)$^a$</td>
<td>300 (44.4)$^b$</td>
<td>28 (21.3)$^c$</td>
<td>88.9 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age (years), mean (SE)</td>
<td>22.7 (0.17)</td>
<td>23.1 (0.12)</td>
<td>23.2 (0.30)</td>
<td>4.60 (2)</td>
<td>.10</td>
</tr>
<tr>
<td>Born in Hong Kong, n (%)</td>
<td>287 (72.0)</td>
<td>497 (73.7)</td>
<td>1111 (83.9)</td>
<td>4.68 (2)</td>
<td>.10</td>
</tr>
<tr>
<td>Education level, mean (SE)</td>
<td>2.08 (0.05)</td>
<td>2.17 (0.04)</td>
<td>2.13 (0.09)</td>
<td>2.17 (2)</td>
<td>.34</td>
</tr>
<tr>
<td>Current student status, n (%)</td>
<td>176 (44.1)</td>
<td>254 (37.7)</td>
<td>48 (36.7)</td>
<td>3.64 (2)</td>
<td>.16</td>
</tr>
<tr>
<td>Married status, n (%)</td>
<td>39 (9.9)</td>
<td>97 (14.4)</td>
<td>13 (10.0)</td>
<td>3.58 (2)</td>
<td>.17</td>
</tr>
<tr>
<td>Family satisfaction (1-5)$^d$, mean (SE)</td>
<td>3.68 (0.04)$^b$</td>
<td>3.60 (0.03)$^b$</td>
<td>3.38 (0.09)$^c$</td>
<td>8.76 (2)</td>
<td>.01</td>
</tr>
<tr>
<td>Nonheterosexual, n (%)</td>
<td>16 (3.9)$^c$</td>
<td>45 (6.7)$^c$</td>
<td>26 (19.9)$^b$</td>
<td>16.3 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sexual knowledge (0-12)$^d$, mean (SE)</td>
<td>9.39 (0.15)$^c$</td>
<td>10.5 (0.08)$^b$</td>
<td>9.83 (0.26)$^c$</td>
<td>46.3 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sex desire (1-5)$^d$, mean (SE)</td>
<td>2.62 (0.06)$^c$</td>
<td>3.30 (0.04)$^b$</td>
<td>3.76 (0.11)$^a$</td>
<td>143.6 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>COVID-19 mental impact (1-5)$^d$, mean (SE)</td>
<td>3.05 (0.05)$^c$</td>
<td>3.35 (0.04)$^b$</td>
<td>3.30 (0.09)$^b$</td>
<td>24.7 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Daily hours on social media, mean (SE)</td>
<td>2.69 (0.11)$^c$</td>
<td>2.98 (0.09)$^b$</td>
<td>3.51 (0.20)$^b$</td>
<td>13.3 (2)</td>
<td>.001</td>
</tr>
<tr>
<td>Sexually harassed others, n (%)</td>
<td>6 (1.6)$^c$</td>
<td>18 (2.7)$^c$</td>
<td>38 (28.5)$^b$</td>
<td>37.3 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sexual coercion, n (%)</td>
<td>11 (2.8)$^c$</td>
<td>24 (3.5)$^c$</td>
<td>32 (24.0)$^b$</td>
<td>23.0 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Compensated dating with sex, n (%)</td>
<td>0 (0.0)$^c$</td>
<td>18 (2.7)$^c$</td>
<td>21 (15.6)$^a$</td>
<td>36.1 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Unsafe sex, n (%)</td>
<td>70 (17.7)$^b$</td>
<td>59 (8.7)$^c$</td>
<td>27 (20.2)$^b$</td>
<td>7.89 (2)</td>
<td>.02</td>
</tr>
<tr>
<td>Multiple sex partners, n (%)</td>
<td>8 (2.1)$^c$</td>
<td>37 (5.5)$^c$</td>
<td>36 (27.4)$^b$</td>
<td>36.9 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Psychological distress (PHQ-4)$^e$ (0-12)$^d$, mean (SE)</td>
<td>1.44 (0.11)$^c$</td>
<td>2.36 (0.10)$^b$</td>
<td>3.39 (0.34)$^a$</td>
<td>59.3 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sexually transmitted infections, n (%)</td>
<td>3 (0.8)$^c$</td>
<td>7 (1.1)$^c$</td>
<td>10 (7.8)$^b$</td>
<td>7.40 (2)</td>
<td>.03</td>
</tr>
<tr>
<td>Drug use, n (%)</td>
<td>3 (0.7)$^c$</td>
<td>21 (3.1)$^b$</td>
<td>17 (12.6)$^a$</td>
<td>19.3 (2)</td>
<td>.001</td>
</tr>
<tr>
<td>Suicidal ideation, n (%)</td>
<td>8 (2.1)$^c$</td>
<td>50 (7.4)$^b$</td>
<td>20 (15.0)$^b$</td>
<td>25.5 (2)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

$^{a,b,c}$Significant post-hoc differences among the 3 latent classes (c < b < a).

$^d$Higher scores indicate better family satisfaction, more negative COVID-19 mental impact, and higher levels of psychological distress, sexual knowledge, and sex desire.

$^e$PHQ-4: Patient Health Questionnaire-4.

Age, place of birth, education level, current student, and marital status were not significantly associated ($P=.13–.98$) with the latent class membership (Table 4). Females had low odds of the active class (OR 0.19, 95% CI 0.09-0.38; $P<.001$) compared with the other 2 classes. Higher COVID-19 mental impact (OR per score 1.54, 95% CI 1.23-1.89; $P<.001$) was associated with high odds of the normative class compared with the abstinent class. Family satisfaction was associated with low odds of the active class (OR per score 0.67-0.68; $P=0.03–.049$) compared with the other 2 classes, while nonheterosexual orientation (OR 3.36-6.82; $P<0.001$) and daily hours spent on social media (OR 1.13-1.19; $P=.01–.04$) were associated with high odds of the active class compared with the other 2 classes. Sexual knowledge was associated with high odds of the normative class compared with the other 2 classes (OR 1.16-1.29; $P=.001–.006$).
Table 4. Multinomial logistic regression for associations of the latent class memberships of online sexual experiences with demographic, family, psychological, and sex-related factors (N=1123).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Active class (reference: abstinent)</th>
<th>Normative class (reference: abstinent)</th>
<th>Active class (reference: normative)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>P value</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>Female (versus male)</td>
<td>0.19 (0.09-0.38)</td>
<td>&lt;.001</td>
<td>0.49 (0.34-0.71)</td>
</tr>
<tr>
<td>Age, per year</td>
<td>1.03 (0.89-1.20)</td>
<td>.66</td>
<td>0.99 (0.90-1.09)</td>
</tr>
<tr>
<td>Born in Hong Kong (versus other places)</td>
<td>1.84 (0.84-4.06)</td>
<td>.13</td>
<td>1.13 (0.70-1.83)</td>
</tr>
<tr>
<td>Education level (1-3)</td>
<td>1.01 (0.70-1.45)</td>
<td>.96</td>
<td>1.02 (0.81-1.27)</td>
</tr>
<tr>
<td>Current student (versus working)</td>
<td>0.97 (0.41-2.31)</td>
<td>.95</td>
<td>0.82 (0.47-1.43)</td>
</tr>
<tr>
<td>Married (versus single)</td>
<td>0.62 (0.26-1.48)</td>
<td>.28</td>
<td>1.01 (0.58-1.74)</td>
</tr>
<tr>
<td>Family satisfaction (1-5)</td>
<td>0.68 (0.47-0.99)</td>
<td>.049</td>
<td>1.02 (0.81-1.29)</td>
</tr>
<tr>
<td>Nonheterosexual (versus heterosexual)</td>
<td>6.82 (2.32-20.1)</td>
<td>&lt;.001</td>
<td>2.03 (0.85-4.83)</td>
</tr>
<tr>
<td>Sexual knowledge (0-12)</td>
<td>1.10 (0.98-1.23)</td>
<td>.09</td>
<td>1.29 (1.18-1.40)</td>
</tr>
<tr>
<td>Sex desire (1-5)</td>
<td>3.69 (2.51-5.42)</td>
<td>&lt;.001</td>
<td>2.22 (1.80-2.73)</td>
</tr>
<tr>
<td>COVID-19 mental impact (1-5)</td>
<td>1.37 (0.99-1.89)</td>
<td>.06</td>
<td>1.54 (1.23-1.89)</td>
</tr>
<tr>
<td>Daily hours on social media, per hour</td>
<td>1.19 (1.04-1.35)</td>
<td>.01</td>
<td>1.05 (0.96-1.15)</td>
</tr>
</tbody>
</table>

*a*OR: odds ratio.

*b*For continuous independent variables, ORs are per score. Higher scores indicate better family satisfaction, more negative COVID-19 mental impact, and higher levels of psychological distress, sexual knowledge, and sex desire.

**Associations Between the Active Class and Behavioral Health Outcomes**

Considering the higher rates of behavioral health outcomes in the active class than in the nonactive classes, the SEM was used to examine the effects of the active class on the behavioral health outcomes. The SEM fitted the data approximately well (RMSEA=0.026, CFI=0.98, SRMR=0.056). Significant factor loadings were found for sexual risk behaviors and PHQ-4 scores. Male sex, being married, nonheterosexual orientation, and higher COVID-19 mental impact were associated (β=0.13-0.27; P=.001-0.02) with more sexual risk behaviors. Male sex, younger age, education level, being married, lower family satisfaction, nonheterosexual orientation, and higher COVID-19 mental impact were associated (β=0.09-0.28; P=.001-0.04) with higher PHQ-4 scores. COVID-19 mental impact was positively associated (β=0.18-0.27; P=.001-0.02) with STIs, drug use, and suicidal ideation.

Figure 1 shows the SEM results. The model covariates and residual correlations between sexual risk behaviors and PHQ-4 scores have not been shown for simplicity. Controlling for the model covariates, the active class was positively associated with sexual risk behaviors and PHQ-4 scores (β=0.13-0.34; P<.001). Sexual risk behaviors were positively associated with STIs and drug use (β=0.71-0.93; P<.001), and PHQ-4 scores were significantly positively associated with suicidal ideation (β=0.40; P<.001). The residual correlations among the 3 behavioral health outcomes were not statistically significant (r=-0.26 to 0.03; P=.35-.91). There was a positive residual correlation between sexual risk behaviors and PHQ-4 scores (r=0.28; P<.001). The active class did not have significant direct effects (β=-0.09 to 0.03; P=.41-.62) on behavioral health outcomes. However, the active class had positive indirect effects on STIs (β=0.31; P<.001) via sexual risk behaviors, drug use (β=0.24; P<.001) via sexual risk behaviors, and suicidal ideation (β=0.05; P<.001) via PHQ-4 scores. These results implied that sexual risk behaviors substantially mediated the effects of the active class on STIs and drug use, and PHQ-4 scores substantially mediated the effects of the active class on suicidal ideation. The model explained 48.8%, 52.7%, and 81.7% of the variances of suicidal ideation, drug use, and STIs, respectively. Compared with the nonactive classes, the active class was indirectly associated with higher prevalences of STIs (Δ [absolute difference in percentage points]=4.8%, 95% CI 0.5%-8.1%), drug use (Δ=7.6%, 95% CI 2.9%-11.8%), and suicidal ideation (Δ=2.5%, 95% CI 0.8%-4.3%).
Figure 1. Standardized coefficients in the structural equation model on the associations among the active class (N=132) compared with the nonactive (abstinent and normative) classes (N=1075) of online sexual experiences (OSEs), sexual risk behaviors, Patient Health Questionnaire-4 (PHQ-4) scores, sexually transmitted infections (STIs), drug use, and suicidal ideation. For simplicity of presentation, model covariates and residual correlations between sexual risk behaviors and PHQ-4 scores are not shown. Standard errors of the coefficients are shown in parenthesis. Factor loadings ($\lambda$) of the latent factors of sexual risk behaviors and PHQ-4 scores are shown with orange dotted arrows. The direct effects ($\beta$) of OSE class on STIs, drug use, and suicidal ideation are shown with pale green dashed arrows. The indirect effects of OSE class on STIs and drug use via sexual risk behaviors are highlighted with red solid arrows. The indirect effects of OSE class on suicidal ideation via PHQ-4 scores are highlighted with blue solid arrows. The residual correlations ($r$) among STIs, drug use, and suicidal ideation are shown with black dash-dotted arrows. Significant indirect effects (red and blue solid arrows) are highlighted in bold.

Discussion

Principal Findings

This is the first study to report results on the latent heterogeneity of OSEs in young adults. Young adults could be classified into the following 3 latent classes: abstinent class with minimal OSEs, normative class with occasional OSEs, and active class with substantive OSEs. The measurement invariance test results supported stability of the 3 classes across sex. The 3 derived OSE classes were validated with distinct profiles in demographic, family, psychological, and sex-related characteristics amid the COVID-19 pandemic in 2022. Moreover, the SEM found significant indirect associations between OSE classes and STIs, drug use, and suicidal ideation via sexual risk behaviors and mental health. In particular, sexual risk behaviors were found to mediate the relationships between OSE classes and STIs and drug use, and mental health was found to mediate the relationship between OSE classes and suicidal ideation.

Comparison With Prior Work

The normative class, comprising the majority of the sample, had mainly nonarousal and solitary arousal OSEs, indicating passive exposure of sex-related information and active pursuit of sex knowledge. This class showed higher sexual knowledge but not increased sexual risk behaviors compared with the other classes and could represent young adults who are curious about sexuality and explore pornographic content online. A cluster randomized controlled trial on internet-based sexuality education programs significantly improved the sexual knowledge of adolescents in mainland China [31]. Our 2021 YSS survey found that internet use was positively associated with sexual knowledge in Hong Kong adolescents amid the COVID-19 pandemic [32]. A national survey in Germany found greater perceived positive outcomes of OSEs (eg, better sexual well-being) than negative outcomes (eg, online dating scams and cyberbullying) in adults [33]. Our findings align with these results and suggest a beneficial relationship between implicit types of OSEs and better sexual knowledge in young adults.
The prevalence of our normative and active classes (66.9%) was similar to the prevalence of online sexual activities (68%) in a previous German study [33]. In our study, males showed a lower prevalence of the abstinent class and a much higher prevalence of the active class than females. This sex difference is consistent with the findings of a meta-analysis [34] on the higher prevalence of online sexual behaviors among young male adults and the literature on higher sensation-seeking and lower impulse control in males in early adulthood [35]. In our sample, there was a significantly higher sex desire (Cohen $d=0.71$; $P<.001$) in males than in females. A longitudinal study on British young adults found significant decreases in sex desire in females but not males during COVID-19 social lockdown restrictions compared with prelockdown levels [36]. Future studies should examine sex differences in OSEs across cultural contexts.

We found that the active class had higher PHQ-4 scores and more sexual risk behaviors than the nonactive classes. These findings agree with previous studies on the positive associations between sexting and depressive symptoms in American youth [4] and between pornography use and sexual risk behaviors in Chinese youth [18]. All of the 3 residual correlations among STIs, drug use, and suicidal ideation were not statistically significant in the SEM. This suggests that the behavioral health outcomes were relatively well explained by the OSE class and latent mediators such that there were no leftover associations among them. The active class was at risk of behavioral health outcomes via sexual risk behaviors and PHQ-4 scores, and our results suggest a mediating role for sexual risk behaviors in the relationship of the active class with STIs and drug use. We recently found a mediating role for PHQ-4 scores between meaning in life and suicidal ideation in Hong Kong young adults [26]. PHQ-4 could be a simple and useful tool in future studies on the relationship between OSEs and risk factors.

This study has several strengths. First, we developed 11 items to measure 3 types of OSEs in young adults [25]. The moderate item-total correlations indicated good reliability and adequate discriminant validity of the items. Second, LCA classified the participants from empirical data into distinct latent classes of OSEs while allowing for potential misclassification errors. The 3 latent classes were validated using the BCH procedure and a 3-step approach [30]. Third, the SEM offered more reliable results by accounting for measurement errors of the latent factors and estimated the direct and indirect effects of OSE classes on behavioral health outcomes.

Early prevention and intervention are necessary to reduce the health burdens on young adults and society. The active class, although the smallest group, had more nonheterosexual individuals, had a worse family relationship, and spent more time on social media than the nonactive classes. Recent studies [37,38] found higher rates of sexual and suicidal risks in nonheterosexual individuals among Canadian and American young adults. Future research can explore risk factors like HIV knowledge and sexual identity stigma in this group [39-41]. Our previous study found that lower family satisfaction was associated with more sexual risk behaviors in Hong Kong adolescents [42]. The COVID-19 pandemic has increased young adults’ social media use, and those with a poor family relationship may rely on the internet for sexual content and be exposed to pornographic content online. Greater internet exposure has been shown to be associated with sexual risk behaviors among young male adults in mainland China [43].

Limitations

Our study had several limitations. First, the cross-sectional design made it difficult to determine the directionality between OSE classes and other variables. Since both OSEs and sexual risk behaviors were measured over the lifetime, their relationships could be bidirectional with reciprocal effects from sexual risk behaviors to greater OSEs. Future longitudinal studies are needed to clarify the temporal relationships among OSEs, sexual risk behaviors, mental health, and behavioral health outcomes. Second, we did not assess the underlying causes for OSEs, and young adults could engage in OSEs for different purposes such as boredom regulation, sexual gratification, and sensation seeking. Further qualitative studies should elucidate and compare the associations and motives for OSEs in the normative and active classes. Third, we did not differentiate between active solicitation and passive exposure in some OSE items, which may have included unwanted or forced behaviors. Further studies should examine the relationships between unwanted OSEs and online sexual harassment. Fourth, our questionnaire did not consider potential confounding factors such as impulsivity, compulsivity, internet and gaming addictions, and peer influences. Future research could examine the comorbidity between problematic OSEs and other risk or protective factors via network analysis. Fifth, the 11 OSE items we developed were not previously validated. Future studies should develop a new scale on OSEs with reference to our items and evaluate the psychometric properties. Sixth, the response rate of the phone survey was only 42.5%, and over half of the eligible participants did not take part in the phone interviews. The relatively low response rate implied potential selection biases. Our study sample of 1205 interviewed people might differ from the 1633 noninterviewed people in terms of OSEs and sexual risk behaviors. It was plausible that young adults with active OSEs could opt out of the survey, which led to underreporting of OSEs and a reduction in the effect size. It is necessary to be cautious when generalizing the present results to the general population of young adults.

Conclusions and Implications

This study provides the first results on the 3 latent classes of OSEs (abstinent, normative, and active) with distinct profiles in demographic and family characteristics, mental health, sex-related factors, and behavioral health outcomes. The active class showed indirect associations with higher rates of STIs and drug use via sexual risk behaviors and higher rates of suicidal ideation via PHQ-4 scores than the nonactive classes. Our results have clinical implications for personalized interventions to help young adults with substantive OSEs.

In Hong Kong, school-based sex education primarily focuses on the biological aspects of sexual and reproductive health among young adults, and sex education could promote sexual knowledge and be protective against STIs [44]. Comprehensive sex education interventions in northwest China have been successful in increasing sexual knowledge and sexual
self-efficacy among adolescents [45]. Recent reviews [46-48] suggest that sex education programs should emphasize the social aspects of sexual health via positive youth development approaches and include up-to-date topics, such as internet pornography, safe use of social media, and caveats of OSEs, to promote media literacy. Social media platforms could be an effective means of delivering sex education programs, and a pilot randomized controlled trial [49] in the United States has shown feasibility in reducing sexual risk behaviors and drug use via mobile health apps among 50 young individuals. Future research is needed to test the cost-effectiveness of mobile health apps in promoting sex knowledge, enhancing self-efficacy for refusing risky OSEs, reducing sexual risk behaviors, and preventing STIs and drug use among young adults.

Acknowledgments
We would like to express our appreciation to Ms Joe Yiu and Mr Sun Chan for providing support for the Youth Sexuality Study and young adults for their participation in the survey. The work was supported by the Family Planning Association of Hong Kong and Department of Health of the Hong Kong SAR government.

Data Availability
The raw data analyzed in this study are available from the first author (TCTF).

Authors’ Contributions
PSFY and THL are co-corresponding authors on this manuscript. Correspondence to THL should be directed to hrmrlnh@hku.hk.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Prevalence of online sexual experiences in male and female participants.
[DOCX File, 15 KB - publichealth_v10i1e50020_app1.docx ]

Multimedia Appendix 2
Fit indices and classification quality of latent class models on online sexual experiences in the whole sample and in male and female participants separately.
[DOCX File, 16 KB - publichealth_v10i1e50020_app2.docx ]

Multimedia Appendix 3
Latent class prevalence of the 3-class model without measurement invariance across sex and the conditional item probabilities of online sexual experiences in male participants.
[DOCX File, 15 KB - publichealth_v10i1e50020_app3.docx ]

Multimedia Appendix 4
Latent class prevalence of the 3-class model without measurement invariance across sex and the conditional item probabilities of online sexual experiences in female participants.
[DOCX File, 15 KB - publichealth_v10i1e50020_app4.docx ]

References


and HIV among youth: Randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 17;8(3):e16251 [FREE Full text] [doi: 10.2196/16251] [Medline: 32181747]

Abbreviations

BCH: Bolck, Croon, and Hagenaars
BIC: Bayesian information criterion
CFI: comparative fit index
LCA: latent class analysis
LMR: Lo-Mendell-Rubin
OR: odds ratio
OSE: online sexual experience
PHQ-4: Patient Health Questionnaire-4
RMSEA: root mean square error of approximation
SEM: structural equation model
SRMR: standardized root mean square residual
STI: sexually transmitted infection
YSS: Youth Sexuality Study
Profiles of Cough and Associated Risk Factors in Nonhospitalized Individuals With SARS-CoV-2 Omicron Variant Infection: Cross-Sectional Online Survey in China

Tingting Xu1, MD; Yuehan Chen1, MD; Wenzhi Zhan1, MD; Kian Fan Chung2,3*, MD, PhD; Zhongmin Qiu4*, MD, PhD; Kewu Huang5, MD, PhD; Ruchong Chen1, MD, PhD; Jiaxing Xie1, MD, PhD; Gang Wang6, MD, PhD; Min Zhang7, MD, PhD; Xuefen Wang8, MD, PhD; Hongmei Yao9, MD, PhD; Xiuling Liao10*, MD, PhD; Yunhui Zhang11, MD, PhD; Guojun Zhang12, MD, PhD; Wei Zhang13, MD, PhD; Dejun Sun14, MD, PhD; Jia Zhu15, MD, PhD; Shujuan Jiang16, MD, PhD; Jiantao Feng17, MD, PhD; Jianping Zhao18, MD, PhD; Gengyun Sun19, MD, PhD; Huaqiong Huang20, MD, PhD; Jianyong Zhang21, MD, PhD; Lingwei Wang22, MD, PhD; Feng Wu23, MD, PhD; Suyun Li24, MD, PhD; Pusheng Xu25, MD, PhD; Chunhua Chi26, MD, PhD; Ping Chen27, MD, PhD; Mei Jiang2, MD, PhD; Wen He1, MD; Lianrong Huang1, MD, PhD; Wei Luo1, MD, PhD; Shiyue Li1, MD, PhD; Nanshan Zhong2, MD, PhD; Kefang Lai1, MD, PhD; China Cough Coalition28*

1State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
2Department of Respiratory Medicine, Royal Brompton and Harefield Hospital, London, United Kingdom
3Experimental Studies Unit, National Heart & Lung Institute, Imperial College London, London, United Kingdom
4Department of Pulmonary and Critical Care Medicine, School of Medicine, Tongji Hospital, Tongji University, Shanghai, China
5Department of Pulmonary and Critical Care Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China
6West China School of Medicine, West China Hospital of Sichuan University, Chengdu, China
7Department of Respiratory and Critical Care Medicine, Shanghai General Hospital, Shanghai JiaoTong University School of Medicine, Shanghai, China
8Department of Respiratory Medicine, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
9Department of Respiratory Medicine, Guiyang Provincial People’s Hospital, Guiyang, Guizhou, China
10The First People’s Hospital of Yunnan Province, Kunming, China
11Department of Respiratory Medicine, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China
12Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Nanchang University, Nanchang, China
13Department of Pulmonary and Critical Care Medicine, the Inner Mongolia Autonomous Region People’s Hospital, Hohhot, China
14Department of Respiratory Medicine, Jiangsu Province Hospital of Chinese Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China
15Department of Pulmonary Medicine, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, China
16Department of Respiratory Medicine, Xiangya Hospital, Central South University, Changsha, China
17Department of Respiratory Medicine, Tongji Hospital, Tongji Medical College of Hua Zhong University of Science and Technology, Wuhan, China
18Department of Respiratory Medicine, The First Affiliated Hospital of Medical University of Anhui, Hefei, China
19Key Laboratory of Respiratory Disease of Zhejiang Province, Department of Respiratory and Critical Care Medicine, Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China
20The Second Department of Pulmonary and Critical Care Medicine, Affiliated Hospital of Zunyi Medical University, Zunyi, China
21Shenzhen People’s Hospital, Shenzhen Institute of Respiratory Diseases, Shenzhen, China
22Department of Pulmonary and Critical Care Medicine, Huizhou Third People’s Hospital, Guangzhou Medical University, Huizhou, China
23The First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou, China
24The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
25Department of Respiratory and Critical Care Medicine, Peking University First Hospital, Beijing, China
26General Hospital of Northern Theater Command, Shenyang, China
27See Acknowledgments

*these authors contributed equally

Corresponding Author:
Abstract

Background: Cough is a common symptom during and after COVID-19 infection; however, few studies have described the cough profiles of COVID-19.

Objective: The aim of this study was to investigate the prevalence, severity, and associated risk factors of severe and persistent cough in individuals with COVID-19 during the latest wave of the Omicron variant in China.

Methods: In this nationwide cross-sectional study, we collected information of the characteristics of cough from individuals with infection of the SARS-CoV-2 Omicron variant using an online questionnaire sent between December 31, 2022, and January 11, 2023.

Results: There were 11,718 (n=7978, 68.1% female) nonhospitalized responders, with a median age of 37 (IQR 30-47) years who responded at a median of 16 (IQR 12-20) days from infection onset to the time of the survey. Cough was the most common symptom, occurring in 91.7% of participants, followed by fever, fatigue, and nasal congestion (68.8%-87.4%). The median cough visual analog scale (VAS) score was 70 (IQR 50-80) mm. Being female (odds ratio [OR] 1.31, 95% CI 1.20-1.43), having a COVID-19 vaccination history (OR 1.71, 95% CI 1.37-2.12), current smoking (OR 0.48, 95% CI 0.41-0.58), chronic cough (OR 2.04, 95% CI 1.69-2.45), coronary heart disease (OR 1.71, 95% CI 1.17-2.52), asthma (OR 1.22, 95% CI 1.02-1.46), and gastroesophageal reflux disease (GERD) (OR 1.21, 95% CI 1.01-1.45) were independent factors for severe cough (VAS>70, 37.4%). Among all respondents, 35.0% indicated having a productive cough, which was associated with risk factors of being female (OR 1.44, 95% CI 1.31-1.57), having asthma (OR 1.84, 95% CI 1.52-2.22), chronic cough (OR 1.44, 95% CI 1.19-1.74), and GERD (OR 1.22, 95% CI 1.01-1.47). Persistent cough (>3 weeks) occurred in 13.0% of individuals, which was associated with the risk factors of having diabetes (OR 2.24, 95% CI 1.30-3.85), asthma (OR 1.70, 95% CI 1.11-2.62), and chronic cough (OR 1.97, 95% CI 1.32-2.94).

Conclusions: Cough is the most common symptom in nonhospitalized individuals with Omicron SARS-CoV-2 variant infection. Being female, having asthma, chronic cough, GERD, coronary heart disease, diabetes, and a COVID-19 vaccination history emerged as independent factors associated with severe cough, productive cough, and persistent cough.

(JMIR Public Health Surveill 2024;10:e47453) doi:10.2196/47453

KEYWORDS
COVID-19; Omicron variant; nonhospitalized; cough

Introduction

The SARS-CoV-2 virus has been spreading globally since 2019, causing over 664 million infections and 6.7 million deaths as of January 14, 2023 [1]. The epidemiology and clinical characteristics of COVID-19 have been widely studied; however, SARS-CoV-2 has continued to mutate from the dominant Alpha to Delta variants and now the Omicron variant and subvariants, and the clinical characteristics of COVID-19 also change according to the particular virus variant [2-4]. The Omicron variant has been the dominant variant in the fourth wave of the global COVID-19 pandemic having infected an estimated 300 million people worldwide as of January 2023. Since December 2022, infection with the Omicron variant has been spreading rapidly across China.

Omicron is recognized as being highly contagious with a high capability for immune escape [5,6]. Over half of patients are reported to have at least one symptom during the infection period, including fever, cough, sore throat, fatigue, dyspnea, headache, and body pain [2,7]. Among these symptoms, cough is one of the most common [8,9]. Experiencing a chronic cough is associated with a significant decrease in quality of life, with symptoms such as incontinence, chest pain, headache, and poor sleep [10]. Studies in the United Kingdom [3] and the United States [4] have shown that the prevalence of acute cough during Omicron infection ranges from approximately 67.4% to 78%. A study conducted in a mobile cabin hospital in Shanghai demonstrated that cough was present in 57.5% of 1139 patients [7]. In addition, nearly 2.4%-7.2% of patients still experience cough as a symptom 1-2 months after Omicron infection is resolved [8,9]. However, few studies have described the specific cough features such as its severity, pattern, and risk factors.

We hypothesized that certain risk factors might be related with severe cough, productive cough, and persistent cough after...
Omicron infection. Additionally, there has been no nationwide survey concerning the clinical characteristics of Omicron infection in China to date. Therefore, we investigated the prevalence, severity, and associated risk factors for cough in nonhospitalized individuals with Omicron infection during the latest COVID-19 wave (2022-2023) in China.

Methods

Study Design and Participants

A nationwide survey was conducted online by China Cough Coalition [11], which is an organization dedicated to promoting the regular diagnosis, management, and research on cough in China, and has been assessing the cough characteristics of the Omicron pandemic. From December 31, 2022, to January 11, 2023, we recruited individuals aged ≥18 years with positive nucleic acid or antigen detection of SARS-CoV-2 or highly suspected symptoms of COVID-19; individuals with mental illness or other conditions that might hinder their ability to complete the survey were excluded. Only nonhospitalized individuals were included in the study.

Data were collected using Wenjuanxing [12], an online questionnaire survey platform in China. Participants were asked to fill out the questionnaire by using their smartphone, and each device was restricted to submitting the questionnaire results only once. The questionnaires were mainly distributed by physicians of China Cough Coalition and the survey link was also available for the public. At the top of the questionnaire, the purpose and contents of the survey were briefly introduced and participants were required to read the informed consent and confirm that they voluntarily agreed to join this survey. Participants had full right to quit the study at any time during the survey.

Ethics Considerations

This study was approved by the Ethics Review Committees of the First Affiliated Hospital of Guangzhou Medical University (ES-2023-012-01). Although verbal consent was provided as described above, the requirement for written informed consent was waived by the institutional ethics board of the hospital. No financial compensation was offered to the subjects for their participation. All data were collected anonymously and promised to be used only in this study.

Questionnaire Items

A standardized and structured questionnaire was designed for this study. Prior to the distribution of the final version of the questionnaire, it was first tested in a pilot study (data not shown). The questionnaire consisted of 21 items, divided into the following six aspects (see Multimedia Appendix 1): (1) baseline data (age, sex, smoking history, vaccination history for COVID-19, and comorbidities); (2) detection of SARS-CoV-2 (date of the positive result of SARS-CoV-2 antigen/nucleic acid testing); (3) general and severe symptoms, including date of onset of COVID-19 symptoms (cough, fever, fatigue, myalgia/arthritis, nasal congestion, headache/dizziness, sore throat, runny nose, hyposmia/hypogeusia, chest tightness, diarrhea, postnasal drip, and conjunctivitis; see Multimedia Appendix 1 for the definition of “severe” symptoms); (4) cough characteristics, including the date of cough onset, cough pattern, type of sputum, duration of cough, timing of cough, and cough severity compared with previous common cold experience; (5) results of chest computed tomography/chest X-ray according to radiology reports if available; and (6) medication, including antipyretics, antibiotics, antiviral drugs, antitussives, nasal drugs, and traditional Chinese medicine. Cough severity was evaluated using a visual analog scale (VAS), with a range of 0-100 mm, where 0 indicates no cough and 100 indicates the worst cough imaginable [13]. Severe cough was defined as VAS > 70. Persistent cough was defined as a cough duration >3 weeks and chronic cough was defined as a cough duration >8 weeks.

Statistical Analysis

Categorical variables are presented as frequency (percentage), while continuous variables are expressed as mean (SD) or median (IQR). Statistical comparisons between groups were performed with unpaired Student t tests for normally distributed data, Mann-Whitney U tests for skewed data, and χ² tests or Fisher exact tests for proportional data. Multivariable logistic regression was performed to detect the risk factors for persistent cough adjusted by age, sex, and duration from infection to the survey, assessed according to estimates of adjusted odds ratios, 95% CIs, and P values [14]. The final model selection was determined according to the backward selection of all factors based on the Akaike information criterion. Statistical analysis was performed using R software (v 4.1.2).

Results

Demographic and General Information of Participants

A total of 12,227 questionnaires were recorded in 31 provinces across China. The number of questionnaires in different provinces is shown in Multimedia Appendix 2. We excluded 78 individuals with incorrect information and 431 hospitalized individuals. Therefore, data were analyzed from a total of 11,718 nonhospitalized individuals, including clinic attendees (15.4%) and individuals not diagnosed or treated in hospitals (84.6%).

The demographic characteristics of the individuals are shown in Table 1. Among the 11,718 individuals with a median age of 37.0 (IQR 30.0-47.0) years, 7978 (68.1%) were female. The median duration from Omicron infection to the survey was 16.0 (IQR 12.0-20.0) days. The majority of individuals were not current smokers and had no comorbidities. The most frequent comorbidities were chronic rhinitis/sinusitis, followed by hypertension, chronic urticaria/allergic dermatitis, and asthma.

Among the total sample, the most common symptoms were cough, followed by fever, fatigue, and nasal congestion. The most common severe symptoms included severe bone pain/myalgia, high fever, and severe nasal congestion (Table 2).
Table 1. Baseline characteristics of nonhospitalized participants with Omicron variant infection (N=11,718).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>7978 (68.1)</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>37.0 (30.0–47.0)</td>
</tr>
<tr>
<td>Positive nucleic acid/antigen test, n (%)</td>
<td>9950 (84.9)</td>
</tr>
<tr>
<td>Duration from infection to survey (days), median (IQR)</td>
<td>16.0 (12.0–20.0)</td>
</tr>
<tr>
<td>Clinic visits, n (%)</td>
<td>1804 (15.4)</td>
</tr>
<tr>
<td>Vaccination(^a), n (%)</td>
<td>11,259 (96.1)</td>
</tr>
</tbody>
</table>

**Last vaccination\(^b\), n (%)**
- ≤6 months: 1707 (15.2)
- >6 months: 9552 (84.8)

**Medication, n (%)**
- Antipyretics: 8262 (70.5)
- Traditional Chinese medicine: 5127 (43.8)
- Antitussive: 2987 (25.5)
- Antibiotics: 2280 (19.5)
- Antiviral drugs: 356 (3.0)
- Nasal drugs: 229 (2.0)
- None: 1007 (8.6)

**Smoking status, n (%)**
- Current smoker: 906 (7.7)
- Nonsmoker: 10,812 (92.3)

**Comorbidities, n (%)**
- COPD\(^c\): 136 (1.2)
- Asthma: 546 (4.7)
- Chronic cough: 495 (4.2)
- Interstitial lung disease: 35 (0.3)
- Chronic rhinitis/sinusitis: 1838 (15.7)
- Hypertension: 843 (7.2)
- Diabetes: 314 (2.7)
- Coronary heart disease: 121 (1.0)
- Cerebrovascular disease: 78 (0.7)
- GERD\(^d\): 521 (4.4)
- Malignant tumor: 191 (1.6)
- Chronic kidney disease: 59 (0.5)
- Chronic urticaria/allergic dermatitis: 644 (5.5)
- None of the above diseases: 7578 (64.7)

\(^a\)Inactivated vaccine for SARS-CoV-2.
\(^b\)Proportions were calculated only for patients with a COVID-19 vaccination history.
\(^c\)COPD: chronic obstructive pulmonary disease.
\(^d\)GERD: gastroesophageal reflux disease.
Table 2. Prevalence of general and severe symptoms in nonhospitalized individuals with Omicron variant infection (N=11,718).

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General symptoms</strong></td>
<td></td>
</tr>
<tr>
<td>Cough</td>
<td>10,741 (91.7)</td>
</tr>
<tr>
<td>Fever</td>
<td>10,243 (87.4)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>9407 (80.3)</td>
</tr>
<tr>
<td>Nasal congestion</td>
<td>8058 (68.8)</td>
</tr>
<tr>
<td>Myalgia/arthralgia</td>
<td>8033 (68.6)</td>
</tr>
<tr>
<td>Headache/dizziness</td>
<td>7740 (66.1)</td>
</tr>
<tr>
<td>Sore throat</td>
<td>7576 (64.7)</td>
</tr>
<tr>
<td>Runny nose</td>
<td>6901 (58.9)</td>
</tr>
<tr>
<td>Hyposmia/hypogeusia</td>
<td>5403 (46.1)</td>
</tr>
<tr>
<td>Chest tightness</td>
<td>4019 (34.3)</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>2628 (22.4)</td>
</tr>
<tr>
<td>Postnasal drip</td>
<td>1887 (16.1)</td>
</tr>
<tr>
<td>Conjunctivitis</td>
<td>571 (4.9)</td>
</tr>
<tr>
<td>Others</td>
<td>712 (6.1)</td>
</tr>
<tr>
<td><strong>Severe symptoms</strong></td>
<td></td>
</tr>
<tr>
<td>Severe bone pain/myalgia</td>
<td>4858 (41.5)</td>
</tr>
<tr>
<td>High fever (≥39.1°C)</td>
<td>4558 (38.9)</td>
</tr>
<tr>
<td>Severe nasal congestion</td>
<td>4129 (35.2)</td>
</tr>
<tr>
<td>Severe sore throat</td>
<td>3851 (32.9)</td>
</tr>
<tr>
<td>Severe headache</td>
<td>3366 (28.7)</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>1867 (15.9)</td>
</tr>
<tr>
<td>None of the above</td>
<td>2438 (20.8)</td>
</tr>
</tbody>
</table>

Cough Characteristics

Among the 10,741 respondents with cough, 44.0% (n=4724) presented with cough during both daytime and nighttime and 36.8% (n=3956) only had the cough during the daytime (Table 3). In terms of the cough pattern, 35.0% (n=3760) of individuals presented with a productive cough; white thick sputum and yellow purulent sputum were common.

Coughs were commonly initiated within 3 days, as reported by 6580 (61.3%) of the 10,741 individuals with cough. At entry to the survey, 38.2% (n=4103) of individuals still had cough, while 61.8% (n=6638) of individuals reported that their cough was relieved. Among the 2209 individuals with a duration from infection to the survey over 3 weeks, the cough was relieved within 1 week in 23.9% (n=528) of individuals, lasted for 1-3 weeks in 52.6% (n=1162) of individuals, developed into a subacute cough (3-8 weeks) in 12.7% (n=281) of individuals, and developed into a chronic cough (>8 weeks) in 0.3% (n=8) of individuals.
Table 3. Characteristics of cough in individuals with cough (N=10,741).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>7466 (69.5)</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>37.0 (30.0-46.0)</td>
</tr>
<tr>
<td>Duration from infection to survey (days), median (IQR)</td>
<td>16.0 (12.0-20.0)</td>
</tr>
<tr>
<td>Cough onset time (days), median (IQR)</td>
<td>3.0 (2.0-4.0)</td>
</tr>
<tr>
<td>Cough duration (weeks), median (IQR)</td>
<td>10.0 (7.0-14.0)</td>
</tr>
<tr>
<td><strong>Cough duration (weeks), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt; 1</td>
<td>3975 (37.0)</td>
</tr>
<tr>
<td>1-3</td>
<td>6424 (59.8)</td>
</tr>
<tr>
<td>3-8</td>
<td>331 (3.1)</td>
</tr>
<tr>
<td>&gt; 8</td>
<td>11 (0.1)</td>
</tr>
<tr>
<td>Cough VAS(^a), median (IQR)</td>
<td>70 (50-80)</td>
</tr>
<tr>
<td><strong>Cough compared to previous common cold, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Less severe</td>
<td>639 (5.9)</td>
</tr>
<tr>
<td>Similar</td>
<td>2525 (23.5)</td>
</tr>
<tr>
<td>More severe</td>
<td>7577 (70.5)</td>
</tr>
<tr>
<td><strong>Cough timing, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Both daytime and nighttime</td>
<td>4724 (44.0)</td>
</tr>
<tr>
<td>Mainly daytime</td>
<td>3956 (36.8)</td>
</tr>
<tr>
<td>Mainly evening (before falling asleep)</td>
<td>1377 (12.8)</td>
</tr>
<tr>
<td>Mainly nighttime (after falling asleep)</td>
<td>684 (6.4)</td>
</tr>
<tr>
<td><strong>Cough pattern, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Nonproductive</td>
<td>6981 (65.0)</td>
</tr>
<tr>
<td>Productive</td>
<td>3760 (35.0)</td>
</tr>
<tr>
<td><strong>Property of sputum(^b), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>White and thick viscous</td>
<td>809 (21.5)</td>
</tr>
<tr>
<td>Yellow and then white</td>
<td>737 (19.6)</td>
</tr>
<tr>
<td>White with a little yellow</td>
<td>644 (17.1)</td>
</tr>
<tr>
<td>Yellow and purulent</td>
<td>613 (16.3)</td>
</tr>
<tr>
<td>White and then yellow</td>
<td>536 (14.3)</td>
</tr>
<tr>
<td>White and watery</td>
<td>286 (7.6)</td>
</tr>
<tr>
<td>Blood-stained</td>
<td>135 (3.6)</td>
</tr>
</tbody>
</table>

\(^a\)VAS: visual analog scale.
\(^b\)Proportions were calculated among patients with productive cough.

**Factors Associated With Severe, Productive, and Persistent Cough**

With a cut-off VAS score of 70, 4388 (37.4%) of the total 11,718 respondents were categorized in the severe cough group and 7330 (62.6%) were categorized in the nonsevere group. Compared with those in the nonsevere group, the group with severe cough had a higher proportion of female participants (\(P<.001\)), a younger age (\(P<.001\)), fewer current smokers (\(P<.001\)), more individuals who had been vaccinated for COVID-19 (\(P<.001\)), and higher proportions of individuals with gastroesophageal reflux disease (GERD) (\(P=.03\)) and chronic cough (\(P<.001\)) as comorbidities (Table 4).

With respect to the different patterns of cough, more female respondents presented with a productive cough than male respondents (\(P<.001\)). In addition, the group with a productive cough included higher proportions of individuals with COPD (\(P=.01\)), asthma (\(P<.001\)), chronic cough (\(P<.001\)), interstitial lung disease (\(P=.03\)), chronic rhinitis/sinusitis (\(P=.003\)), and GERD (\(P=.008\)) as comorbidities (Table 5).
The group with a persistent cough included higher proportions of individuals with their most recent vaccination injected longer than 6 months before ($P = .03$) and higher proportions of comorbidities, including asthma ($P = .005$), chronic cough ($P < .001$), chronic rhinitis/sinusitis ($P = .02$), and diabetes ($P = .002$) compared to the group with acute cough (Table 6). Multivariate analysis showed that being female; being younger; having received vaccination for COVID-19; having asthma, chronic cough, coronary heart disease, or GERD as comorbidities; and being a nonsmoker were independent factors associated with more severe cough (Figure 1A). Being female, of younger age, and having asthma, chronic cough, and GERD as comorbidities were independent factors for productive cough (Figure 1B). The most recent vaccination more than 6 months ago, and having asthma, chronic cough, and diabetes were independent factors for persistent cough (Figure 1C).

Table 4. Comparison of characteristics between individuals with severe and nonsevere cough. a

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=11,718)</th>
<th>Nonsevere cough (n=7330)</th>
<th>Severe cough (n=4388)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>7978 (68.1)</td>
<td>4753 (64.8)</td>
<td>3225 (73.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>37.0 (30.0-47.0)</td>
<td>38.0 (30.0-48.0)</td>
<td>36.0 (29.0-45.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Duration from infection to survey (days), median (IQR)</td>
<td>16.0 (12.0-20.0)</td>
<td>16.0 (12.0-20.0)</td>
<td>16.0 (12.0-20.0)</td>
<td>.06</td>
</tr>
<tr>
<td>Vaccination b, n (%)</td>
<td>11259 (96.1)</td>
<td>6991 (95.4)</td>
<td>4268 (97.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Last vaccination c, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.60</td>
</tr>
<tr>
<td>≤6 months</td>
<td>1707 (15.2)</td>
<td>1070 (15.3)</td>
<td>637 (14.9)</td>
<td></td>
</tr>
<tr>
<td>&gt;6 months</td>
<td>9552 (84.8)</td>
<td>5921 (84.7)</td>
<td>3631 (85.1)</td>
<td></td>
</tr>
<tr>
<td>Comorbidities, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD d</td>
<td>136 (1.2)</td>
<td>96 (1.3)</td>
<td>40 (0.9)</td>
<td>.06</td>
</tr>
<tr>
<td>Asthma</td>
<td>546 (4.7)</td>
<td>325 (4.4)</td>
<td>221 (5.0)</td>
<td>.15</td>
</tr>
<tr>
<td>Chronic cough</td>
<td>495 (4.2)</td>
<td>231 (3.2)</td>
<td>264 (6.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interstitial lung disease</td>
<td>35 (0.3)</td>
<td>16 (0.2)</td>
<td>19 (0.4)</td>
<td>.06</td>
</tr>
<tr>
<td>Chronic rhinitis/sinusitis</td>
<td>1838 (15.7)</td>
<td>1155 (15.8)</td>
<td>683 (15.6)</td>
<td>.80</td>
</tr>
<tr>
<td>Hypertension</td>
<td>843 (7.2)</td>
<td>549 (7.5)</td>
<td>294 (6.7)</td>
<td>.12</td>
</tr>
<tr>
<td>Diabetes</td>
<td>314 (2.7)</td>
<td>209 (2.9)</td>
<td>105 (2.4)</td>
<td>.15</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>121 (1.0)</td>
<td>70 (1.0)</td>
<td>51 (1.2)</td>
<td>.33</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>78 (0.7)</td>
<td>47 (0.6)</td>
<td>31 (0.7)</td>
<td>.76</td>
</tr>
<tr>
<td>GERD e</td>
<td>521 (4.4)</td>
<td>302 (4.1)</td>
<td>219 (5.0)</td>
<td>.03</td>
</tr>
<tr>
<td>Malignant tumor</td>
<td>191 (1.6)</td>
<td>119 (1.6)</td>
<td>72 (1.6)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>59 (0.5)</td>
<td>42 (0.6)</td>
<td>17 (0.4)</td>
<td>.22</td>
</tr>
<tr>
<td>Chronic urticaria/allergic dermatitis</td>
<td>644 (5.5)</td>
<td>389 (5.3)</td>
<td>255 (5.8)</td>
<td>.26</td>
</tr>
<tr>
<td>None of the above diseases</td>
<td>7578 (64.7)</td>
<td>4760 (64.9)</td>
<td>2818 (64.2)</td>
<td>.44</td>
</tr>
<tr>
<td>Smoking status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Current smoker</td>
<td>906 (7.7)</td>
<td>718 (9.8)</td>
<td>188 (4.3)</td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>10812 (92.3)</td>
<td>6612 (90.2)</td>
<td>4200 (95.7)</td>
<td></td>
</tr>
</tbody>
</table>

a Severe cough was defined as a visual analog scale score >70; the score for individuals without cough was set to 0.
bReceived an inactivated vaccine for SARS-CoV-2.
cProportions were calculated only for patients with a COVID-19 vaccination history.
dCOPD: chronic obstructive pulmonary disease.
eGERD: gastroesophageal reflux disease.
# Table 5. Comparison of characteristics between individuals with or without productive cough.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=10,741)</th>
<th>Nonproductive cough (n=6981)</th>
<th>Productive cough (n=3760)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>7466 (69.5)</td>
<td>4676 (67.0)</td>
<td>2790 (74.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>37.0 (30.0-46.0)</td>
<td>37.0 (30.0-47.0)</td>
<td>36.0 (29.0-45.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Duration from infection to survey (days), median (IQR)</td>
<td>16.0 (12.0-20.0)</td>
<td>16.0 (12.0-20.0)</td>
<td>16.0 (12.0-20.0)</td>
<td>.53</td>
</tr>
<tr>
<td>Vaccination(\text{a}), n (%)</td>
<td>10361 (96.5)</td>
<td>6739 (96.5)</td>
<td>3622 (96.3)</td>
<td>.62</td>
</tr>
<tr>
<td><strong>Last vaccination(\text{b}), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.39</td>
</tr>
<tr>
<td>≤6 months</td>
<td>1555 (15.0)</td>
<td>996 (14.8)</td>
<td>559 (15.4)</td>
<td></td>
</tr>
<tr>
<td>&gt;6 months</td>
<td>8806 (85.0)</td>
<td>5743 (85.2)</td>
<td>3063 (84.6)</td>
<td></td>
</tr>
<tr>
<td>Comorbidities, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD(\text{c})</td>
<td>112 (1.0)</td>
<td>60 (0.9)</td>
<td>52 (1.4)</td>
<td>.01</td>
</tr>
<tr>
<td>Asthma</td>
<td>496 (4.6)</td>
<td>250 (3.6)</td>
<td>246 (6.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chronic cough</td>
<td>470 (4.4)</td>
<td>263 (3.8)</td>
<td>207 (5.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interstitial lung disease</td>
<td>33 (0.3)</td>
<td>15 (0.2)</td>
<td>18 (0.5)</td>
<td>.03</td>
</tr>
<tr>
<td>Chronic rhinitis/sinusitis</td>
<td>1729 (16.1)</td>
<td>1070 (15.3)</td>
<td>659 (17.5)</td>
<td>.003</td>
</tr>
<tr>
<td>Hypertension</td>
<td>750 (7.0)</td>
<td>484 (6.9)</td>
<td>266 (7.1)</td>
<td>.82</td>
</tr>
<tr>
<td>Diabetes</td>
<td>272 (2.5)</td>
<td>171 (2.4)</td>
<td>101 (2.7)</td>
<td>.50</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>105 (1.0)</td>
<td>65 (0.9)</td>
<td>40 (1.1)</td>
<td>.57</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>67 (0.6)</td>
<td>43 (0.6)</td>
<td>24 (0.6)</td>
<td>.99</td>
</tr>
<tr>
<td>GERD(\text{d})</td>
<td>494 (4.6)</td>
<td>293 (4.2)</td>
<td>201 (5.3)</td>
<td>.008</td>
</tr>
<tr>
<td>Malignant tumor</td>
<td>178 (1.7)</td>
<td>113 (1.6)</td>
<td>65 (1.7)</td>
<td>.73</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>50 (0.5)</td>
<td>31 (0.4)</td>
<td>19 (0.5)</td>
<td>.77</td>
</tr>
<tr>
<td>Chronic urticaria/allergic dermatitis</td>
<td>609 (5.7)</td>
<td>396 (5.7)</td>
<td>213 (5.7)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>None of the above diseases</td>
<td>6936 (64.6)</td>
<td>4610 (66)</td>
<td>2326 (61.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Smoking status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.13</td>
</tr>
<tr>
<td>Current smoker</td>
<td>679 (6.3)</td>
<td>460 (6.6)</td>
<td>219 (5.8)</td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>10,062 (93.7)</td>
<td>6521 (93.4)</td>
<td>3541 (94.2)</td>
<td></td>
</tr>
</tbody>
</table>

\(\text{a}\)Received an inactivated SARS-CoV-2 vaccine.

\(\text{b}\)Proportions were calculated only among patients with a COVID-19 vaccination history.

\(\text{c}\)COPD: chronic obstructive pulmonary disease.

\(\text{d}\)GERD: gastroesophageal reflux disease.
Table 6. Comparison of characteristics between individuals with persistent cough and acute cough.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=10,741)</th>
<th>Acute cough(^a) (n=10,399)</th>
<th>Persistent cough (n=342)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>7466 (69.5)</td>
<td>7213 (69.4)</td>
<td>253 (74.0)</td>
<td>.08</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>37.0 (30.0–46.0)</td>
<td>37.0 (30.0–46.0)</td>
<td>38.5 (32.0–48.0)</td>
<td>.003</td>
</tr>
<tr>
<td>Duration from infection to survey (days), median (IQR)</td>
<td>16.0 (12.0–20.0)</td>
<td>16.0 (12.0–20.0)</td>
<td>26.0 (23.0–29.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vaccination(^b), n (%)</td>
<td>10361 (96.5)</td>
<td>10029 (96.4)</td>
<td>332 (97.1)</td>
<td>.63</td>
</tr>
<tr>
<td><strong>Last vaccination(^c), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>≤6 months</td>
<td>1555 (15.0)</td>
<td>1520 (15.2)</td>
<td>35 (10.5)</td>
<td></td>
</tr>
<tr>
<td>&gt;6 months</td>
<td>8806 (85.0)</td>
<td>8509 (84.8)</td>
<td>297 (89.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Comorbidities, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD(^d)</td>
<td>112 (1.0)</td>
<td>108 (1.0)</td>
<td>4 (1.2)</td>
<td>.78</td>
</tr>
<tr>
<td>Asthma</td>
<td>496 (4.6)</td>
<td>469 (4.5)</td>
<td>27 (7.9)</td>
<td>.005</td>
</tr>
<tr>
<td>Chronic cough</td>
<td>470 (4.4)</td>
<td>438 (4.2)</td>
<td>32 (9.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interstitial lung disease</td>
<td>1729 (16.1)</td>
<td>1668 (16.0)</td>
<td>61 (17.8)</td>
<td>.42</td>
</tr>
<tr>
<td>Chronic rhinitis/sinusitis</td>
<td>33 (0.3)</td>
<td>29 (0.3)</td>
<td>4 (1.2)</td>
<td>.02</td>
</tr>
<tr>
<td>Hypertension</td>
<td>750 (7.0)</td>
<td>728 (7.0)</td>
<td>22 (6.4)</td>
<td>.77</td>
</tr>
<tr>
<td>Diabetes</td>
<td>272 (2.5)</td>
<td>254 (2.4)</td>
<td>18 (5.3)</td>
<td>.002</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>105 (1.0)</td>
<td>100 (1.0)</td>
<td>5 (1.5)</td>
<td>.39</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
<td>67 (0.6)</td>
<td>62 (0.6)</td>
<td>5 (1.5)</td>
<td>.06</td>
</tr>
<tr>
<td>GERD(^e)</td>
<td>494 (4.6)</td>
<td>476 (4.6)</td>
<td>18 (5.3)</td>
<td>.64</td>
</tr>
<tr>
<td>Malignant tumor</td>
<td>178 (1.7)</td>
<td>168 (1.6)</td>
<td>10 (2.9)</td>
<td>.10</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>50 (0.5)</td>
<td>49 (0.5)</td>
<td>1 (0.3)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Chronic urticaria/allergic dermatitis</td>
<td>609 (5.7)</td>
<td>587 (5.6)</td>
<td>22 (6.4)</td>
<td>.62</td>
</tr>
<tr>
<td>None of the above diseases</td>
<td>6936 (64.6)</td>
<td>6743 (64.8)</td>
<td>193 (56.4)</td>
<td>.002</td>
</tr>
<tr>
<td><strong>Smoking status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Current smoker</td>
<td>679 (6.3)</td>
<td>666 (6.4)</td>
<td>13 (3.8)</td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>10062 (93.7)</td>
<td>9733 (93.6)</td>
<td>329 (96.2)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Acute cough: duration of cough ≤7 days and without cough at the time of this survey.  
\(^b\)Received an inactivated SARS-CoV-2 vaccine.  
\(^d\)Proportions were calculated only among patients with a COVID-19 vaccination history.  
\(^e\)GERD: gastroesophageal reflux disease.
Figure 1. Forest plot of risk factors for severe cough (VAS>70), productive cough, and persistent cough. (A) Adjusted odds ratios (ORs) of factors for severe cough (VAS>70) versus nonsevere cough (VAS≤70). (B) Adjusted ORs of risk factors for productive cough versus nonproductive cough. (C) Adjusted ORs of risk factors for persistent cough versus acute cough. Adjusted by age, sex, and duration from infection to survey. COPD: chronic obstructive pulmonary disease; GERD: gastroesophageal reflux disease; VAS: visual analog scale.

Discussion

To the best of our knowledge, this is the first study with a large sample size to investigate the characteristics of cough and associated factors in nonhospitalized individuals infected with the SARS-CoV-2 Omicron variant in China. In this study, the prevalence of cough was 92.6%, followed by fever, fatigue, myalgia/arthritis, nasal congestion, headache/dizziness, sore throat, and runny nose. Productive cough (35.0%) with white thick sputum or yellow purulent sputum was common, which developed into a persistent cough in a subset of the surveyed individuals. Female sex, being a current smoker, and having comorbidities (such as asthma, chronic cough, GERD, and/or diabetes) were factors associated with severe, productive, and persistent cough.

Early reports showed that the prevalence of cough during COVID-19 infection ranged from 44% to 72.5% [15], with a higher prevalence associated with Omicron variant infection than with Delta variant infection [4], although one survey from the United Kingdom [3] indicated that the prevalence of cough was similar between infections with the Omicron and Delta variants. In this study, the prevalence of cough was 92.6%, which was higher than that reported in previous studies [2-4,7], especially those based on data collected prior to emergence of the Omicron variant [16-18], suggesting that the prevalence of cough might be higher with the Omicron variant than with previous variants. In addition to cough, other symptoms such as fever, fatigue, myalgia, nasal congestion, and sore throat were also common, and over 60% of individuals reported having severe symptoms in our study.

With 15 amino acid substitutions in the receptor-binding domain, Omicron has been identified as being highly transmissible and possessing a remarkable capability of immune escape [19]. It has been reported that Omicron caused less damage to the lungs compared with Delta [20] and was more likely to target the upper airways [21]. The vagal sensory nerves play essential roles in regulating cough, which terminate primarily in the larynx, trachea, carina, and large intrapulmonary bronchi [22]. This may explain the greater prevalence of cough among patients with Omicron variant infection. The other possibility for the
higher proportion of cough and other symptoms identified in this study might be related to the reduced herd immunity against SARS-CoV-2. After the outbreak of COVID-19 in early 2020, a very strict prevention and control policy was implemented in China; thus, most people have not been exposed to COVID-19 in recent years [23]. In addition, the latest pandemic outbreak in China occurred during the winter, when viruses multiply more rapidly and spread more quickly on the one hand and the immune response of the airway is decreased on the other hand.

Previous studies have reported that compared to female patients, male patients with COVID-19 are at a higher risk of hospitalization, with worse outcomes and a higher mortality rate, independent of age [24]. In our study, all participants had mild-to-moderate disease and had not been hospitalized for COVID-19. We found that the female participants reported a higher proportion of cough and more severe cough. There was also a female predominance in the group reporting a chronic cough, which could be related to the heightened cough sensitivity among female patients [25]. This female preponderance in the chronic cough population and enhanced cough sensitivity may be explained by sex-related differences in the central processing of cough sensation [26]. However, further studies are needed to determine whether female patients with COVID-19 also present with increased cough sensitivity.

In China, 25.6% of adults (50.7% of men and 1.9% of women) are smokers [27]. The high proportion of nonsmoking female respondents in our study may explain why the proportion of smokers in our data was lower than that of the general population. Smoking has been reported as a risk factor for severe symptoms and/or progression of COVID-19 [28,29], and the prevalence of chronic cough and sputum was reported to be greater among current smokers [30]. Interestingly, smokers tended to report a nonsevere cough in the current survey. Many cigarette smokers usually have a chronic cough, but they scarcely seek medical attention unless a change in the pattern or intensity of their cough is noticed [31]. Cough reflex sensitivity and sensitivity to airway irritation were reported to be significantly decreased in current smokers compared to those of a similar population of nonsmokers [32,33]. Smoking was also found to be accompanied by increased activation in brain regions known to be involved in both cough sensory processing and cough suppression. Therefore, the smoking-induced sensitization of central cough neural circuits could be offset by concurrently enhanced central suppression [32].

Cough is a frequent and common complaint of patients with respiratory diseases such as asthma, chronic obstructive pulmonary disease, and interstitial lung disease. Nevertheless, cough could also be a sole or predominant symptom in extrapulmonary disorders, particularly GERD [34,35]. Most of the available evidence indicates that patients with respiratory diseases or cardiovascular diseases have the greatest odds of cough [9,36,37]. We also found an association of comorbidities, including asthma, chronic cough, coronary heart disease, GERD, and/or diabetes, with severe cough, persistent cough, and productive cough. Many studies have suggested a significant increase in the severity and mortality of COVID-19 in people with diabetes [38]. Self-reported chronic cough/phlegm was more common in patients with diabetes compared to that of the general population [39]. Glititin, a dipeptidyl peptidase-4 inhibitor used to treat diabetes mellitus, can cause cough as a side effect, which may be attributed to the persistent cough of COVID-19 [40]. In addition, diabetes, as an important risk factor, was found to influence the clinical severity of concomitant viral, bacterial, and fungal infections [41]. Therefore, attention should be paid to the possibility of coinfection or secondary bacterial infection in cases of persistent cough among patients with diabetes and COVID-19.

Approximately one-third of the individuals surveyed in this study reported having a productive cough, indicating the hypersecretion of mucus. In addition, the group with productive cough had a low proportion of individuals with respiratory comorbidities, which could not fully explain the excessive phlegm production. Excessive mucus production and accumulation is a common response to respiratory viral infections such as respiratory syncytial virus, rhinovirus, influenza virus, and SARS virus [42]. Recently, Kato et al [43] reported that SARS-CoV-2 induced the production of mucus (MUC5AC and MUC5B) throughout all airway regions via epidermal growth factor receptor/interleukin-1 receptor signaling activation in patients with COVID-19 [43]. This overproduction of airway mucin may increase the viscosity of the sputum, rendering it difficult to cough up. In addition, respiratory viruses infecting the human airway would impair mucociliary clearance, activate transient receptor potential vanilloid 1, induce the hypertrophy and metaplasia of goblet cells in the airway epithelium and the consequent enlargement of bronchial submucosal glands, leading to excessive fluid secretion and mucus plugs formation [42,44]. Intriguingly, we also observed that many individuals, especially the female participants, coughed up yellow or purulent sputum, suggesting coexistence of a bacterial infection. A previous study also suggested female sex as a susceptibility factor for bacterial infection [45]. However, it remains unclear whether these mechanisms play a pivotal role in mucus hypersecretion during Omicron infection.

Over 91% of individuals in our study had received a vaccine against SARS-CoV-2; however, we did not find a protective effect of the vaccine toward alleviating cough and mucus hypersecretion. The inactivated vaccines against SARS-CoV-2 were used in China; however, less than 15% of individuals had received a vaccine <6 months before the latest SARS-CoV-2 wave. Malik et al [46] demonstrated that the neutralizing antibody titer declines to near or below the seropositive threshold of a protective effect 6 months after vaccination with the inactivated vaccine. In addition, SARS-CoV-2 has been constantly mutating and most of the mutations are associated with the spike protein, a vital target for the action of vaccines [46]. Although a booster dose of the inactivated vaccine was associated with a lower risk of Omicron infection [47], compared with the ancestor strain, the production of neutralizing antibodies against Omicron was considerably impaired [48,49]. Another possible effect of the vaccine is the risk of antibody-dependent enhancement. According to in vitro studies, antibody-dependent enhancement might occur in COVID-19 infection [50,51], but current evidence indicates that this is an unlikely side effect induced by the vaccine [52,53].
There are some limitations of our study. First, all participants were nonhospitalized individuals with COVID-19, which may not fully reflect the cough profiles of severe cases of SARS-CoV-2 infection during the Omicron outbreak. Second, the data were obtained via online survey. Participants with low literacy or those of older age may have been neglected given the difficulties in filling out the survey and/or using a smartphone. In addition, clinical features and medical history were assessed by questionnaires, which may be prone to recall bias. However, most of the participants had been infected with SARS-CoV-2 within 1 month prior to completing the survey and many participants were still in the acute phase of COVID-19 at the time of the survey; hence, the impact of recall bias should be small. Lastly, we are following up with these respondents as we have only analyzed cross-sectional data; since many of the participants were still in acute phase of Omicron infection, the prevalence of persistent cough should be surveyed further.

In conclusion, over 90% of nonhospitalized individuals infected with the Omicron variant of SARS-CoV-2 presented with cough as a major symptom, with approximately one-third of individuals having a severe or productive cough, commonly with white thick and yellow purulent sputum, which developed into a persistent cough in a portion of these individuals. Female sex, chronic cough, asthma, GERD, diabetes, smoking, and/or vaccination for COVID-19 were associated with severe cough, productive cough, and/or persistent cough. Further studies are warranted to investigate the mechanism, treatment, and long-term prognosis of cough associated with Omicron variant infection, which will provide insight into how to treat SARS-CoV-2–induced cough.

Acknowledgments
The study was conducted under the support of China Cough Coalition. The authors thank Lan Li, Minghang Wang, Yi Jiang, Tao Bian, Qiang Chen, Qingyue Wu, Jianling Ma, Hua Xie, Qiao Zhang, Jianjun Zhao, Jianwen Fei, Qing Yang, Wei Han, Xuemei Wei, Zhaozhong Cheng, Qingwei Meng, Ling Liu, Chunhua Wei, Huiming Yin, Yuanrong Dai, Yan Wang, Xuekui Du, Liping Chen, Luqing Wei, Xiaowen Han, Liming Cai, Fuhong Li, Yuejuan Feng, Yong Lin, Fenfeng Qian, Linglei Kong, Jianxin Sun, Limei Geng, Xiaoli Hu, Jiantao Hu, and all committee members of China Cough Coalition. The corresponding author KL is on behalf of the Group China Cough Coalition. We also sincerely thank Chunxing Guo, Yawen He, Qiinyi Zhang, Qingchun Tang, Haodong Wu, Minting Liang, and all the health care staff for assistance with participant recruitment. We are grateful to all the patients who participated in this study. This study was supported by the National Key R&D Program of China (2021YFF0604000). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability
The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Survey on cough characteristics and prognosis of COVID-19: Questionnaire V1.0.
[DOCX File, 23 KB - publichealth_v10i1e47453_app1.docx ]

Multimedia Appendix 2
Number of questionnaires from 31 provinces of the Chinese mainland. A darker blue color indicates more questionnaires.
[PNG File, 302 KB - publichealth_v10i1e47453_app2.png ]

References
neutralizing antibodies. Nature 2022 Feb;602(7898):657-663 [FREE Full text] [doi: 10.1038/s41586-021-04385-3] [Medline:
35016194]
14. Lee SW. Methods for testing statistical differences between groups in medical research: statistical standard and guideline
10.3201/eid2802.212422
10. Marchant JM, Newcombe PA, Juniper EF, Sheffield JK, Stathis SL, Chang AB. What is the burden of chronic cough for
36224705]
36238984]
2. Canning BJ, Chang AB, Bolser DC, Smith JA, McGarvey L, CHEST Expert Cough Panel. Anatomy and


Abbreviations

GERD: gastroesophageal reflux disease
MUC: mucin
VAS: visual analog scale
Prevalence and Correlates of Dietary and Nutrition Information Seeking Through Various Web-Based and Offline Media Sources Among Japanese Adults: Web-Based Cross-Sectional Study

Kentaro Murakami¹, BEd, MSc, PhD; Nana Shinozaki², MPH, RD, PhD; Tsuyoshi Okuhara³, MPH, PhD; Tracy A McCaffrey⁴, BSc, PhD; M Barbara E Livingstone⁵, BEd, MSc, PhD

¹Department of Social and Preventive Epidemiology, School of Public Health, The University of Tokyo, Tokyo, Japan
²Department of Nutritional Epidemiology and Behavioural Nutrition, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
³Department of Health Communication, School of Public Health, The University of Tokyo, Tokyo, Japan
⁴Department of Nutrition, Dietetics and Food, Monash University, Clayton, Australia
⁵Nutrition Innovation Centre for Food and Health, School of Biomedical Sciences, Ulster University, Coleraine, United Kingdom

Corresponding Author:
Kentaro Murakami, BEd, MSc, PhD
Department of Social and Preventive Epidemiology
School of Public Health
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, 113 0033
Japan
Phone: 81 3 5841 7872
Fax: 81 3 5841 7873
Email: kenmrkm@m.u-tokyo.ac.jp

Abstract

Background: The advent of the internet has changed the landscape of available nutrition information. However, little is known about people’s information-seeking behavior toward healthy eating and its potential consequences.

Objective: We aimed to examine the prevalence and correlates of nutrition information seeking from various web-based and offline media sources.

Methods: This cross-sectional study included 5998 Japanese adults aged 20 to 79 years participating in a web-based questionnaire survey (February and March 2023). The dependent variable was the regular use of web-based and offline media as a reliable source of nutrition information. The main independent variables included health literacy, food literacy, and diet quality, which were assessed using validated tools, as well as sociodemographic factors (sex, age, education level, and nutrition- and health-related occupations).

Results: The top source of nutrition information was television (1973/5998, 32.89%), followed by web searches (1333/5998, 22.22%), websites of government and medical manufacturers (997/5998, 16.62%), newspapers (901/5998, 15.02%), books and magazines (697/5998, 11.62%), and video sites (eg, YouTube; 634/5998, 10.57%). Multivariable logistic regression showed that higher health literacy was associated with higher odds of using all the individual sources examined; odds ratios (ORs) for 1-point score increase ranged from 1.27 (95% CI 1.09-1.49) to 1.81 (95% CI 1.57-2.09). By contrast, food literacy was inversely associated with the use of television (OR 0.65, 95% CI 0.55-0.77), whereas it was positively associated with the use of websites of government and medical manufacturers (OR 1.98, 95% CI 1.62-2.44), books and magazines (OR 2.09, 95% CI 1.64-2.66), and video sites (OR 1.53, 95% CI 1.19-1.96). Furthermore, diet quality was positively associated with the use of newspapers (OR 1.02, 95% CI 1.01-1.03) and books and magazines (OR 1.03, 95% CI 1.02-1.04). Being female was associated with using television and books and magazines, whereas being male was associated with using websites of government and medical manufacturers, newspapers, and video sites. Age was positively associated with using newspapers and inversely associated with using websites of government and medical manufacturers and video sites. People with higher education were more likely to refer to websites of government and medical manufacturers and newspapers but less likely to use television and video sites. Dietitians were more likely to use websites of government and medical manufacturers and books and magazines than the general public but less likely to use television and video sites.
Conclusions: We identified various web-based and offline media sources regularly used by Japanese adults when seeking nutrition information, and their correlates varied widely. A lack of positive associations between the use of the top 2 major sources (television and web searches) and food literacy or diet quality is highlighted. These findings provide useful insights into the potential for developing and disseminating evidence-based health promotion materials.

(Keywords) nutrition; diet; information seeking; health literacy; food literacy; diet quality; Japan

Introduction

Background

On a global scale, poor diet quality is a major risk factor for premature mortality and morbidity, accounting for 22% of total deaths and 15% of disability-adjusted life years annually; these estimates are even higher in East Asian countries, including Japan (30% and 21%, respectively) [1]. As diet is a key modifiable risk factor, it is not surprising that the focus on prevention of chronic diseases has become a priority [2], mainly by reducing the intake of foods high in sodium and added sugars (eg, ultraprocessed foods) [3] and increasing the intake of plant-based foods, including whole grains, nuts, legumes, fruits, and vegetables [4,5]. This robust evidence base connects food to human health and should satisfy the growing public demand for dietary and nutrition information.

However, the advent of the internet has permanently changed the landscape of information and its use [6,7]. In particular, the introduction of Web 2.0, in which people can create, edit, and share content, has irrevocably changed communication from that via initial static web pages (Web 1.0) and traditional media (eg, newspapers, television, and radio) [8,9]. For a large proportion of people, the internet has become an essential tool for improving health-related knowledge and behavior, including nutrition [10,11]. For example, a national survey estimated that 78% of Japan’s general population uses the internet, with 73% of them seeking health information [12]. However, owing to the lack of regulations and ease of creating and sharing web-based information, consumers have access to an abundance of web-based information of variable quality and accuracy [13,14]. Consequently, nutrition and health professionals increasingly struggle with disseminating evidence-based nutrition information in a web-based media landscape filled with unreliable and contradictory nutrition information (ie, misinformation) [14-17].

It is reasonable to assume that the level of health literacy (ie, the cognitive and social skills that determine an individual’s motivation and ability to access, understand, and use information in ways that promote and maintain health [18]) is probably related to the search for nutrition information and thus the acquisition of nutrition knowledge [19]. However, an appropriate level of health literacy does not automatically lead to proper nutrition knowledge, particularly regarding aspects related to correct dietary decision-making [20]. Consequently, increasing attention is now being directed toward food literacy, defined as “a collection of inter-related knowledge, skills and behaviors required to plan, manage, select, prepare and eat food to meet needs and determine intake” [21].
Methods

Study Procedure and Participants
This cross-sectional study was based on data obtained from a web-based questionnaire survey. This paper was prepared in accordance with the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) [55], where applicable. Figure 1 shows a flow diagram of participant selection. The target sample consisted of 6600 Japanese adults aged 20 to 79 years, including not only the general public but also health professionals allied to nutrition (e.g., dietitians, registered dietitians, physicians, and dentists). Exclusion criteria were age outside the range of 20 to 79 years and working as an unrelated health professional (e.g., veterinarians, dental hygienists, assistant nurses, clinical psychologists, and nurse practitioners).

![Flow diagram of participant selection](https://publichealth.jmir.org/2024/1/e54805)

Registered panelists aged 20 to 79 years of the internet research agency, Rakuten Insight (n=2,603,155)

<table>
<thead>
<tr>
<th>Random selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered panelists aged 20 to 79 years to whom an invitation email was sent (n=676,329)</td>
</tr>
<tr>
<td>Individuals who did not answer the screening questions (n=599,484)</td>
</tr>
<tr>
<td>Registered panelists aged 20 to 79 years who answered the screening questions (n=76,845)</td>
</tr>
<tr>
<td>Individuals who did not or could not proceed to the main survey (n=69,123)</td>
</tr>
<tr>
<td>Registered panelists aged 20 to 79 years who proceeded to the main survey (n=7722)</td>
</tr>
<tr>
<td>Individuals who declined participation during the main survey (n=1122)</td>
</tr>
</tbody>
</table>

Registered panelists aged 20 to 79 years who completed the main survey (n=6600)

<table>
<thead>
<tr>
<th>Number of participants allocated to each category of nutrition- and health-related occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>• None (i.e., general public; not including health professionals, namely dietitians, registered dietitians, physicians, dentists, and other health professionals, and those working with nongovernmental qualification related to food and nutrition or in media)</td>
</tr>
<tr>
<td>• Male individuals aged 20 to 29 years (n=275)</td>
</tr>
<tr>
<td>• Male individuals aged 30 to 39 years (n=275)</td>
</tr>
<tr>
<td>• Male individuals aged 40 to 49 years (n=275)</td>
</tr>
<tr>
<td>• Male individuals aged 50 to 59 years (n=275)</td>
</tr>
<tr>
<td>• Male individuals aged 60 to 69 years (n=275)</td>
</tr>
<tr>
<td>• Male individuals aged 70 to 79 years (n=275)</td>
</tr>
<tr>
<td>• Female individuals aged 20 to 29 years (n=275)</td>
</tr>
<tr>
<td>• Female individuals aged 30 to 39 years (n=275)</td>
</tr>
<tr>
<td>• Female individuals aged 40 to 49 years (n=275)</td>
</tr>
<tr>
<td>• Female individuals aged 50 to 59 years (n=275)</td>
</tr>
<tr>
<td>• Female individuals aged 60 to 69 years (n=275)</td>
</tr>
<tr>
<td>• Female individuals aged 70 to 79 years (n=275)</td>
</tr>
<tr>
<td>• Health professionals</td>
</tr>
<tr>
<td>• Dietitian and registered dietitian (n=660)</td>
</tr>
<tr>
<td>• Physician (n=528)</td>
</tr>
<tr>
<td>• Dentist (n=132)</td>
</tr>
<tr>
<td>• Nurse, midwife, or public health nurse (n=528; combined with other health professionals in analysis)</td>
</tr>
<tr>
<td>• Pharmacist (n=396; combined with other health professionals in analysis)</td>
</tr>
<tr>
<td>• Others</td>
</tr>
<tr>
<td>• Nongovernmental qualification related to food and nutrition (n=660)</td>
</tr>
<tr>
<td>• Media (n=396)</td>
</tr>
<tr>
<td>Individuals whose data were considered unreliable (n=602)</td>
</tr>
</tbody>
</table>

Registered panelists aged 20 to 79 years who were included in the present analysis (N=5998)
Data collection was conducted by an internet research agency, Rakuten Insight [56]. With 2.2 million monitors nationwide, Rakuten Insight monitors fraudulent registrants, such as impersonators and duplicate registrants, by sharing the basic registration information of members. Furthermore, as part of the quality control of survey results, Rakuten Insight implemented an automated computerized checking system. Participants were recruited from among the registered panelists of the research agency. An email with an invitation to participate and a web page link to the survey was sent to a randomized list (n=676,329, 26.1%) of registered panelists aged 20 to 79 years (n=2,603,155; unfortunately, information on the sociodemographic characteristics of all the registered panelists in Rakuten Insight is not available [56]). Initially, a study summary was provided, and only individuals who agreed to participate could proceed to the screening stage (76,845/676,329, 11.36%). As recruitment was stratified by age, sex, and occupation, participants could proceed to the main survey only if there was a vacancy in the relevant category (eg, male individuals aged 20 to 79 years; Figure 1). As a result, among the 76,845 individuals in the screening stage, 7722 (10.05%) proceeded to the main survey, of whom 1122 (14.53%) did not complete all questions. Data collection started on February 10, 2023, and ended on March 16, 2023, when quotas for the main survey were filled (n=6600). The sample size was determined primarily based on feasibility and financial limitations. For analysis, we excluded individuals whose answers were considered unreliable based on their answer to the following question: “This question is for the purpose of investigating ‘misplaced’ answers when responding to the survey. Please select neither agree or disagree from the following options.” As a result, of the 6600 individuals who completed the survey, 314 (4.76%) individuals whose responses were strongly agree, agree, disagree, or strongly disagree were excluded. Further, we excluded individuals whose data were considered unreliable with regard to body height (≥200 cm; 2/6600, 0.03%) or energy intake (<800 or >4200 kcal/day for male individuals and <500 or >3500 kcal/day for female individuals [57]; 286/6600, 4.33%). Consequently, the final analysis sample comprised 5998 individuals (male individuals: n=2687, 44.8%; female individuals: n=3311, 55.2%) aged 20 to 79 years. The respondents included in the present analysis (N=5998) differed somewhat from those excluded from the analysis (n=602). The excluded respondents were more likely to be male and have higher means of age and BMI, lower household income, and nongovernmental qualifications related to food and nutrition.

Ethical Considerations

The study was conducted according to the guidelines of the Declaration of Helsinki, and all procedures involving human participants were approved by the Ethics Committee of the University of Tokyo Faculty of Medicine (protocol code: 20222288NI; date of approval: January 13, 2023). Informed consent was obtained on the website from all individuals involved in the study. The participants were compensated with standard Rakuten Insight incentives (points redeemable for cash or merchandise) for completing the survey.

Assessment of Sociodemographic Characteristics

All the questions used in this study were prepared by the first, second, and third authors. Sex (assigned at birth) was self-selected as either male or female. Age (in years) was also self-reported. Body weight and height were self-reported and used to calculate BMI (in kg/m²), with 3 weight status categories: underweight (<18.5), normal weight (≥18.5 to <25), and overweight (≥25) [58]. The following variables were also used in this study (categorization shown in parentheses): education level (junior high or high school, junior college or technical school, university or higher, and others), household income (<4 million Japanese yen, 4 to 7 million Japanese yen, >7 million Japanese yen, and unknown or do not want to answer; US $1=JPY 148.22), employment status (none, student, part-time job, and full-time job), marital status (unmarried, married, and do not want to answer), living alone (no and yes), presence of chronic diseases (eg, hypertension, hyperlipidemia, and diabetes; no and yes), and smoking status (never, past, and current). In addition, based on the reported postal code of the home address, each participant was grouped into 1 of 6 regions (Hokkaido and Tohoku, Kanto, Hokuriku and Tokai, Kinki, Chugoku and Shikoku, and Kyushu). Participants were also grouped into 1 of 3 municipality levels (ward, city, and town and village). The category missing was created for those whose reported postal codes were incomplete for the municipality level variable. Further, participants were categorized according to nutrition and health-related occupations (Figure 1): none (ie, the general public), nongovernmental qualification related to food and nutrition, media, dietitian and registered dietitian, physician and dentist, and other health professional (ie, nurse, midwife, public health nurse, and pharmacist).

Assessment of Health Literacy

Health literacy was assessed using the Communicative and Critical Health Literacy scale, which was developed and validated for the Japanese population [48]. On the basis of a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree), each participant was asked whether they were able to (1) collect health-related information from various sources, such as newspapers, books, television, and the internet; (2) extract the information they wanted; (3) understand and communicate the obtained information; (4) consider the credibility of the information; and (5) make decisions based on the information, specifically in the context of health-related issues. The health literacy score was calculated as the average of all items, meaning that the higher the score, the higher the health literacy (with possible scores ranging from 1 to 5) [48].

Assessment of Food Literacy

As described in detail in Multimedia Appendix 1 [49], food literacy was assessed using the Japanese version of the 29-item Dutch Self-Perceived Food Literacy scale [49]. This is an expert-based, theory-driven, and validated tool for measuring food literacy with respect to healthy eating and focuses on 8 domains: food preparation skills (6 items), resilience and resistance (6 items), healthy snack styles (4 items), social and conscious eating (3 items), examining food labels (2 items), daily food planning (2 items), healthy budgeting (2 items), and healthy food stockpiling (4 items) [49]. Participants were asked
to answer all the questions based on a 5-point Likert scale (1=not at all or never, 5=always or always). The food literacy score was calculated as the average of all items, with negative items reversed, indicating that the higher the score, the higher the food literacy (with possible scores ranging from 1 to 5) [49].

Assessment of Diet Quality

As a measure of diet quality, the Healthy Eating Index (HEI)-2020 [50] was used. The HEI-2020 is an established, 100-point scale to assess compliance with the 2020 to 2025 Dietary Guidelines for Americans [2], with a higher score indicating a better quality of overall diet. The HEI-2020 consists of 9 adequacy components (eg, fruits, vegetables, and whole grains) and 4 moderation components (eg, sodium and added sugars). The efficacy of the HEI-2015, which completely aligns with the HEI-2020 [50], in assessing the overall diet quality of Japanese people has been supported by our previous analyses [59,60]. Dietary intake information was collected using a validated, short version of the Meal-based Diet History Questionnaire [51,52,61]. A detailed description of the diet quality assessment is presented in Multimedia Appendix 2 [2,50-52,59-63].

Assessment of Dietary and Nutrition Information Seeking Through Various Media Sources

The assessment of dietary and nutrition information seeking through various web-based and offline media sources was conducted using a series of 2 questions, as informed by previous studies [11,25,27,28,30,35-38,64]. Participants were first asked the following question: “Which of the following sources do you routinely use as a source of information about diet and nutrition? Please select all that apply (as many as you want).” The list of sources of dietary and nutrition information shown was as follows: (1) television, (2) radio, (3) newspapers, (4) books and magazines, (5) municipal newsletters, (6) websites of government and medical manufacturers, (7) web searches, (8) news applications, (9) video sites (eg, YouTube [Google LLC]), and (10) social networking sites (eg, Twitter [Twitter Inc], Instagram [Meta Platforms Inc], and Facebook [Meta Platforms Inc]); 3 other face-to-face information sources (ie, hospitals and drug stores, family, and friends) were also shown, which are beyond the scope of this analysis. These were based on items used in a national survey conducted by the Ministry of Internal Affairs and Communications [65]. Participants were then asked the second question: “Do you think that dietary and nutrition information obtained from the sources listed below is reliable?” The list of dietary and nutrition information sources displayed was customized for each participant and contained only the sources the participant chose in the first question. For each of the dietary and nutrition information sources, a 5-point Likert scale category was provided (strongly disagree, agree, neither agree or disagree, agree, and strongly agree). Sources used to seek dietary and nutrition information were defined based on participants’ responses (agree or strongly agree) regarding the reliability of the sources.

Statistical Analysis

All statistical analyses were performed using the SAS statistical software (version 9.4; SAS Institute Inc). Descriptive data are presented as frequencies and percentages of participants for categorical variables and means and SDs for continuous variables. Associations among the top media sources (defined as a priori as >10% prevalence) used for seeking dietary and nutrition information were examined using a chi-square test. Associations of dietary and nutrition information seeking through the top media sources with participant characteristics were also examined using a chi-square test. Finally, multivariable odds ratios and their 95% CIs for dietary and nutrition information seeking through each media source according to participant characteristics were calculated using multiple logistic regression. For each media source, a model was constructed, which included dietary and nutrition information seeking through the media source as the dependent variable and sex (reference: male), age (continuous), weight status (reference: normal weight), education level (reference: junior high or high school), household income (reference: <4 million Japanese yen; US $1=JPY 148.22), employment status (reference: none), marital status (reference: unmarried), living alone (reference: no), presence of chronic disease (reference: no), smoking status (reference: never), region (reference: Kanto), municipality level (reference: ward), nutrition- and health-related occupations (reference: none, ie, general public), health literacy score (continuous), food literacy score (continuous), HEI-2020 score (continuous), and dietary and nutrition information seeking through 5 other media sources (reference: no for each) as the independent variables. The analysis was repeated separately for the general public and health professionals allied to nutrition. We considered 2-tailed P values <.05 statistically significant. We decided to treat age, health literacy score, and HEI-2020 score as continuous variables in our final analyses after observing similar findings when these variables were treated as categorical variables (20 to 39, 40 to 59, and 60 to 79 years for age and quartiles for others; data not shown).

Results

Basic Characteristics of Study Participants

This analysis included 5998 individuals (male individuals: n=2687, 44.8%; female individuals: n=3311, 55.2%) aged 20 to 79 years (Table 1).
Table 1. Basic characteristics of the study participants (N=5998).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2687 (44.8)</td>
</tr>
<tr>
<td>Female</td>
<td>3311 (55.2)</td>
</tr>
<tr>
<td><strong>Age (y), mean (SD)</strong></td>
<td>46.8 (15.1)</td>
</tr>
<tr>
<td><strong>Weight status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>783 (13.05)</td>
</tr>
<tr>
<td>Normal weight</td>
<td>4171 (69.54)</td>
</tr>
<tr>
<td>Overweight</td>
<td>1044 (17.41)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Junior high or high school</td>
<td>951 (15.86)</td>
</tr>
<tr>
<td>Junior college or technical school</td>
<td>1394 (23.24)</td>
</tr>
<tr>
<td>University or higher</td>
<td>3623 (60.4)</td>
</tr>
<tr>
<td>Other</td>
<td>30 (0.5)</td>
</tr>
<tr>
<td><strong>Household income, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;4 million Japanese yen</td>
<td>1130 (18.84)</td>
</tr>
<tr>
<td>4 to 7 million Japanese yen</td>
<td>1596 (26.61)</td>
</tr>
<tr>
<td>&gt;7 million Japanese yen</td>
<td>2297 (38.3)</td>
</tr>
<tr>
<td>Unknown or do not want to answer</td>
<td>975 (16.26)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>1010 (16.84)</td>
</tr>
<tr>
<td>Student</td>
<td>87 (1.45)</td>
</tr>
<tr>
<td>Part-time job</td>
<td>1025 (17.09)</td>
</tr>
<tr>
<td>Full-time job</td>
<td>3876 (64.62)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>2305 (38.43)</td>
</tr>
<tr>
<td>Married</td>
<td>3629 (60.5)</td>
</tr>
<tr>
<td>Do not want to answer</td>
<td>64 (1.07)</td>
</tr>
<tr>
<td><strong>Living alone, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>4682 (78.06)</td>
</tr>
<tr>
<td>Yes</td>
<td>1316 (21.94)</td>
</tr>
<tr>
<td><strong>Presence of chronic disease, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3849 (64.17)</td>
</tr>
<tr>
<td>Yes</td>
<td>2149 (35.83)</td>
</tr>
<tr>
<td><strong>Smoking status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>3886 (64.79)</td>
</tr>
<tr>
<td>Past</td>
<td>1213 (20.22)</td>
</tr>
<tr>
<td>Current</td>
<td>899 (14.99)</td>
</tr>
<tr>
<td><strong>Region, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Hokkaido and Tohoku</td>
<td>608 (10.14)</td>
</tr>
<tr>
<td>Kanto</td>
<td>2377 (39.63)</td>
</tr>
<tr>
<td>Tokai and Hokuriku</td>
<td>901 (15.02)</td>
</tr>
</tbody>
</table>
The mean BMI was 22.1 (SD 3.6) kg/m²; 69.54% (4171/5998) of the participants were categorized as having normal weight. Most participants had a high educational background (university or higher: 3623/5998, 60.4%) and full-time jobs (3876/5998, 64.62%) and were never (3886/5998, 64.79%) or past (1213/5998, 20.22%) smokers. Owing to the sampling design, approximately half (3021/5998, 50.37%) of the participants were members of the general public, whereas the remainder were in nutrition- and health-related occupations, such as dietitians and registered dietitians (631/5998, 10.52%) and physicians and dentists (602/5998, 10.04%).

When participants were allowed to choose with multiple options, the prevalence of diet and nutrition information seeking through various media sources among Japanese adults was as follows: 32.89% (1973/5998) for television, 5.3% (318/5998) for radio, 15.02% (901/5998) for newspapers, 11.62% (697/5998) for books and magazines, 3.52% (211/5998) for municipal newsletters, 16.62% (997/5998) for websites of government and medical manufacturers, 22.22% (1333/5998) for web searches, 9.85% (591/5998) for news applications, 10.57% (634/5998) for video sites (eg, YouTube), and 8.65% (519/5998) for social networking sites (eg, Twitter, Instagram, and Facebook). Thus, the major sources of information (>10% of prevalence) were television (1973/5998, 32.89%), web searches (1333/5998, 22.22%), websites of government and medical manufacturers (997/5998, 16.62%), newspapers (901/5998, 15.02%), books and magazines (697/5998, 11.62%), and video sites (eg, YouTube; 634/5998, 10.57%). These were then examined in subsequent analyses as the top media sources. Dietary and nutrition information seeking through these top 6 media sources was significantly associated with each other, except for a null association between dietary and nutrition information seeking through newspapers and dietary and nutrition information seeking through video sites (Multimedia Appendix 3). Dietary and nutrition information seeking via the top 6 media sources differed significantly according to participant characteristics, with a few exceptions, such as weight status, region, and municipality level (Multimedia Appendix 4).

**Associations of Participant Characteristics With Diet and Nutrition Information Seeking Through the Top 6 Media Sources**

Multivariable odds ratios (with their 95% CIs) for dietary and nutrition information seeking through each media source are shown in Table 2 (for television, web searches, and websites of government and medical manufacturers) and Table 3 (for newspapers, books and magazines, and video sites).
Table 2. Associations of participant characteristics with diet and nutrition information seeking through the top 3 media sources (ie, television, web searches, and websites of government and medical manufacturers) among Japanese adults (N=5998).a

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Television, OR (95% CI)b</th>
<th>Web searches, OR (95% CI)d</th>
<th>Websites of government and medical manufacturers, OR (95% CI)f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex (reference: male)</td>
<td>1.68 (1.44-1.97)</td>
<td>0.95 (0.80-1.14)</td>
<td>0.78 (0.64-0.95)</td>
</tr>
<tr>
<td>Age (per 1-year increment)</td>
<td>1.00 (0.996-1.01)</td>
<td>1.00 (0.998-1.01)</td>
<td>0.987 (0.980-0.994)</td>
</tr>
<tr>
<td>Weight status (reference: normal weight)f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>0.94 (0.78-1.13)</td>
<td>1.31 (1.07-1.60)</td>
<td>1.07 (0.86-1.34)</td>
</tr>
<tr>
<td>Overweight</td>
<td>0.98 (0.83-1.16)</td>
<td>0.93 (0.77-1.12)</td>
<td>1.07 (0.87-1.33)</td>
</tr>
<tr>
<td>Education level (reference: junior high or high school)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior college or technical school</td>
<td>0.84 (0.69-1.03)</td>
<td>1.08 (0.86-1.36)</td>
<td>1.12 (0.84-1.50)</td>
</tr>
<tr>
<td>University or higher</td>
<td>0.77 (0.64-0.93)</td>
<td>0.89 (0.72-1.09)</td>
<td>1.41 (1.08-1.83)</td>
</tr>
<tr>
<td>Other</td>
<td>0.36 (0.13-1.01)</td>
<td>0.48 (0.14-1.71)</td>
<td>1.37 (0.41-4.61)</td>
</tr>
<tr>
<td>Household income (reference: &lt;4 million JPY)f</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 to 7 million JPY</td>
<td>1.01 (0.84-1.22)</td>
<td>1.06 (0.86-1.31)</td>
<td>1.07 (0.84-1.37)</td>
</tr>
<tr>
<td>&gt;7 million JPY</td>
<td>1.10 (0.90-1.34)</td>
<td>0.87 (0.69-1.09)</td>
<td>1.12 (0.86-1.44)</td>
</tr>
<tr>
<td>Unknown or do not want to answer</td>
<td>0.90 (0.72-1.12)</td>
<td>0.80 (0.62-1.02)</td>
<td>0.95 (0.71-1.26)</td>
</tr>
<tr>
<td>Employment status (reference: none)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1.41 (0.82-2.43)</td>
<td>0.39 (0.19-0.81)</td>
<td>1.33 (0.69-2.55)</td>
</tr>
<tr>
<td>Part-time job</td>
<td>1.14 (0.92-1.41)</td>
<td>1.07 (0.84-1.37)</td>
<td>0.75 (0.56-1.01)</td>
</tr>
<tr>
<td>Full-time job</td>
<td>1.02 (0.83-1.26)</td>
<td>1.02 (0.81-1.30)</td>
<td>1.05 (0.80-1.39)</td>
</tr>
<tr>
<td>Marital status (reference: unmarried)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1.23 (1.04-1.44)</td>
<td>1.07 (0.89-1.29)</td>
<td>0.83 (0.68-1.01)</td>
</tr>
<tr>
<td>Do not want to answer</td>
<td>0.86 (0.44-1.69)</td>
<td>0.50 (0.21-1.20)</td>
<td>0.33 (0.11-1.02)</td>
</tr>
<tr>
<td>Living alone (reference: no)</td>
<td>0.80 (0.66-0.97)</td>
<td>1.17 (0.95-1.44)</td>
<td>0.96 (0.76-1.20)</td>
</tr>
<tr>
<td>Presence of chronic disease (reference: no)</td>
<td>0.91 (0.79-1.04)</td>
<td>1.00 (0.86-1.16)</td>
<td>1.30 (1.10-1.54)</td>
</tr>
<tr>
<td>Smoking status (reference: never)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>0.87 (0.73-1.02)</td>
<td>1.05 (0.88-1.26)</td>
<td>1.01 (0.83-1.24)</td>
</tr>
<tr>
<td>Current</td>
<td>0.93 (0.77-1.12)</td>
<td>1.03 (0.84-1.26)</td>
<td>0.84 (0.66-1.06)</td>
</tr>
<tr>
<td>Region (reference: Kanto)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hokkaido and Tohoku</td>
<td>1.01 (0.81-1.25)</td>
<td>0.87 (0.69-1.11)</td>
<td>0.94 (0.72-1.22)</td>
</tr>
<tr>
<td>Tokai and Hokuriku</td>
<td>1.12 (0.93-1.35)</td>
<td>1.00 (0.82-1.23)</td>
<td>0.88 (0.70-1.10)</td>
</tr>
<tr>
<td>Kinki</td>
<td>1.06 (0.89-1.26)</td>
<td>0.92 (0.76-1.12)</td>
<td>0.96 (0.77-1.18)</td>
</tr>
<tr>
<td>Chugoku and Shikoku</td>
<td>1.18 (0.93-1.50)</td>
<td>0.89 (0.68-1.16)</td>
<td>0.91 (0.68-1.22)</td>
</tr>
<tr>
<td>Kyushu</td>
<td>1.27 (1.02-1.58)</td>
<td>0.76 (0.59-0.98)</td>
<td>0.89 (0.67-1.16)</td>
</tr>
<tr>
<td>Municipality level (reference: ward)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>0.91 (0.80-1.03)</td>
<td>1.11 (0.96-1.28)</td>
<td>1.02 (0.87-1.20)</td>
</tr>
<tr>
<td>Town and village</td>
<td>0.91 (0.69-1.21)</td>
<td>0.82 (0.59-1.14)</td>
<td>1.22 (0.87-1.71)</td>
</tr>
<tr>
<td>Missing</td>
<td>0.89 (0.55-1.43)</td>
<td>0.94 (0.54-1.64)</td>
<td>0.97 (0.53-1.75)</td>
</tr>
<tr>
<td>Nutrition- and health-related occupations (reference: none, ie, general public)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nongovernmental qualification related to food and nutrition</td>
<td>0.61 (0.48-0.79)</td>
<td>0.87 (0.66-1.14)</td>
<td>1.02 (0.76-1.37)</td>
</tr>
<tr>
<td>Media</td>
<td>1.18 (0.91-1.54)</td>
<td>0.80 (0.59-1.10)</td>
<td>0.76 (0.53-1.09)</td>
</tr>
<tr>
<td>Dietitian and registered dietitian</td>
<td>0.57 (0.44-0.72)</td>
<td>0.88 (0.67-1.14)</td>
<td>2.60 (2.00-3.37)</td>
</tr>
<tr>
<td>Physician and dentist</td>
<td>0.55 (0.43-0.71)</td>
<td>1.17 (0.90-1.52)</td>
<td>1.29 (0.97-1.71)</td>
</tr>
<tr>
<td></td>
<td>Television, OR(^b) (95% CI)(^c)</td>
<td>Web searches, OR (95% CI)(^d)</td>
<td>Websites of government and medical manufacturers, OR (95% CI)(^e)</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Other health professional</td>
<td>0.89 (0.74-1.09)</td>
<td>0.92 (0.74-1.16)</td>
<td>1.43 (1.12-1.83)</td>
</tr>
<tr>
<td>Health literacy score</td>
<td>1.40 (1.25-1.57)</td>
<td>1.60 (1.40-1.82)</td>
<td>1.81 (1.57-2.09)</td>
</tr>
<tr>
<td>(per 1-point increment)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food literacy score</td>
<td>0.65 (0.55-0.77)</td>
<td>0.85 (0.70-1.02)</td>
<td>1.98 (1.62-2.44)</td>
</tr>
<tr>
<td>(per 1-point increment)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy Eating Index-2020</td>
<td>1.00 (0.993-1.01)</td>
<td>1.00 (0.994-1.01)</td>
<td>1.01 (0.999-1.02)</td>
</tr>
<tr>
<td>(per 1-point increment)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Odds ratio (OR) for diet and nutrition information seeking through each media source in comparison with the reference category of each variable. Statistically significant values are italicized (\(P<.05\)).

\(^b\)OR: odds ratio.

\(^c\)Model with diet and nutrition information seeking through television as the dependent variable and variables listed in the first column and diet and nutrition information seeking through 5 other media sources (web searches, websites of government and medical manufacturers, newspapers, books and magazines, and video sites; no or yes for each) as the independent variables.

\(^d\)Model with diet and nutrition information seeking through web searches as the dependent variable and variables listed in the first column and diet and nutrition information seeking through 5 other media sources (television, websites of government and medical manufacturers, newspapers, books and magazines, and video sites; no or yes for each) as the independent variables.

\(^e\)Model with diet and nutrition information seeking through websites of government and medical manufacturers as the dependent variable and variables listed in the first column and diet and nutrition information seeking through 5 other media sources (television, web searches, newspapers, books and magazines, and video sites; no or yes for each) as the independent variables.

\(^f\)Underweight, normal weight, and overweight were defined having BMIs of <18.5, ≥18.5 to <25, and ≥25 kg/m\(^2\), respectively.

\(^g\)US $1=JPY 148.22.
| Table 3. Associations of participant characteristics with diet and nutrition information seeking through the top 4th to 6th media sources (ie, newspapers, books and magazines, and video sites) among Japanese adults (N=5998)\(^{a}\). |
|---|---|---|
| | Newspapers, OR\(^{b}\) (95% CI)\(^{c}\) | Books and magazines, OR (95% CI)\(^{d}\) | Video sites (eg, YouTube), OR (95% CI)\(^{e}\) |
| Female sex (reference: male) | 0.69 (0.56-0.86) | 1.26 (1.001-1.59) | 0.66 (0.53-0.83) |
| Age (per 1-year increment) | 1.06 (1.05-1.07) | 1.00 (0.99-1.01) | 0.97 (0.96-0.98) |
| Weight status (reference: normal weight)\(^{f}\) | | | |
| Underweight | 1.29 (0.998-1.66) | 0.79 (0.61-1.03) | 0.84 (0.64-1.11) |
| Overweight | 0.70 (0.56-0.88) | 1.31 (1.02-1.67) | 0.90 (0.69-1.15) |
| Education level (reference: junior high or high school) | | | |
| Junior college or technical school | 0.95 (0.72-1.26) | 1.13 (0.82-1.55) | 0.96 (0.72-1.27) |
| University or higher | 1.45 (1.14-1.86) | 1.17 (0.87-1.57) | 0.67 (0.52-0.88) |
| Other | 2.10 (0.69-6.41) | 1.42 (0.40-5.06) | 1.23 (0.35-4.28) |
| Household income (reference: <4 million JPY)\(^{g}\) | | | |
| 4 to 7 million JPY | 1.01 (0.78-1.31) | 1.11 (0.84-1.46) | 1.03 (0.78-1.36) |
| >7 million JPY | 0.98 (0.74-1.29) | 1.15 (0.85-1.53) | 1.12 (0.83-1.51) |
| Unknown or do not want to answer | 0.85 (0.63-1.14) | 0.85 (0.61-1.18) | 0.82 (0.58-1.15) |
| Employment status (reference: none) | | | |
| Student | 1.32 (0.50-3.48) | 2.07 (0.99-4.31) | 0.92 (0.47-1.79) |
| Part-time job | 1.25 (0.95-1.66) | 1.10 (0.79-1.52) | 0.84 (0.60-1.16) |
| Full-time job | 1.26 (0.95-1.66) | 1.15 (0.84-1.59) | 0.75 (0.55-1.02) |
| Marital status (reference: unmarried) | | | |
| Married | 1.04 (0.82-1.31) | 0.86 (0.68-1.09) | 0.82 (0.64-1.03) |
| Do not want to answer | 0.67 (0.23-1.98) | 0.95 (0.37-2.44) | 1.35 (0.56-3.25) |
| Living alone (reference: no) | 0.51 (0.38-0.68) | 1.12 (0.86-1.46) | 1.15 (0.88-1.50) |
| Presence of chronic disease (reference: no) | 0.96 (0.80-1.15) | 1.03 (0.85-1.25) | 1.03 (0.84-1.27) |
| Smoking status (reference: never) | | | |
| Past | 0.88 (0.71-1.09) | 1.02 (0.81-1.30) | 1.44 (1.14-1.83) |
| Current | 0.79 (0.61-1.02) | 0.70 (0.52-0.95) | 1.33 (1.03-1.73) |
| Region (reference: Kanto) | | | |
| Hokkaido and Tohoku | 1.47 (1.11-1.96) | 1.09 (0.81-1.47) | 1.27 (0.94-1.72) |
| Tokai and Hokuriku | 1.27 (0.98-1.63) | 1.07 (0.82-1.39) | 1.06 (0.81-1.40) |
| Kinki | 1.16 (0.91-1.47) | 0.94 (0.74-1.21) | 0.93 (0.71-1.21) |
| Chugoku and Shikoku | 1.74 (1.28-2.37) | 0.96 (0.68-1.35) | 1.19 (0.84-1.68) |
| Kyushu | 1.41 (1.05-1.90) | 0.84 (0.61-1.17) | 1.36 (0.996-1.86) |
| Municipality level (reference: ward) | | | |
| City | 1.07 (0.90-1.28) | 0.90 (0.75-1.08) | 1.00 (0.82-1.21) |
| Town and village | 1.32 (0.92-1.89) | 0.84 (0.56-1.25) | 1.15 (0.77-1.72) |
| Missing | 0.80 (0.39-1.64) | 1.02 (0.53-1.95) | 1.14 (0.60-2.14) |
| Nutrition- and health-related occupations (reference: none, ie, general public) | | | |
| Nongovernmental qualification related to food and nutrition | 1.21 (0.87-1.69) | 1.64 (1.20-2.24) | 1.23 (0.91-1.66) |
| Media | 1.49 (1.07-2.07) | 1.35 (0.94-1.96) | 0.51 (0.32-0.81) |
| Dietitian and registered dietitian | 0.79 (0.55-1.12) | 1.97 (1.46-2.64) | 0.54 (0.37-0.79) |
| Physician and dentist | 0.63 (0.46-0.87) | 1.03 (0.72-1.46) | 1.03 (0.72-1.46) |
Table 1. Odds ratios (OR) for diet and nutrition information seeking through each media source in comparison with the reference category of each variable. Statistically significant values are italicized (P < .05).

<table>
<thead>
<tr>
<th>Media Source</th>
<th>Odds Ratio (OR) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspapers</td>
<td>0.64 (0.47-0.88)</td>
</tr>
<tr>
<td>Health literacy score</td>
<td>1.27 (1.09-1.49)</td>
</tr>
<tr>
<td>Food literacy score</td>
<td>1.14 (0.91-1.43)</td>
</tr>
<tr>
<td>Healthy Eating Index-2020</td>
<td>1.02 (1.01-1.03)</td>
</tr>
</tbody>
</table>

Note: OR: odds ratio.

Information seeking through watching television was associated with female sex, lower education, being married, living with someone, living in the Kyushu region (compared with living in the Kanto region), belonging to the general public (compared with having a nongovernmental qualification related to food and nutrition, being a dietitian or registered dietitian, and being a physician or dentist), a higher score of health literacy, and a lower score of food literacy. Web searches were associated with underweight (compared with normal weight), nonemployment status (compared with students), living in the Kanto region (compared with living in the Kyushu region), and a higher score of health literacy. Use of websites of government and medical manufacturers was associated with male sex; younger age; higher education; the presence of chronic disease; being a dietitian or registered dietitian, physician or dentist, or other health professional (compared with belonging to the general public); and higher scores for health literacy and food literacy. Seeking information from newspapers was associated with male sex; older age; higher education; living with someone; living in the Hokkaido and Tohoku, Chugoku and Shikoku, or Kyushu regions; working in the media (compared with belonging to the general public); and higher scores for health literacy and the HEI-2020. In addition, overweight and being a physician or dentist or other health professional (compared with belonging to the general public) were less likely to seek information from newspapers. Information seeking through books and magazines was associated with female sex; overweight; never smoking (compared with current smoking); working with nongovernmental qualifications related to food and nutrition or as a dietitian or registered dietitian (compared with belonging to the general public); and higher scores for health literacy, food literacy, and the HEI-2020. The use of video sites was associated with male sex, younger age, lower education, past and current smoking, belonging to the general public (compared with working in the media, as a dietitian or registered dietitian, or as other health professional), and higher scores for health literacy and food literacy. The analysis was repeated separately for the general public (Multimedia Appendix 5) and health professionals allied to nutrition (Multimedia Appendix 6). In general, similar findings were observed in approximately 90% (81/90) of the associations examined, although some did not reach statistical significance.

Discussion

Principal Findings

To our knowledge, this is the first study to comprehensively examine the prevalence and correlates of dietary and nutrition information seeking from web-based and offline media sources. In this cross-sectional study of 5998 Japanese adults aged 20 to 79 years, the principal dietary and nutrition information platforms used by >10% of participants were television (1973/5998, 32.89%), web searches (1333/5998, 22.22%), websites of government and medical manufacturers (997/5998, 16.62%), newspapers (901/5998, 15.02%), and video sites (697/5998, 11.62%), and books and magazines (634/5998, 10.57%). After adjustment for potential confounding factors, health literacy was positively associated with the use of each of these sources. By contrast, food literacy showed an inverse association with watching television, whereas it was positively associated with the use of websites of government and medical manufacturers, books and magazines, and video sites. In addition, the media sources that were positively associated with diet quality were websites and books and magazines. Regarding sociodemographic variables, being female was associated with watching television and reading books and magazines, whereas being male was more likely to be associated with using websites of government and medical manufacturers, newspapers, and video sites. Older people were more likely to read newspapers for nutrition information, whereas younger people were more attracted to websites of government and medical manufacturers and video sites. People with higher...
education levels were more likely to use websites of government and medical manufacturers and newspapers but less likely to rely on television and video sites. Compared with the general public, dietitians and registered dietitians were more likely to use websites of government and medical manufacturers and books and magazines but less likely to use television and video sites.

**Comparison With Prior Work**

To our knowledge, only a few studies have investigated the lifestyle (especially dietary) profiles of people seeking dietary and nutrition information from various media sources [25,26,35-40]. This type of study is of utmost importance, given not only that our eating habits have a significant impact on our health [1] but also that the general media landscape is filled with dietary misinformation [14-17]. In this study, after adjustment for potential confounding factors, nutrition information seeking from newspapers and books and magazines (but not television, web searches, websites of government and medical manufacturers, and video sites) was positively associated with diet quality. Thus, the association between nutrition information seeking and diet quality was only observed for offline media. This may be explained by the nature of offline media use, which permits the skimming, reading, and rereading of information, thereby encouraging an active learning process [26,35,38] that may lead to favorable dietary behaviors. Positive associations between the use of offline media and the consumption of fruits and vegetables have been consistently observed in previous studies [26,35,38], with the exception of 1 study [25]. Conversely, previous studies failed to find a significant association between the use of web-based media and fruit and vegetable consumption [25,26,36,37], with the exception of 1 study [35]. Null associations were also observed in this study despite our classification of internet use into several categories (ie, web searches, websites of government and medical manufacturers, video sites, news applications, and social networking sites). With regard to watching television, previous studies have generally not shown favorable associations with eating behaviors [25,26,35,38], which is consistent with the present findings. This fits well with the suggestion that television creates only a passive learning environment, which is unsuitable for the delivery of meaning, knowledge, or context [26,38].

In line with previous studies [19], we found that health literacy was positively associated with the use of all media sources for nutrition information seeking. A novel finding of our study was that food literacy was positively associated with nutrition information seeking using websites of government and medical manufacturers, books and magazines, and video sites. Although we are not aware of previous research on this topic, the findings are plausible, given that gathering information from these sources is an active process and is, therefore, more prevalent in health-information–oriented, health-concerned individuals [38], including people with higher food literacy. Interestingly, we found an inverse association between nutrition information seeking using television and food literacy. The exact reason is unclear, but this might be attributable to the definition and components of the food literacy score measured in this study, which do not necessarily include information seeking skills but rather represent a collection of interrelated knowledge, skills, and behaviors needed for healthy eating [21]. For people with lower food literacy scores and who are unlikely to be interested in nutrition-related topics, television may be the easiest way to obtain dietary and nutrition information.

In this study, the top dietary and nutrition information source was television, followed by web searches, websites of government and medical manufacturers, newspapers, books and magazines, and video sites. This is consistent with a finding from a national survey conducted in 2019 that showed that a larger portion of the general Japanese population considers traditional media, such as television (52%), books and magazines (23%), and newspapers (23%), as information sources that influence their dietary habits, compared with the internet (17% for websites and 8% for social media; multiple choices allowed) [34]. This was particularly apparent in older adults [34]. This tendency to rely on traditional mass media (particularly television and newspapers) for seeking health information has been repeatedly observed in Asian populations, such as Japanese [30-34] and Hong Kong Chinese populations [22,54]. The reason is unknown, but it may be because Asian cultures often prioritize values of obedience, duty, and in-group harmony [66,67].

A unique characteristic of this study is its inclusion of health professionals allied to nutrition, such as dietitians and registered dietitians. This design allowed us to describe interesting pictures of and build knowledge on how the general public and health professionals related to nutrition use and benefit from dietary and nutrition information seeking differently. As expected, we found that compared with the general public, dietitians and registered dietitians were more likely to refer to websites of government and medical manufacturers and books and magazines and less likely to use television or video sites for nutrition information. Taken together with our findings on food literacy and diet quality, this study does suggest that dietary and nutrition information obtained from websites of government and medical manufacturers and books and magazines would be relatively high in quality and usefulness, whereas information obtained from television (and video sites) would be relatively low in quality and useless. This interpretation is generally consistent with previous studies on the quality of nutrition-related information [13,68-74]. Nevertheless, no empirical evidence on the quality of web-based or offline dietary and nutrition information written in Japanese is available, and such investigations are warranted [17].

Studies have shown that being female and having a higher educational background are the strongest correlates of seeking health information from web-based sources [12,22,29,53,54]. Similar associations have also been suggested with regard to offline health information seeking [22,29,54], except for null or inverse associations between education and television exposure [38,75,76]. Studies also suggest that younger age is associated with web-based health information seeking [12,22,53], whereas older age is associated with offline health information seeking [22,28]. Our findings are generally consistent with these previous findings, except for findings regarding the role of sex. We found that being female was associated with the use of television and books and magazines, whereas being male was associated with the use of websites of...
In summary, although health-related information seeking varies across cultures and countries [23,80], all previous studies on nutrition information-seeking behaviors have been conducted in Western countries [11,25-28,35-38]. Therefore, our study of Japanese adults provides valuable insights into information-seeking behavior in relation to healthy eating and its potential consequences. This information may be used in future public health strategies aimed at improving health.

**Limitations**

This study has several limitations. First, it was conducted using a cross-sectional design. Because the temporality of the associations is unknown, it is not possible to address causality or its direction. For example, we cannot determine whether obtaining information about diet and nutrition from books and magazines contributes to improved food literacy and diet quality, or vice versa. More prospective research is needed on this topic. Second, our sample was confined to individuals registered with an internet research agency and, therefore, limited to people with access to the internet, and only a small percentage of these individuals expressed interest in participating in the study. Thus, compared with a nationally representative population, there is likely to be a higher use of web-based media for gathering information on diet and nutrition than what is observed online. In addition, by design, approximately half (2977/5998, 49.63%) of our participants were health professionals allied to nutrition. Consequently, the education level and household income of our participants were higher than those of a nationally representative sample (education: 54,760,800/100,247,800, 54.63% for junior high school or high school, 20,808,900/100,247,800, 20.76% for junior college or technical school, and 24,678,100/100,247,800, 24.62% for university or higher [81]; household income: 2943/5641, 45% for <4 million Japanese yen, 1760/6541, 26.9% for ≥4 to <7 million Japanese yen, and 1831/6541, 28% for ≥7 million Japanese yen [82]). Moreover, our participants appeared different from a nationally representative sample, at least in terms of mean (SD) height, weight, BMI, and HEI-2020 score (male individuals: 167.7, SD 6.9 cm; 67.4, SD 12.0 kg; 23.9, SD 3.6 kg/m²; and 51.3, SD 9.0, respectively; female individuals: 154.3, SD 6.7 cm; 53.6, SD 9.2 kg; 22.5, SD 3.7 kg/m²; and 52.9, SD 9.2, respectively) [34,59]. Taken together, although our sample size was large, the participants were not a nationally representative sample of the general Japanese population. For example, our study may have missed some socially susceptible groups (eg, people without homes or those receiving income support). By including these people, we may have found different correlates for seeking information about diet and nutrition from various web-based and offline media sources. Further research in a more representative sample is warranted.

Third, as with any survey based on a self-administered questionnaire, all the variables used in this study were based on self-report and may thus be subject to biased reporting despite our use of validated tools to assess health literacy [48], food literacy [49], and diet quality [50-52]. Regarding information-seeking behavior, several methodological studies have shown that self-reports of web-based activities [83], mobile phone use [84], and news exposure on television [85] often diverge from real behaviors. Thus, the self-reported responses regarding dietary and nutrition information seeking obtained in this study might be poor reflections of the truth. If so, the associations between media use for seeking dietary and nutrition information and the sociodemographic and lifestyle variables we observed may be generally diluted compared with the true associations. Therefore, the present findings should be interpreted with caution in this context, and efforts should be made to develop a validated scale to assess nutrition information-seeking behavior for use in future research. Fourth, as in many other studies [11,25-28,35-38,86], exposure to dietary and nutrition information sources was rather crudely measured in this study. The dichotomous nature of our measure did not permit a more refined measurement of variables, including an individual’s frequency of use (eg, yearly, weekly, and daily); the content, types, and quality of information participants were exposed to; and for whom the participants were seeking information. An in-depth analysis that incorporates these variables should be conducted to uncover the complexity of dietary and nutrition information-seeking behavior.

Fifth, although this study was conducted during a certain time window (February and March 2023), previous studies have suggested seasonal variations in dieting and weight loss attempts [87] and the frequency of internet searches for fitness and weight loss information [88]. Therefore, certain sociodemographic and lifestyle characteristics associated with media use for seeking dietary and nutrition information may manifest differently over time [86]. As such, it should be noted that this study provides only a snapshot in time. Future research should be designed to capture this dynamic nature of nutrition information-seeking behavior and its correlates. Finally, although we adjusted for various variables, we cannot rule out the possibility of residual confounding. Despite these limitations, the major strength of this study is its inclusion of a wide variety of dietary and nutrition information sources. Unlike a substantial majority of recent studies [11,23,36,37,86], our data allowed for differentiation among several web-based information sources (eg, web searches, websites of government and medical manufacturers, and video sites). This study also investigated more conventional, offline information sources (eg, television, newspapers, and books and magazines), which are still widely valued and used, in Asian countries at least [22,30-34,54]. People who use these information sources do not necessarily have the same approach and thus might not have the same profiles [89], which was clearly observed in this study.
Practical Implications

Our findings have implications for public health. First, the study demonstrates that a wide variety of sources are regularly used for seeking dietary and nutrition information in Japan. However, it was clear that each source was preferred by users with distinct characteristics. On the basis of these differences in potential users, we speculated that each media source should have suitable topics and optimal information dissemination strategies. For example, it is ideal to use television, an entertainment-oriented medium, for constantly disseminating easy-to-understand messages for chronic disease prevention (eg, “make half your plate fruits & vegetables”) [2] to instigate health-related conversations with family and friends for people with lower levels of formal education [38,76]. This may be particularly relevant, considering the multitasking agenda of women, who are frequent users of television and who play key roles as health managers and family caregivers [79,90,91]. In addition, newspapers (as well as books and magazines) should perhaps include only a small number of high-quality articles on the importance of overall dietary patterns, a topic for which the most powerful evidence is accumulated in nutritional epidemiology [1,2,4]. These articles should be written by reputable journalists with adequate training on issues related to scientific methodology, food, and health [92]. If successful, newspapers may play a vital role in chronic disease prevention and management [12] among middle-aged and older individuals with a higher educational background [38,76]. Further, for every web-based media source, nutrition professionals are ultimately responsible for and able to contribute to the creation and promotion of evidence-based content on topics that are popular [16,17,93] among the public [13,94]. Simultaneously, consideration should be given to the practicality and utility of the information [6] as well as to the affective needs (eg, enjoyment) and cognitive needs (eg, interest) of information [15], all of which are essential to the effective dissemination of dietary and nutrition information.

Ultimately, however, it is the consumers who decide from what and where they learn and what they eat. Research shows that although consumers say that they primarily look for the source, the presence of a professional design, and a variety of other criteria when assessing the credibility of a website, internet users do not in practice check the “about us” sections of websites, try to find out who the authors or owners of sites are, or read disclaimers or disclosure statements [95]. As an initial step, the general public should be sufficiently informed that there is a huge amount of misinformation (and potentially disinformation, ie, false information that is intentional spread) about nutrition, particularly on the internet and social media [7]. More fundamentally, critical thinking, or the ability to objectively analyze facts to make decisions successfully [96], appears to be an essential skill in the present age of highly developed information [6,7]. Excellence in this type of thinking does not come naturally but must be systematically cultivated [7]. Therefore, systematic learning of critical thinking skills should be introduced in the compulsory education system, with a focus on understanding the scientific process while eliminating the sole focus on results, thereby empowering consumers to evaluate evidence and understand the evolving nature of science [6].

Conclusions

This cross-sectional study identified various web-based and offline media sources regularly used by Japanese adults seeking nutrition information. Except for health literacy, each media source had unique correlates, suggesting large differences in potential users, relevant topics, and optimal information dissemination strategies among media sources. Perhaps the most concerning finding is the lack of positive associations between the use of 2 major sources (television and web searches) and food literacy and diet quality. By contrast, a promising finding is the positive associations of the use of websites of government and medical manufacturers, newspapers, books and magazines, and video sites with food literacy or diet quality. The present findings are a valuable scientific contribution to the development of effective promotional tactics and strategies for healthy eating.

Acknowledgments

This work was supported by a grant (22FA1022) from the Ministry of Health, Labour and Welfare, Japan. The Ministry of Health, Labour and Welfare had no role in the design of the study, analysis of the data, or writing of this paper.

Data Availability

The data sets used or analyzed in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

KM contributed to the concept and design of the survey and data collection and management, formulated the research question, analyzed and interpreted the data, prepared the first draft of the manuscript, and had primary responsibility for the final content. NS contributed to the concept and design of the survey, interpreted the data, and provided critical input for the final draft of the manuscript. TO contributed to the concept and design of the survey and provided critical input for the final draft of the manuscript. TAM and MBEL provided critical input for the final draft of the manuscript. All authors have read and approved the final manuscript.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Detailed description of the assessment of food literacy.
[DOCX File, 47 KB - publichealth_v10i1e54805_app1.docx]

Multimedia Appendix 2
Detailed description of the assessment of diet quality.
[DOCX File, 231 KB - publichealth_v10i1e54805_app2.docx]

Multimedia Appendix 3
Associations among the top 6 media sources used by Japanese adults for seeking diet and nutrition information (N=5998).
[DOCX File, 54 KB - publichealth_v10i1e54805_app3.docx]

Multimedia Appendix 4
Prevalence of diet and nutrition information seeking through the top 6 media sources among Japanese adults, according to participant characteristics (N=5998).
[DOCX File, 65 KB - publichealth_v10i1e54805_app4.docx]

Multimedia Appendix 5
Associations of participant characteristics with diet and nutrition information seeking through the top 6 media sources among Japanese adults belonging to the general public (3021/5998, 50.37%).
[DOCX File, 62 KB - publichealth_v10i1e54805_app5.docx]

Multimedia Appendix 6
Associations of participant characteristics with diet and nutrition information seeking through the top 6 media sources among Japanese adults with nutrition and health-related occupations (2977/5998, 49.63%).
[DOCX File, 63 KB - publichealth_v10i1e54805_app6.docx]

References


28. Ruani MA, Reiss MJ, Kalea AZ. Diet-nutrition information seeking, source trustworthiness, and eating behavior changes:
27. Sexton-Dhamu MJ, Livingstone KM, Pendergast FJ, Worsley A, McNaughton SA. Individual, social-environmental and
26. Baranowski T, Cullen KW, Nicklas T, Thompson D, Baranowski J. Are current health behavioral change models helpful


55. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

56. Home page. Rakuten Insight. URL: https://insight.rakuten.co.jp/ [accessed 2023-12-20]


Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
HEI: Healthy Eating Index

©Kentaro Murakami, Nana Shinozaki, Tsuyoshi Okuhara, Tracy A McCaffrey, M Barbara E Livingstone. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 14.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Original Paper

Digital Health Literacy of the Population in Germany and Its Association With Physical Health, Mental Health, Life Satisfaction, and Health Behaviors: Nationally Representative Survey Study

Lars König1,2, Dr rer nat; Adelheid Kuhlmey1, Prof Dr phil; Ralf Suhr1,2, PD Dr med

1Institut für Medizinische Soziologie und Rehabilitationswissenschaft, Charité – Universitätsmedizin Berlin, Berlin, Germany
2Stiftung Gesundheitswissen, Berlin, Germany

Corresponding Author:
Lars König, Dr rer nat
Institut für Medizinische Soziologie und Rehabilitationswissenschaft
Charité – Universitätsmedizin Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30419549262
Email: lars.koenig@charite.de

Abstract

Background: Digital health literacy, also known as eHealth literacy, describes the ability to seek, find, understand, and apply health information from the internet to address health problems. The World Health Organization calls for actions to improve digital health literacy. To develop target group–specific digital health literacy interventions, it is necessary to know the digital health literacy of the general population and relevant subgroups.

Objective: This study aims to representatively assess the digital health literacy of the population in Germany and relevant subgroups. The results are meant to facilitate the development of target group–specific digital health literacy interventions. Additionally, this study further explores the associations between digital health literacy and physical health, mental health, life satisfaction, and diverse health behaviors.

Methods: Study participants were drawn from a representative panel of the German-speaking population with internet access. To further increase the representativeness of the sample, survey weights were calculated using an iterative proportional fitting procedure. Participants answered a series of questionnaires regarding their digital health literacy, physical health, mental health, life satisfaction, and diverse health behaviors. Two-sided independent sample t tests were conducted to determine the significant differences between societal subgroups. Pearson correlation coefficients were calculated to explore the correlates of digital health literacy.

Results: Digital health literacy is unevenly distributed within German society. The results of this study suggest that people with a low level of formal education and people with a low social status would benefit from digital health literacy interventions that address their competencies in the domains of information seeking and information appraisal. Furthermore, the results suggest that older people would likely benefit from digital health literacy interventions that address their competencies in the domains of information seeking and also information appraisal. Regarding sex, this study suggests that men might benefit from digital health literacy interventions that specifically address their competencies in the domain of information seeking. Furthermore, digital health literacy is weakly positively correlated with physical health, mental health, life satisfaction, exercise routines, fruit consumption, and vegetable consumption.

Conclusions: Overall, the results of this study demonstrate that digital health literacy is associated with diverse health outcomes and behaviors. Furthermore, the results provide a starting point for the development of target group–specific digital health literacy interventions.

(JMIR Public Health Surveill 2024;10:e48685) doi:10.2196/48685

https://publichealth.jmir.org/2024/1/e48685

JMIR Public Health Surveill 2024 | vol. 10 | e48685 | p.1179
(page number not for citation purposes)
KEYWORDS
digital health; digital health literacy; eHealth; eHealth literacy; health behaviors; health literacy; life satisfaction; mental health; physical health; representative survey

Introduction
Digital health literacy, also known as eHealth literacy, can be defined as “the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem” [1]. Nowadays, these skills seem particularly important for two reasons. First, people from around the world regularly use the internet to acquire health information [2,3]. Second, when searching the internet for health information, people are often confronted with misinformation [4,5]. Furthermore, recent studies have found that evaluating the reliability of health information on the internet is not just difficult for the general population but also for better educated and younger subgroups like university students [6-8].

Besides its central role in the context of evaluating health information on the internet, digital health literacy seems important because it is associated with various health intentions and behaviors [9]. Studies in diverse contexts have shown, for example, that people with high digital health literacy demonstrate better physical exercise routines [10], eat more nutritionally balanced diets [11], and protect themselves better against viruses [12]. Furthermore, they have better cancer screening practices [13], are more confident about finding cancer information [14], and are less likely to fall prey to conspiracy beliefs [12]. Given these potential positive health effects, it is no surprise that the World Health Organization calls for actions to improve digital health literacy [15].

To develop target group-specific digital health literacy interventions, it is necessary to know the digital health literacy of the general population and relevant subgroups. Various instruments have been developed to measure digital health literacy [16]. One of the most widely used instruments is the eHealth Literacy Scale [17], which has already been translated into diverse languages, including Dutch [18], Chinese [19], and Korean [20]. A German version of the eHealth Literacy Scale was published in 2014 [21]. This instrument, however, was recently criticized because of content-related and methodological shortcomings, and a revised German eHealth Literacy Scale was published in 2022 [9].

This study aims to representatively assess the digital health literacy of the population in Germany and relevant subgroups, using the revised German eHealth Literacy Scale [9]. The results are meant to facilitate the development of target group-specific digital health literacy interventions. Additionally, this study further explores the associations between digital health literacy and physical health, mental health, life satisfaction, and diverse health behaviors (exercise routines, fruit consumption, vegetable consumption, soft drink consumption, alcohol consumption, cigarette consumption, and drug consumption).

Methods

Ethical Considerations
Before data collection, a detailed study protocol that included information on the procedures, measures, and statistical analyses was submitted to the ethics committee of the Berlin Medical Association. The ethics committee consisted of 2 medical doctors, a lawyer, a statistician, and a layperson. The ethics committee had no ethical or professional objections to the study protocol (reference Eth-39/22). Before the study started, participants provided their informed consent to take part in the study. Participants had the opportunity to opt out of the study at any time during the study. The independent, nonprofit foundation Stiftung Gesundheitswissen did not compensate the participants for their participation, and participants were informed that the foundation would only receive anonymized data.

Survey Methodology
The market research Institute Forsa Gesellschaft für Sozialforschung und Statistische Analysen mbH (Forsa) was responsible for data acquisition [22]. The survey was conducted using the representative web-based panel forsa-omninet, which is a representative panel for the German-speaking population with internet access and currently has around 100,000 participants. A continuous recruiting process adds new participants to the panel every month. In addition, the composition of the panel is continuously monitored based on key characteristics (eg, region, age, and sex), and recruitment is adjusted accordingly. Data acquisition took place from September 22 to October 12, 2022. A random sample was drawn from the representative web-based panel. All panelists selected for the survey were invited via email. The invitation email provided information on the topic and purpose of the survey. If necessary, the selected panelists were reminded about their participation in 2 further emails. In total, 3927 panelists were invited to take part in the survey. All responses were checked for plausibility, and a comprehensive speeder analysis was conducted based on the response times. A total of 2000 panelists completed the survey. This corresponds to a response rate of about 50.9% (2000/3927), which is quite high for a relatively long survey that is completed on a voluntary basis. On average, panelists took 20 minutes to finish the survey. The sample characteristics of the participants who were invited to participate in the study (invited sample) and the participants who were included in the data analyses before (unweighted sample) and after the weighting procedure (weighted sample) can be found in the Results section. Please note that the unweighted and weighted samples are based on the 1996 study participants who answered all items from the revised German eHealth Literacy Scale. Furthermore, please note that due to sample weighting and rounding, sample sizes may vary, and percentages may exceed or fall below 100%. There are differences between these groups. For example, 50.6% (1988/3927) of the invited sample...
All things considered, how satisfied are you with your life to assess life satisfaction [24]. Participants answered the question on scales ranging from 0 (not at all satisfied) to 10 (completely satisfied).

Health Behavior

Health behavior was assessed by asking participants questions about their health behavior within a typical week. Participants were asked about their exercise routines (On average, how many days a week do you exercise?), fruit consumption (On average, how many days a week do you eat fruit?), vegetable consumption (On average, how many days a week do you eat vegetables?), soft drink consumption (On average, how many days a week do you drink sugary soft drinks?), alcohol consumption (On average, how many days a week do you drink alcohol?), cigarette consumption (On average, how many days a week do you smoke cigarettes?), and drug consumption (On average, how many days a week do you use illegal drugs?). Participants answered the questions on scales ranging from 0 (0 days) to 7 (7 days). This study was part of a larger study, and therefore the raw data set contains further variables that have not been described because they exceed the scope of this study.

Statistical Analyses

All statistical analyses were conducted with the statistical software SPSS (version 29.0.0.0; IBM Corp). Cronbach α was calculated for the 2 subscales of the revised German eHealth Literacy Scale to ensure the quality of the measures. Two-sided independent sample t tests were conducted to determine the significant differences between the subgroups. For all 2-tailed t tests, unequal variance was assumed. Throughout the analyses, group differences were considered significant if P<.05, which should be taken into consideration when interpreting the results. Interested readers can find detailed information on when it is appropriate to adjust significance thresholds (eg, disjunction testing vs conjunction testing vs individual testing) elsewhere [25]. Pearson correlations were calculated to explore the associations between digital health literacy and physical health, mental health, life satisfaction, and health behaviors. Most of the instruments used in this study relied on Likert-like scales and therefore produced ordinal data. There has been a long debate about whether data from Likert-like scales should be analyzed using parametric statistics or less sensitive and less powerful nonparametric statistics (eg, 2-sample t test vs Wilcoxon rank sum test, Pearson coefficient of correlation vs Spearman rank correlation) [26-28]. Previous research has demonstrated that parametric statistics can be quite robust against violations of its assumptions, especially with large sample sizes [26-28]. One article concluded that “Parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of ‘coming to the wrong conclusion.’” These findings are consistent with empirical literature “dating back nearly 80 years” [26]. Hence, parametric statistics were chosen for this analysis to avoid an unnecessary loss of information. The final data analyses were based on the 1996 study participants who answered all items from the revised German eHealth Literacy Scale.

Digital Health Literacy

The revised German eHealth Literacy Scale was used to assess digital health literacy [9]. The instrument consists of 8 items that comprise 2 subscales assessing competencies in the domain of information seeking (4 items; eg, I know how to find helpful health resources on the Internet) and information appraisal (4 items; eg, I have the skills I need to evaluate the health resources I find on the Internet). Participants rated all items on scales ranging from 1 (strongly disagree) to 5 (strongly agree). A total score was generated for each subscale by calculating the mean.

Physical Health and Mental Health

Physical health was assessed by asking participants “Overall, how do you currently rate your physical health?” Mental health was assessed by asking participants “Overall, how do you currently rate your mental health?” Participants answered the questions on scales ranging from 0 (very bad) to 10 (very good).

Life Satisfaction

The 1-item General Life Satisfaction Short Scale (L-1) was used to assess life satisfaction [24]. Participants answered the question “All things considered, how satisfied are you with your life...
Subgroup Analysis

In Germany, the Health Literacy Survey Germany project has assessed the general health literacy of the population in Germany and relevant subgroups [8]. To simplify and facilitate comparisons between different studies, similar subgroups were chosen for this study. The subgroups were divided along the lines of level of education (low, middle, and high), social status (low, middle, and high), age (16-29 years, 30-45 years, 46-64 years, and 65 years and older), chronic disease (no and yes), migration background (no and yes), and sex (male and female). The German version of the MacArthur Scale was used to assess social status [29]. Scores were classified as low (1-4), middle (5-7), and high (8-10). Educational degrees (eg, low=no degree, middle=high school degree, and high=university degree) were used to assess participants’ level of education. Educational degrees were classified as low (ohne Haupt-/Volksschulabschluss; Haupt-/Volksschulabschluss; Mittlere Reife, Realschulabschluss, Fachschulreife; Abschluss der Polytechnischen Oberschule), middle (Fachhochschulreife, Abschluss einer Fachoberschule; Abitur, allgemeine oder fachgebundene Hochschulreife), and high (Fach-/Hochschulstudium).

Results

Sample Characteristics

To increase the representativeness of the sample, survey weights were calculated, and the sample was weighted accordingly. Table 1 provides invited, unweighted, and weighted sample characteristics by sex, age, state, level of education, migration background, chronic disease, and social status.
Table 1. Nationally representative survey of the population in Germany: invited, unweighted, and weighted sample characteristics by sex, age, state, level of education, migration background, chronic disease, and social status.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Invited sample, n (%)</th>
<th>Unweighted sample, n (%)</th>
<th>Weighted sample, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1988 (50.6)</td>
<td>955 (47.8)</td>
<td>978 (49)</td>
</tr>
<tr>
<td>Female</td>
<td>1939 (49.4)</td>
<td>1041 (52.2)</td>
<td>1019 (51)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-29</td>
<td>825 (21)</td>
<td>459 (23)</td>
<td>353 (17.7)</td>
</tr>
<tr>
<td>30-45</td>
<td>904 (23)</td>
<td>433 (21.7)</td>
<td>475 (23.8)</td>
</tr>
<tr>
<td>46-64</td>
<td>1256 (32)</td>
<td>598 (30)</td>
<td>654 (32.7)</td>
</tr>
<tr>
<td>65 years and older</td>
<td>942 (24)</td>
<td>506 (25.4)</td>
<td>514 (25.8)</td>
</tr>
<tr>
<td>State</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schleswig-Holstein</td>
<td>138 (3.5)</td>
<td>91 (4.6)</td>
<td>70 (3.5)</td>
</tr>
<tr>
<td>Hamburg</td>
<td>90 (2.3)</td>
<td>43 (2.2)</td>
<td>43 (2.2)</td>
</tr>
<tr>
<td>Lower Saxony</td>
<td>381 (9.7)</td>
<td>208 (10.4)</td>
<td>191 (9.6)</td>
</tr>
<tr>
<td>Bremen</td>
<td>31 (0.8)</td>
<td>11 (0.6)</td>
<td>16 (0.8)</td>
</tr>
<tr>
<td>Northrhine-Westphalia</td>
<td>844 (21.5)</td>
<td>384 (19.2)</td>
<td>430 (21.5)</td>
</tr>
<tr>
<td>Hesse</td>
<td>298 (7.6)</td>
<td>166 (8.3)</td>
<td>151 (7.6)</td>
</tr>
<tr>
<td>Rhineland-Palatinate</td>
<td>196 (5)</td>
<td>88 (4.4)</td>
<td>99 (4.9)</td>
</tr>
<tr>
<td>Baden-Württemberg</td>
<td>534 (13.6)</td>
<td>300 (15)</td>
<td>265 (13.3)</td>
</tr>
<tr>
<td>Bavaria</td>
<td>643 (16.4)</td>
<td>370 (18.5)</td>
<td>316 (15.8)</td>
</tr>
<tr>
<td>Saarland</td>
<td>37 (0.9)</td>
<td>20 (1.0)</td>
<td>24 (1.2)</td>
</tr>
<tr>
<td>Berlin</td>
<td>176 (4.5)</td>
<td>89 (4.5)</td>
<td>88 (4.4)</td>
</tr>
<tr>
<td>Brandenburg</td>
<td>113 (2.9)</td>
<td>47 (2.4)</td>
<td>61 (3.1)</td>
</tr>
<tr>
<td>Mecklenburg-West Pomerania</td>
<td>70 (1.8)</td>
<td>30 (1.5)</td>
<td>39 (2)</td>
</tr>
<tr>
<td>Saxony</td>
<td>184 (4.7)</td>
<td>69 (3.5)</td>
<td>97 (4.9)</td>
</tr>
<tr>
<td>Saxony-Anhalt</td>
<td>98 (2.5)</td>
<td>43 (2.2)</td>
<td>53 (2.7)</td>
</tr>
<tr>
<td>Thuringia</td>
<td>94 (2.4)</td>
<td>37 (1.9)</td>
<td>52 (2.6)</td>
</tr>
<tr>
<td>Level of education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1091 (54.7)</td>
<td>1105 (55.3)</td>
</tr>
<tr>
<td>Middle</td>
<td>N/A</td>
<td>399 (20)</td>
<td>372 (18.7)</td>
</tr>
<tr>
<td>High</td>
<td>N/A</td>
<td>494 (24.7)</td>
<td>511 (25.6)</td>
</tr>
<tr>
<td>Missing values</td>
<td>N/A</td>
<td>12 (0.6)</td>
<td>8 (0.4)</td>
</tr>
<tr>
<td>Migration background</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>1867 (93.5)</td>
<td>1873 (93.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>N/A</td>
<td>129 (6.5)</td>
<td>124 (6.2)</td>
</tr>
<tr>
<td>Chronic disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>1257 (63)</td>
<td>1230 (61.6)</td>
</tr>
<tr>
<td>Yes</td>
<td>N/A</td>
<td>714 (35.8)</td>
<td>742 (37.2)</td>
</tr>
<tr>
<td>Missing values</td>
<td>N/A</td>
<td>25 (1.3)</td>
<td>25 (1.2)</td>
</tr>
<tr>
<td>Social status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>N/A</td>
<td>312 (15.6)</td>
<td>315 (15.8)</td>
</tr>
<tr>
<td>Middle</td>
<td>N/A</td>
<td>1393 (69.8)</td>
<td>1385 (69.4)</td>
</tr>
<tr>
<td>High</td>
<td>N/A</td>
<td>291 (14.6)</td>
<td>296 (14.8)</td>
</tr>
</tbody>
</table>
Quality of Measures

Using the unweighted data, Cronbach α was calculated for the 2 subscales of the revised German eHealth Literacy Scale to ensure the quality of the measures. Widely used conventions define Cronbach α values of .7 and higher as acceptable [30,31]. Both scales surpassed this widely used Cronbach α threshold (information seeking: .897 and information appraisal: .839).

Level of Education

On average, people with a high level of education had higher perceived competency levels in the domains of information seeking (t_{3,415}=1068.213; P<.001; d=0.177) and information appraisal (t_{6,406}=1118.823; P<.001; d=0.327) than people with a low level of education. Perceived competency levels did not differ significantly between people with a high level of education and people with a middle level of education (information seeking: t_{1,062}=809.769; P=.29; d=0.072 and information appraisal: t_{0.414}=818.391; P=.66; d=0.029). Furthermore, perceived competency levels in the domain of information seeking did not differ significantly between people with a middle level of education and people with a low level of education (t_{1,920}=701.099; P=.05; d=0.110). However, people with a middle level of education had higher perceived competency levels in the domain of information appraisal than people with a low level of education (t_{5,456}=750.726; P<.001; d=0.301).

Social Status

On average, people with a high social status had higher perceived competency levels in the domains of information seeking (t_{4,947}=606.910; P<.001; d=0.399) and information appraisal (t_{6,583}=598.474; P<.001; d=0.530) than people with a low social status. Furthermore, people with a high social status had higher perceived competency levels than people with a middle social status (information seeking: t_{2,222}=437.791; P=.03; d=0.140 and information appraisal: t_{2,751}=460.093; P=.006; d=0.166). Additionally, people with a middle social status had higher perceived competency levels than people with a low social status (information seeking: t_{4,089}=438.922; P<.001; d=0.271 and information appraisal: t_{5,526}=437.177; P<.001; d=0.368).

Age

On average, perceived competency levels in the domain of information seeking did not differ significantly between people aged 16-29 years and people aged 65 years and older (t_{1,880}=826.226; P=.06; d=0.127). However, people aged 16-29 years had higher perceived competency levels in the domain of information appraisal than people aged 65 years and older (t_{4,787}=831.799; P<.001; d=0.320). People aged 16-29 years had higher perceived competency levels than people aged 46-64 years (information seeking: t_{2,253}=761.604; P=.02; d=0.146 and information appraisal: t_{4,002}=783.957; P<.001; d=0.257). Perceived competency levels did not differ significantly between people aged 16-29 years and people aged 30-45 years (information seeking: t_{0.256}=754.114; P=.80; d=–0.018 and information appraisal: t_{0.184}=762.075; P=.85; d=–0.013). People aged 30-45 years had higher perceived competency levels than people aged 46-64 years (information seeking: t_{2,766}=1062.057; P=.006; d=0.165 and information appraisal: t_{2,542}=1069.427; P<.001; d=0.270) and people aged 65 years and older (information seeking: t_{2,997}=979.125; P=.02; d=0.145 and information appraisal: t_{5,301}=979.619; P<.001; d=0.335). However, perceived competency levels did not differ significantly between people aged 46-64 years and people aged 65 years and older (information seeking: t_{0.818}=1048.014; P=.86; d=–0.011 and information appraisal: t_{1,236}=1058.857; P=.22; d=0.074).

Chronic Disease

On average, perceived competency levels in the domains of information seeking (U=1548.068; P=.71; d=–0.017) and information appraisal (U=1512.596; P=.12; d=0.072) did not differ significantly between people without a chronic disease and people with a chronic disease.

Migration Background

On average, perceived competency levels in the domains of information seeking (t_{0.369}=139.851; P=.71; d=0.034) and information appraisal (t_{0.604}=142.612; P=.55; d=–0.052) did not differ significantly between people without a migration background and people with a migration background.

Sex

On average, women had higher perceived competency levels in the domain of information seeking than men (t_{2,333}=1992.953; P=.02; d=0.104). However, perceived competency levels in the domain of information appraisal did not differ significantly between women and men (t_{0.016}=1993.874; P=.99; d=0.001). Table 2 shows the digital health literacy of the population in Germany by level of education, social status, age, chronic disease, migration background, and sex.
Table 2. Nationally representative survey of the population in Germany: digital health literacy by the level of education, social status, age, chronic disease, migration background, and sex. Statistically significant differences are indicated by superscript letters.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Information seeking, mean (SD)</th>
<th>Information appraisal, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population in Germany</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>3.7940 (0.92522)</td>
<td>3.7680 (0.85482)</td>
</tr>
<tr>
<td><strong>Level of education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (a)</td>
<td>3.7323 (0.95846)</td>
<td>3.6483 (0.89606)</td>
</tr>
<tr>
<td>Middle (b)</td>
<td>3.8350 (0.86613)</td>
<td>3.9076 (0.75513)</td>
</tr>
<tr>
<td>High (c)</td>
<td>3.8983 (0.88436)</td>
<td>3.9303 (0.78663)</td>
</tr>
<tr>
<td><strong>Social status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (d)</td>
<td>3.5640 (0.99864)</td>
<td>3.4817 (0.93104)</td>
</tr>
<tr>
<td>Middle (e)</td>
<td>3.8148 (0.90731)</td>
<td>3.7974 (0.84026)</td>
</tr>
<tr>
<td>High (f)</td>
<td>3.9414 (0.88264)</td>
<td>3.9347 (0.76645)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-29 (g)</td>
<td>3.8649 (0.86732)</td>
<td>3.9073 (0.77874)</td>
</tr>
<tr>
<td>30-45 (h)</td>
<td>3.8804 (0.85653)</td>
<td>3.9174 (0.78404)</td>
</tr>
<tr>
<td>46-64 (i)</td>
<td>3.7329 (0.92275)</td>
<td>3.6939 (0.85758)</td>
</tr>
<tr>
<td>65 years and older (j)</td>
<td>3.7433 (1.01659)</td>
<td>3.6286 (0.92788)</td>
</tr>
<tr>
<td><strong>Chronic disease</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (k)</td>
<td>3.7915 (0.92113)</td>
<td>3.7943 (0.84147)</td>
</tr>
<tr>
<td>Yes (l)</td>
<td>3.8073 (0.93183)</td>
<td>3.7326 (0.87641)</td>
</tr>
<tr>
<td><strong>Migration background</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (m)</td>
<td>3.7960 (0.92588)</td>
<td>3.7652 (0.85894)</td>
</tr>
<tr>
<td>Yes (n)</td>
<td>3.7646 (0.91835)</td>
<td>3.8099 (0.79211)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (o)</td>
<td>3.7448 (0.91756)</td>
<td>3.7677 (0.84414)</td>
</tr>
<tr>
<td>Female (p)</td>
<td>3.8413 (0.93051)</td>
<td>3.7683 (0.86538)</td>
</tr>
</tbody>
</table>

**Correlates of Information Seeking**

Participants’ perceived competency levels in the domain of information seeking were weakly positively and significantly correlated with their physical health, mental health, life satisfaction, exercise routines, fruit consumption, and vegetable consumption. Furthermore, information seeking was weakly negatively and significantly correlated with alcohol consumption. There were no significant correlations between information seeking and soft drink consumption, cigarette consumption, and drug consumption. Table 3 provides further information about the correlation coefficients and their significance levels.
Table 3. Nationally representative survey of the population in Germany: correlations (Pearson r and 2-tailed P value) between digital health literacy (information seeking and information appraisal) and physical health, mental health, life satisfaction, and health behaviors.

<table>
<thead>
<tr>
<th></th>
<th>Information seeking</th>
<th>Information appraisal</th>
<th>Physical health</th>
<th>Mental health</th>
<th>Life satisfaction</th>
<th>Exercise routine</th>
<th>Fruit consumption</th>
<th>Vegetable consumption</th>
<th>Soft drink consumption</th>
<th>Alcohol consumption</th>
<th>Cigarette consumption</th>
<th>Drug consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Information seeking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td></td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Information appraisal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.794b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Physical health</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.99b</td>
<td>0.150b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mental health</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.109b</td>
<td>0.156b</td>
<td>0.510b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Life satisfaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.126b</td>
<td>0.157b</td>
<td>0.528b</td>
<td>0.698b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exercise routines</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.070b</td>
<td>0.096b</td>
<td>0.216b</td>
<td>0.123b</td>
<td>0.132b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>.002</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fruit consumption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.061b</td>
<td>0.056c</td>
<td>0.077b</td>
<td>0.124b</td>
<td>0.127b</td>
<td>0.224b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>.007</td>
<td>.026</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vegetable consumption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>0.099b</td>
<td>0.145b</td>
<td>0.084b</td>
<td>0.085b</td>
<td>0.111b</td>
<td>0.218b</td>
<td>0.485b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Soft drink consumption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>−0.036</td>
<td>−0.026</td>
<td>−0.056*</td>
<td>−0.088b</td>
<td>−0.100b</td>
<td>−0.120b</td>
<td>−0.190b</td>
<td>−0.161b</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>.104</td>
<td>.251</td>
<td>.012</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alcohol consumption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>−0.049*</td>
<td>−0.019</td>
<td>0.073b</td>
<td>0.126b</td>
<td>0.120b</td>
<td>−0.018</td>
<td>−0.023</td>
<td>0.024</td>
<td>−0.085b</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>.028</td>
<td>.389</td>
<td>.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.414</td>
<td>.305</td>
<td>.285</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cigarette consumption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>−0.011</td>
<td>−0.037</td>
<td>−0.112b</td>
<td>−0.101b</td>
<td>−0.132b</td>
<td>−0.136b</td>
<td>−0.129b</td>
<td>−0.061b</td>
<td>0.100b</td>
<td>.000</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>.626</td>
<td>.102</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.007</td>
<td>&lt;.001</td>
<td>1.000</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td><strong>Drug consumption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>−0.015</td>
<td>.000</td>
<td>0.009</td>
<td>−0.031</td>
<td>−0.042</td>
<td>0.054c</td>
<td>−0.010</td>
<td>0.017</td>
<td>0.090b</td>
<td>−0.004</td>
<td>0.152b</td>
<td>1</td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>.514</td>
<td>.984</td>
<td>.682</td>
<td>.166</td>
<td>.059</td>
<td>.016</td>
<td>.659</td>
<td>.457</td>
<td>&lt;.001</td>
<td>.860</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Correlates of Information Appraisal

Participants’ perceived competency levels in the domain of information appraisal were weakly positively and significantly correlated with their physical health, mental health, life satisfaction, exercise routines, fruit consumption, and vegetable consumption. There were no significant correlations between information appraisal and soft drink consumption, alcohol consumption, cigarette consumption, and drug consumption. Table 3 provides further information about the correlation coefficients and their significance levels.

Discussion

Principal Findings

To guide the development of target group–specific digital health literacy interventions, it is necessary to know the digital health literacy of different subgroups within a society. For Germany, such data have not been collected with appropriate and methodological sound measures. Therefore, this study representatively assessed the digital health literacy of the population in Germany and relevant subgroups. The results suggest that people with a low level of education and people with a low social status would benefit from digital health literacy interventions that address their competencies in the domains of information seeking and information appraisal. These findings align with previous research, which suggests that individuals with lower levels of education and lower social status tend to have lower general health literacy [8,32]. Furthermore, the results suggest that older people would likely benefit from digital health literacy interventions that address their competencies in the domains of information seeking and information appraisal. Once again, these findings are in line with results from general health literacy research, which highlights that older individuals may constitute a vulnerable group deserving special attention [8,32,33].

Previous research has found that people with chronic diseases and people with migration backgrounds have lower health literacy levels [8,34,35]. Such differences, however, were not found in this study. People with a chronic disease and people without a chronic disease did not significantly differ in regard to their digital health literacy. Neither did people with a migration background and people without a migration background significantly differ in regard to their digital health literacy. The last finding is in line with results from a study conducted in Israel, which suggests that the migration status of an individual may be linked to general health literacy but not necessarily to digital health literacy [36]. The authors of the study hypothesize that this finding might be attributed to the fact that certain language barriers are less prominent on the internet compared to real-world settings. Another contributing factor could be that the study sample might have contained individuals with migration backgrounds who possess particularly strong language skills. Regarding sex, this study suggests that men might benefit from digital health literacy interventions that specifically address their competencies in the domain of information seeking. These findings corroborate earlier research results which indicate that women have higher (digital) health literacy than men, even though the differences are often small [7,8,37].

Besides comparing the digital health literacy of different societal groups, this study explored the associations between digital health literacy and diverse health-related constructs and behaviors. The results suggest that digital health literacy is weakly positively correlated with physical health, mental health, life satisfaction, exercise routines, fruit consumption, and vegetable consumption. Moreover, higher perceived competency levels in the domain of information seeking were weakly negatively correlated with alcohol consumption. When interpreting these results, however, it must be stressed that the found associations were rather weak.

In line with previous research, information appraisal was more strongly associated with mental health, physical health, and life satisfaction than information seeking [9]. Interestingly, the results suggest that digital health literacy is more strongly associated with health perceptions (eg, mental health) than with health behaviors (eg, exercise routines). As a reviewer (EN) of this paper pointed out, this finding might be explained by methodological considerations. For example, 2 health perception variables (eg, “Overall, how do you currently rate your mental health?”) might show stronger associations with each other than a health perception variable and a health behavior variable (eg, “On average, how many days a week do you exercise?”). Future studies should develop adequate research designs to test this hypothesis.

Limitations and Future Directions

The results of this study demonstrate the importance of digital health literacy in society. There are, however, limitations that must be kept in mind when interpreting the results. Three of these limitations seem especially important and need to be discussed in more detail. The first limitation concerns the study design. This study used a cross-sectional study design, and therefore no causal inference can be drawn [23]. This study found, for example, that participants’ digital health literacy is positively and significantly correlated with their physical health. From this finding, however, one cannot conclude that digital health literacy causes physical health. To draw a causal inference like this, future studies could, for example, use an experimental study design and actively manipulate the digital health literacy of the study participants before assessing their physical health [38].

The second limitation concerns the data collection method of this study. Data collection took place via a web-based survey platform. Hence, participants needed basic technical skills and access to the internet to participate in the study. It is known, however, that older people use the internet less frequently than younger people [39]. Therefore, this study can only be representative of the German population with internet access.
People without any experience with the internet probably have lower digital health literacy scores. Hence, the data collection method should be taken into account when comparing the results of this study to other digital health literacy studies that might have used other data collection methods such as in-person or telephone interviews. There is another problem that might have affected the representativeness of the study results, which has already been raised in the survey methodology section and concerns the nonresponders and potential selection bias. As a reminder, 50.6% (1988/3927) of the invited sample were men, but only 47.8% (955/1996) of the unweighted sample were men. After the weighting procedure, 49% (978/1997) of the sample were counted as men. During the analyses, it was found that, on average, women had higher perceived competency levels in the domain of information seeking than men. One might argue that men with an above-average interest in the topic of digital health literacy were more likely to complete the study, and their response data were weighted. Hence, the found difference between women and men might be even more pronounced within the German population. Such an argument can also be made regarding other demographic factors. Since we have no information about the digital health literacy of the nonresponders, we currently cannot test this hypothesis. Nevertheless, it is important to bear in mind that there were demographic differences between responders and nonresponders, as can be seen in Table 1, and this might have affected the results, even after the weighting procedure.

The third limitation concerns the types of measures that were used in this study. Besides multi-item measures, this study also used single-item measures. Single-item measures are often criticized because they usually cannot assess complex constructs (eg, personality) and may be less reliable under specific circumstances [40-42]. In many situations, however, there are good arguments in favor of the use of single-item measures, and previous research has shown that these measures can be reliable and valid [40,43,44]. Nevertheless, it would be interesting to substitute the single-item measures with multi-item measures in future studies to investigate whether these changes would alter the general direction of the results. Furthermore, this study also relied on self-report measures. Self-report measures, however, are often criticized because they may produce inaccurate results [45-47]. When asked about their vegetable consumption within a typical week, for example, respondents might give an inaccurate answer because they usually do not track their eating habits very carefully. When asked about their drug consumption within a typical week, respondents might be inclined to lie because they do not want to make a bad impression or admit to themselves that they might have a drug problem. Therefore, to verify these results, it would be interesting to repeat this study with behavioral, observational, and performance measures instead of self-report measures.

**Conclusions**

Overall, the results of this study demonstrate that digital health literacy is associated with diverse health outcomes and behaviors. Furthermore, the results provide a starting point for the development of target group–specific digital health literacy interventions.

---

**Acknowledgments**

This research was funded by the independent, nonprofit foundation Stiftung Gesundheitswissen. Generative artificial intelligence (eg, ChatGPT and DeepL) has occasionally been used to generate ideas for English phrasing (eg, “What is the most used English translation for the German word ‘bevölkerungsrepräsentativ’?”). The authors want to thank the reviewers and editors for their constructive and valuable comments which improved the manuscript significantly. Forsa was responsible for the recruitment of study participants, data acquisition, and the calculation of survey weights.

**Data Availability**

The data sets generated and analyzed during this study are available from the independent, nonprofit foundation Stiftung Gesundheitswissen on reasonable request.

**Authors’ Contributions**

LK and RS contributed to the conceptualization of the study and the study design. LK and RS were responsible for the analysis and interpretation of data. LK was responsible for drafting the first version of the manuscript. LK, RS, and AK were responsible for fact-checking and revising the manuscript critically for important intellectual content. LK, RS, and AK reviewed the final manuscript and approved its publication.

**Conflicts of Interest**

LK and RS are employees of the independent, nonprofit foundation Stiftung Gesundheitswissen.

**References**

2. Baumann E, Czerwinski F, Rosset M, Seelig M, Suhr R. Wie informieren sich die Menschen in Deutschland zum Thema Gesundheit? Erkenntnisse aus der ersten Welle von HINTS Germany. [How do people in Germany seek health information?](https://publichealth.jmir.org/2024/1/e48685)
Insights from the first wave of HINTS Germany. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020;63(9):1151-1160 [FREE Full text] [doi: 10.1007/s00103-020-03192-x] [Medline: 32666180]


Abbreviations

Forsa: forsa Gesellschaft für Sozialforschung und statistische Analysen mbH
Mediating Effect of Tobacco Dependence on the Association Between Maternal Smoking During Pregnancy and Chronic Obstructive Pulmonary Disease: Case-Control Study

Jinxuan Li$^{1,2,3,4,5,6,7}$, MD; Jianying Xu$^{8}$, MD; Lan Yang$^{9}$, MD; Yongjian Xu$^{10}$, MD; Xiayang Zhang$^{11}$, MD; Chunxue Bai$^{12}$, MD; Jian Kang$^{13}$, MD; Pixin Ran$^{14}$, MD; Huahao Shen$^{15}$, MD; Fuqiang Wen$^{16}$, MD; Kewu Huang$^{17}$, MD; Wanzhen Yao$^{18}$, MD; Tieying Sun$^{19,20}$, MD; Guangliang Shan$^{21}$, MD; Ting Yang$^{2,3,4,5,22}$, MD; Yingxiang Lin$^{17}$, MD; Jianguo Zhu$^{20}$, MD; Ruiling Wang$^{8}$, MD; Zhihong Shi$^{9}$, MD; Jianping Zhao$^{10}$, MD; Xianwei Ye$^{11}$, MD; Yuanlin Song$^{12}$, MD; Qiuju Wang$^{13}$, MD; Gang Hou$^{2,3,4,5,22}$, MD; Yumin Zhou$^{14}$, MD; Wen Li$^{15}$, MD; Liren Ding$^{15}$, MD; Hao Wang$^{16}$, MD; Yahong Chen$^{18}$, MD; Yanfei Guo$^{19,20}$, MD; Fei Xiao$^{20}$, MD; Yong Lu$^{17}$, MD; Xiaoxia Peng$^{23,24}$, MD; Biao Zhang$^{21}$, MD; Zuomin Wang$^{25}$, MD; Hong Zhang$^{17}$, MD; Xiaoting Bu$^{17}$, MD; Xiaolei Zhang$^{2,3,4,5,22}$, MD; Li An$^{17}$, MD; Shu Zhang$^{17}$, MD; Zixi Cao$^{17}$, MD; Qingyuan Zhan$^{2,3,4,5,22}$, MD; Yuanting Yang$^{17}$, MD; Lirong Liang$^{26,27}$, MD; Bin Cao$^{2,3,4,5,22}$, MD; Huaping Dai$^{2,3,4,5,22}$, MD; Kian Fan Chung$^{28}$, MD; Zhongming Chen$^{29}$, PhD; Jiang He$^{30}$, MD; Sinan Wu$^{2,3,4,5,31}$, MD; Dan Xiao$^{2,3,4,5,6,7^*}$, MD; Chen Wang$^{2,3,4,5,7,22^*}$, MD; China Pulmonary Health Study Group$^{22}$

1. China-Japan Friendship School of Clinical Medicine, Capital Medical University, Beijing, China
2. National Center for Respiratory Medicine, Beijing, China
3. State Key Laboratory of Respiratory Health and Multimorbidity, Beijing, China
4. National Clinical Research Center for Respiratory Diseases, Beijing, China
5. Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Beijing, China
6. Department of Tobacco Control and Prevention of Respiratory Diseases, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, China
7. WHO Collaborating Centre for Tobacco Cessation and Respiratory Diseases Prevention, Beijing, China
8. Department of Pulmonary and Critical Care Medicine, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Taiyuan, China
9. Department of Pulmonary and Critical Care Medicine, First Affiliated Hospital of Xi’an Jiaotong University, Xi’an, China
10. Department of Pulmonary and Critical Care Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
11. Department of Pulmonary and Critical Care Medicine, Guizhou Provincial People’s Hospital, Guiyang, China
12. Department of Pulmonary and Critical Care Medicine, Zhongshan Hospital, Fudan University, Shanghai, China
13. Department of Pulmonary and Critical Care Medicine, First Hospital of China Medical University, Shenyang, China
14. State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Diseases, Guangzhou Institute of Respiratory Diseases, First Affiliated Hospital, Guangzhou Medical University, Guangzhou, China
15. Department of Pulmonary and Critical Care Medicine, Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China
16. State Key Laboratory of Biotherapy of China and Department of Respiratory and Critical Care Medicine, West China Hospital of Sichuan University, Chengdu, China
17. Beijing Key Laboratory of Respiratory and Pulmonary Circulation Disorders, Department of Pulmonary and Critical Care Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China
18. Department of Pulmonary and Critical Care Medicine, Peking University Third Hospital, Beijing, China
19. Department of Respiratory and Critical Care Medicine, Beijing Hospital, Beijing, China
20. National Center of Gerontology, Beijing, China
21. Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing, China
22. Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, China-Japan Friendship Hospital, Beijing, China
23. Center for Clinical Epidemiology and Evidence-Based Medicine, Beijing Children’s Hospital, Capital Medical University, Beijing, China
24. National Center for Children’s Health, Beijing, China
25. Department of Stomatology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China
26. Department of Epidemiology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China
27. Beijing Institute of Respiratory Medicine, Beijing, China
28. National Heart and Lung Institute, Imperial College London and Royal Brompton and Harefield NHS Trust, London, United Kingdom
Abstract

Background: Maternal smoking during pregnancy (MSDP) is a known risk factor for offspring developing chronic obstructive pulmonary disease (COPD), but the underlying mechanism remains unclear.

Objective: This study aimed to explore whether the increased COPD risk associated with MSDP could be attributed to tobacco dependence (TD).

Methods: This case-control study used data from the nationwide cross-sectional China Pulmonary Health study, with controls matched for age, sex, and smoking status. TD was defined as smoking within 30 minutes of waking, and the severity of TD was assessed using the Fagerstrom Test for Nicotine Dependence. COPD was diagnosed when the ratio of forced expiratory volume in 1 second to forced vital capacity was <0.7 in a postbronchodilator pulmonary function test according to the 2017 Global Initiative for Chronic Obstructive Lung Disease criteria. Logistic regression was used to examine the correlation between MSDP and COPD, adjusting for age, sex, BMI, educational attainment, place of residence, ethnic background, occupation, childhood passive smoking, residential fine particulate matter, history of childhood pneumonia or bronchitis, average annual household income, and medical history (coronary heart disease, hypertension, and diabetes). Mediation analysis examined TD as a potential mediator in the link between MSDP and COPD risk. The significance of the indirect effect was assessed through 1000 iterations of the “bootstrap” method.

Results: The study included 5943 participants (2991 with COPD and 2952 controls). Mothers of the COPD group had higher pregnancy smoking rates (COPD: n=305, 10.20%; controls: n=211, 7.10%; P<.001). TD was more prevalent in the COPD group (COPD: n=582, 40.40%; controls: n=478, 33.90%; P<.001). After adjusting for covariates, MSDP had a significant effect on COPD (β=.097; P<.001). There was an association between MSDP and TD (β=.074; P<.001) as well as between TD and COPD (β=.048; P=.007). Mediation analysis of TD in the MSDP-COPD association showed significant direct and indirect effects (direct: β=.094; P<.001 and indirect: β=.004; P=.03). The indirect effect remains present in the smoking population (direct: β=.120; P<.001 and indirect: β=.002; P=.03).

Conclusions: This study highlighted the potential association between MSDP and the risk of COPD in offspring, revealing the mediating role of TD in this association. These findings contribute to a deeper understanding of the impact of prenatal tobacco exposure on lung health, laying the groundwork for the development of relevant prevention and treatment strategies.

(JMIR Public Health Surveill 2024;10:e53170) doi:10.2196/53170

KEYWORDS
chronic obstructive pulmonary disease; tobacco dependence; smoking; mediating effects; lung function

Introduction

Chronic obstructive pulmonary disease (COPD) is a global health challenge, ranking among the leading causes of both morbidity and mortality worldwide [1]. It was estimated that COPD affected a substantial portion of the global population with a prevalence rate of 10.3% among individuals, and the number of individuals afflicted with COPD has been on the rise [2]. Notably, China accounts for nearly a quarter of global COPD cases, with a 67% increase in prevalence among individuals aged 40 years and older between 2012 and 2015 [3].

COPD is influenced by various factors, including genetics, lifestyle, environmental factors, and other influencing factors.
Smoking, recognized as the most significant cause of COPD, leads to its development in approximately half of all smokers [6,7]. The prenatal and perinatal phases are crucial for lung development, with maternal smoking linked to offspring susceptibility to various health issues [8-10]. A systematic review suggested that maternal smoking during pregnancy (MSDP) is associated with an increased risk of COPD [11]. MSDP may directly impact fetal development, influencing the formation of the central nervous system and respiratory system [12]. Given the vulnerability of embryos and fetuses to the external environment, maternal smoking may exert a more pronounced effect on fetal development [13].

Despite widespread recognition of the adverse effects of MSDP on offspring COPD [11], the underlying mechanisms remain incompletely elucidated. Tobacco dependence (TD) is a complex condition influenced by genetic and environmental factors, classified as a mental disorder according to the International Classification of Diseases, and TD should be recognized as a lethal noncommunicable disease [14]. Previous studies have identified TD as an independent risk factor for atherosclerosis [15]. Notably, TD is prevalent among smokers, often associated with heightened smoking intensity and lower cessation rates [16], and approximately 40% of smokers experience impaired lung function and develop COPD [17]. A meta-analysis indicates an elevated risk of smoking and TD in offspring associated with MSDP [18]. Given these associations, our hypothesis posits that TD plays a pivotal mediating role in the association between MSDP and COPD.

Therefore, this study aims to investigate whether TD serves as a mediator in the association between MSDP and the risk of COPD in offspring. To achieve this, we used data from the national cross-sectional China Pulmonary Health (CPH) study [3] and conducted a case-control study to explore the role of TD in the association between MSDP and COPD. The findings will contribute to the development of targeted interventions focused on TD.

**Methods**

**Participants and Study Design**

The study population was drawn from the national cross-sectional CPH study [3,19,20], the largest study assessing the burden of COPD in China. The CPH study, encompassing a vast nationwide cross-sectional examination, encompassed 57,779 Chinese adults aged 20 years and older across 10 provinces in China. This extensive study incorporated 80 urban and 80 rural areas, using a multistage stratified cluster sampling approach. The details of recruitment for the CPH can be found elsewhere [3]. Trained health workers in local community health centers administered standardized questionnaire surveys to all participants, capturing essential information on sociodemographic status, medical history, lifestyle, and more. The questionnaire information involved in this study can be obtained in Table S1 in Multimedia Appendix 1.

In this study, a case-control design was used, and the workflow is illustrated in Figure 1. Inclusion criteria for the case group were as follows: (1) provision of signed informed consent, (2) age of 18 years or older, and (3) availability of complete questionnaire information. Individuals with other conditions such as asthma, tuberculosis, lung cancer, lung inflammation, and missing information were excluded. Control participants were matched for age, sex, and smoking from cohort. The matching tolerance can be found in Table S2 in Multimedia Appendix 1. Ultimately, a total of 2991 patients in the COPD group and 2952 controls were included in this study.
**Definition of MSDP and Evaluation of TD**

MSDP was self-reported through a questionnaire, with all participants asked to respond to the question, “Did your mother smoke during pregnancy? (yes or no).” Upon awakening, smoking within the first 30 minutes was defined as indicative of TD, based on prior study [21]. The severity of TD was assessed using the Fagerstrom Test for Nicotine Dependence (FTND) [22]. FTND is a standardized tool for evaluating the degree of physical nicotine addiction. It consists of six items: (1) How soon after waking do you smoke your first cigarette (within 5 minutes, 6 to 30 minutes, 31 to 60 minutes, and after 60 minutes)? (2) Do you find it difficult to refrain from smoking in places where it is prohibited (yes or no)? (3) Which cigarette would you find most difficult to give up (the first one in the morning or any other)? (4) How many cigarettes do you smoke per day (10 or less, 11 to 20, 21 to 30, and 31 or more)? (5) Do you smoke more frequently during the first hour after wakening compared to the rest of the day (yes or no)? and (6) Do you smoke when you are so ill that you are confined to bed most of the day (yes or no)? The FTND generates a total score ranging from 0 to 10, with scores calculated as the sum of individual items. Yes or no items are scored as 0 or 1, and multiple-choice items are scored as 0 to 3. The questionnaire provided a comprehensive assessment of cigarette consumption, compulsion, and dependence. A higher total FTND score indicates a greater degree of TD. All participants reported their smoking status (yes or no). Smoking was defined as having consumed 100 cigarettes in one’s lifetime, encompassing both current and former smokers.

To address potential recall bias in self-reported MSDP, we implemented rigorous quality control measures. Our questionnaire design emphasized clarity and comprehensibility, and interviewers received detailed training to ensure precise definitions of smoking behavior. Memory aids were provided to aid accurate recall, and privacy protection measures were emphasized to minimize bias.
Definition of COPD
All study participants underwent postbronchodilator pulmonary function tests, including measurements of forced expiratory volume in 1 second (FEV$_1$) and forced vital capacity (FVC) and FEV$_1$/FVC ratio. COPD was diagnosed in accordance with the criteria set forth by the 2017 Global Initiative for Chronic Obstructive Lung Disease [23] if the participant’s FEV$_1$/FVC ratio was <0.7.

Covariates
This study incorporated several covariates, including age (years), sex (male or female), BMI (calculated as weight in kilograms divided by the square of height in meters), educational attainment (primary and below, junior middle school, senior high school, or bachelor and above), place of residence (urban or rural), ethnic background (Han Chinese or other), occupation (farmer, worker, or other), childhood passive smoking, residential fine particulate matter (PM$_{2.5}$) exposure, history of childhood pneumonia or bronchitis (yes or no), average annual household income, and medical history (coronary heart disease, hypertension, and diabetes).

Statistical Analysis
For normally distributed continuous variables, descriptive statistics are presented as mean (SD). Categorical variables are expressed as numbers and percentages. The Kolmogorov-Smirnov test was used to assess the normality of continuous variables. Comparisons of continuous variables were conducted using the Mann-Whitney U test, while categorical variables were compared using the chi-square test.

Mediation analysis, as outlined by Baron and Kenny [24], was used to investigate the potential mediating role of TD in the association between MSDP and the risk of COPD. The association between MSDP and TD was modeled using logistic regression, whereas the association with FTND scores was assessed using linear regression. The total effect of the initial variable on the outcome was defined as the sum of its direct and indirect effects. The indirect effect was further defined as the product of the initial variable’s impact on the intermediate variable and the intermediate variable’s impact on the outcome, with adjustments made for the initial variable. The statistical significance of the indirect effect was assessed through 1000 repetitions of a bootstrap procedure. Regarding the sensitivity analysis, we conducted a similar analysis by defining TD as a total FTND score ≥4. The sample size for this study was determined using the Monte Carlo method [25]. The details of the sample size calculation process are provided in Table S3 in Multimedia Appendix 1. All statistical analyses were conducted using R (version 4.2.1; R Foundation for Statistical Computing), with the use of packages such as bruceR, interactions, and mediation to estimate mediating effects. A significance level of $P=0.05$ (2-tailed) was considered statistically significant.

Ethical Considerations
The study received approval from the ethics review committee of Beijing Capital Medical University (11-KE-42) and other collaborating institutes. Written informed consent was obtained from all study participants in accordance with the principles outlined in the Declaration of Helsinki. To ensure privacy and confidentiality, the data used in this study underwent anonymization and deidentification processes.

Results
Participants
A total of 5943 participants were enrolled in this study, comprising 2991 individuals with COPD and 2952 matched controls. The demographic characteristics of all participants are elaborated in Table 1. In the control group, the mean age was 59.17 (SD 11.44) years, while in the COPD group, it was 59.50 (SD 11.66) years. Among controls, there were 1783 (60.40%) male participants, and in the COPD group, 1816 (60.70%) male participants. Smoking was observed in 1442 (48.20%) individuals within the COPD group and 1412 (47.80%) individuals in the control group. For matching purposes, no statistically significant differences were detected between the 2 groups in terms of participants’ age, sex, and smoking. However, a higher proportion of mothers in the COPD group reported MSDP compared to the control group (COPD: n=305, 10.20% and controls: n=211, 7.10%; $P<0.001$).
Table 1. Demographic characteristics of participants in this study.

<table>
<thead>
<tr>
<th></th>
<th>Controls (n=2952)</th>
<th>COPD$^a$ (n=2991)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>59.17 (11.44)</td>
<td>59.50 (11.66)</td>
<td>.27$^b$</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td>.82$^c$</td>
</tr>
<tr>
<td>Female</td>
<td>1169 (39.60)</td>
<td>1175 (39.30)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1783 (60.40)</td>
<td>1816 (60.70)</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m$^2$), n (%)</td>
<td></td>
<td></td>
<td>&lt;.001$^c$</td>
</tr>
<tr>
<td>&lt;18.5</td>
<td>96 (3.30)</td>
<td>129 (4.30)</td>
<td></td>
</tr>
<tr>
<td>18.5-25</td>
<td>1682 (57)</td>
<td>1849 (61.80)</td>
<td></td>
</tr>
<tr>
<td>≥25</td>
<td>1174 (39.80)</td>
<td>1013 (33.90)</td>
<td></td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001$^c$</td>
</tr>
<tr>
<td>Primary and below</td>
<td>971 (32.90)</td>
<td>1227 (41)</td>
<td></td>
</tr>
<tr>
<td>Junior middle school</td>
<td>988 (33.50)</td>
<td>987 (33)</td>
<td></td>
</tr>
<tr>
<td>Senior high school</td>
<td>620 (21)</td>
<td>493 (16.50)</td>
<td></td>
</tr>
<tr>
<td>Bachelor and above</td>
<td>373 (12.60)</td>
<td>284 (9.50)</td>
<td></td>
</tr>
<tr>
<td>Place of residence, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001$^c$</td>
</tr>
<tr>
<td>City</td>
<td>1977 (67)</td>
<td>1875 (62.7)</td>
<td></td>
</tr>
<tr>
<td>Countryside</td>
<td>975 (33)</td>
<td>1116 (37.3)</td>
<td></td>
</tr>
<tr>
<td>Nation, n (%)</td>
<td></td>
<td></td>
<td>.02$^c$</td>
</tr>
<tr>
<td>Han Chinese</td>
<td>2897 (98.10)</td>
<td>2906 (97.20)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>55 (1.90)</td>
<td>85 (2.80)</td>
<td></td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001$^c$</td>
</tr>
<tr>
<td>Farmer</td>
<td>794 (26.90)</td>
<td>1172 (39.20)</td>
<td></td>
</tr>
<tr>
<td>Worker</td>
<td>850 (28.80)</td>
<td>678 (22.70)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1308 (44.30)</td>
<td>1141 (38.10)</td>
<td></td>
</tr>
<tr>
<td>MSDP$^d$, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001$^c$</td>
</tr>
<tr>
<td>No</td>
<td>2741 (92.90)</td>
<td>2686 (89.80)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>211 (7.10)</td>
<td>305 (10.20)</td>
<td></td>
</tr>
<tr>
<td>Smoking, n (%)</td>
<td></td>
<td></td>
<td>.79$^c$</td>
</tr>
<tr>
<td>No</td>
<td>1540 (52.20)</td>
<td>1549 (51.80)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1412 (47.80)</td>
<td>1442 (48.20)</td>
<td></td>
</tr>
<tr>
<td>GOLD$^e$, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001$^f$</td>
</tr>
<tr>
<td>1</td>
<td>—</td>
<td>1685 (56.30)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>1103 (36.90)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>172 (5.80)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>31 (1)</td>
<td></td>
</tr>
<tr>
<td>FEV$_1$, mean (SD)</td>
<td>2.67 (0.66)</td>
<td>2.13 (0.70)</td>
<td>&lt;.001$^b$</td>
</tr>
<tr>
<td>FEV$_1$/FVC$^b$, ratio, mean (SD)</td>
<td>80.44 (5.63)</td>
<td>61.18 (9.00)</td>
<td>&lt;.001$^b$</td>
</tr>
<tr>
<td>AAHL$^j$, mean (SD)</td>
<td>1.34 (2.39)</td>
<td>1.11 (1.18)</td>
<td>&lt;.001$^b$</td>
</tr>
<tr>
<td>PM$_2.5$, mean (SD)</td>
<td>71.85 (13.64)</td>
<td>73.81 (15.95)</td>
<td>&lt;.001$^b$</td>
</tr>
<tr>
<td>Childhood passive smoking, n (%)</td>
<td></td>
<td></td>
<td>.049$^e$</td>
</tr>
</tbody>
</table>
### Differences of TD in Smokers Between the 2 Groups

As depicted in Table 2, the prevalence of TD was notably higher in the COPD group in comparison to the control group (COPD: n=582, 40.40%; controls: n=478, 33.90%; \(P=.001\)). Patients with COPD, when contrasted with the control group, exhibited a shorter time interval between waking up and the desire to smoke their first cigarette (Q1: \(P=.004\)) and encountered greater difficulty refraining from smoking in public places (Q2: \(P=.006\)). Among the COPD group, 19.40% (n=280) reported a tendency to smoke more frequently during the early hours after waking, whereas in the control group, this behavior was observed in 16.50% (n=233; \(P=.048\)). Additionally, the proportion of individuals smoking when unwell was higher in the COPD group than in the control group (Q6: \(P<.001\)). Q3 showed no statistically significant difference between the 2 groups, but with a \(P\) value <0.1, suggesting the potential presence of marginal effects.

---

<table>
<thead>
<tr>
<th>History of childhood pneumonia or bronchitis, n (%)</th>
<th>Controls (n=2952)</th>
<th>COPD(^a) (n=2991)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>1184 (40.1)</td>
<td>1276 (42.7)</td>
<td>(&lt;.001^c)</td>
</tr>
<tr>
<td>Yes</td>
<td>1768 (59.9)</td>
<td>1715 (57.3)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coronary heart disease, n (%)</th>
<th>Controls (n=2952)</th>
<th>COPD(^a) (n=2991)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>2842 (96.3)</td>
<td>2789 (93.2)</td>
<td>(.11^c)</td>
</tr>
<tr>
<td>Yes</td>
<td>110 (3.7)</td>
<td>202 (6.8)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypertension, n (%)</th>
<th>Controls (n=2952)</th>
<th>COPD(^a) (n=2991)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>2888 (97.8)</td>
<td>2906 (97.2)</td>
<td>(.007^c)</td>
</tr>
<tr>
<td>Yes</td>
<td>64 (2.2)</td>
<td>85 (2.8)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diabetes, n (%)</th>
<th>Controls (n=2952)</th>
<th>COPD(^a) (n=2991)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>2851 (96.6)</td>
<td>2895 (96.8)</td>
<td>(.70^c)</td>
</tr>
<tr>
<td>Yes</td>
<td>101 (3.4)</td>
<td>96 (3.2)</td>
<td></td>
</tr>
</tbody>
</table>

---

\(^a\)COPD: chronic obstructive pulmonary disease.
\(^b\)\(P\) value for 2-tailed \(t\) test.
\(^c\)\(P\) value for chi-square test.
\(^d\)MSDP: maternal smoking during pregnancy.
\(^e\)GOLD: Global Initiative for Chronic Obstructive Lung Disease.
\(^f\)Not available.
\(^g\)FEV\(_1\): forced expiratory volume in 1 second.
\(^h\)FVC: forced vital capacity.
\(^i\)AAHI (average annual household income) is measured in 10,000 Yuan (a currency exchange rate of 1 Chinese Yuan (CNY)=US $0.16147 is applicable).
\(^j\)PM\(_{2.5}\): fine particulate matter.
Table 2. Differences between the COPD\(^a\) and control group in the FTND\(^b\) among smokers.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Controls (n=1412), n (%)</th>
<th>COPD (n=1442), n (%)</th>
<th>(P) value(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1(^d) After 60 minutes</td>
<td>784 (55.50)</td>
<td>728 (50.50)</td>
<td>.004</td>
</tr>
<tr>
<td>31 to 60 minutes</td>
<td>150 (10.60)</td>
<td>133 (9.20)</td>
<td></td>
</tr>
<tr>
<td>6 to 30 minutes</td>
<td>186 (13.20)</td>
<td>235 (16.30)</td>
<td></td>
</tr>
<tr>
<td>Within 5 minutes</td>
<td>292 (20.70)</td>
<td>347 (24.10)</td>
<td></td>
</tr>
<tr>
<td>Q2(^e) No</td>
<td>1122 (79.50)</td>
<td>1082 (75)</td>
<td>.006</td>
</tr>
<tr>
<td>Yes</td>
<td>290 (20.50)</td>
<td>360 (25)</td>
<td></td>
</tr>
<tr>
<td>Q3(^f) Any other</td>
<td>1059 (75)</td>
<td>1035 (71.80)</td>
<td>.06</td>
</tr>
<tr>
<td>The first one in the morning</td>
<td>353 (25)</td>
<td>407 (28.20)</td>
<td></td>
</tr>
<tr>
<td>Q4(^g) 10 or less</td>
<td>1228 (87)</td>
<td>1247 (86.50)</td>
<td>.62</td>
</tr>
<tr>
<td>11 to 20</td>
<td>141 (10)</td>
<td>139 (9.60)</td>
<td></td>
</tr>
<tr>
<td>21 to 30</td>
<td>19 (1.30)</td>
<td>22 (1.50)</td>
<td></td>
</tr>
<tr>
<td>31 or more</td>
<td>24 (1.70)</td>
<td>34 (2.40)</td>
<td></td>
</tr>
<tr>
<td>Q5(^h) No</td>
<td>1179 (83.50)</td>
<td>1162 (80.60)</td>
<td>.048</td>
</tr>
<tr>
<td>Yes</td>
<td>233 (16.50)</td>
<td>280 (19.40)</td>
<td></td>
</tr>
<tr>
<td>Q6(^i) No</td>
<td>1221 (86.50)</td>
<td>1168 (81)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>191 (13.50)</td>
<td>274 (19)</td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>1.92 (2.03)</td>
<td>2.25 (2.14)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TD(^j) No</td>
<td>934 (66.10)</td>
<td>860 (59.60)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>478 (33.90)</td>
<td>582 (40.40)</td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\)COPD: chronic obstructive pulmonary disease.  
\(^{b}\)FTND: Fagerstrom Test for Nicotine Dependence.  
\(^{c}\)\(P\) value for chi-square test.  
\(^{d}\)Q1: How soon after you wake up do you smoke your first cigarette?  
\(^{e}\)Q2: Do you find it difficult to refrain from smoking in places where it is forbidden?  
\(^{f}\)Q3: Which cigarette would you find most difficult to give up?  
\(^{g}\)Q4: How many cigarettes do you smoke per day?  
\(^{h}\)Q5: Do you smoke more frequently during the first hour after waking compared to the rest of the day?  
\(^{i}\)Q6: Do you smoke when you are so ill that you are confined to bed most of the day?  
\(^{j}\)TD: tobacco dependence.

**Direct and Indirect Effects of MSDP on COPD in All Participants**

The results of the mediation analysis for all participants are presented in Table 3. After adjusting for age, sex, BMI, educational attainment, place of residence, ethnicity, occupation, childhood passive smoking, residential PM\(_{2.5}\) exposure, history of childhood pneumonia or bronchitis, average annual household income, and medical history (coronary heart disease, hypertension, and diabetes), the total effect of MSDP on COPD was statistically significant (\(\beta=.097; P<.001\)). Furthermore, the direct effect of MSDP on COPD remained significant after adjusting for TD and other covariates (\(\beta=.094; P<.001\)). The indirect effect of TD was also found to be statistically significant.
The sensitivity analysis revealed similar results (Table S4 in Multimedia Appendix 1).

### Table 3. Mediating effect of TD\(^a\) (mediating variable) on the association between MSDP\(^b\) (independent variable) and COPD\(^c\) (dependent variable) in all participants\(^d\).

<table>
<thead>
<tr>
<th>Path</th>
<th>(\beta) (SE)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSDP on COPD(^e)</td>
<td>.097 (0.023)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>MSDP on TD(^f)</td>
<td>.074 (0.017)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TD on COPD given MSDP(^g)</td>
<td>.048 (0.018)</td>
<td>.007</td>
</tr>
<tr>
<td>Indirect: MSDP on COPD(^h)</td>
<td>.004 (0.002)</td>
<td>.03</td>
</tr>
<tr>
<td>Direct: MSDP on COPD given TD(^i)</td>
<td>.094 (0.023)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(a\)TD: tobacco dependence.
\(b\)MSDP: maternal smoking during pregnancy.
\(c\)COPD: chronic obstructive pulmonary disease.
\(d\)All analyses adjusted for age, sex, BMI, educational attainment, place of residence, ethnicity, occupation, childhood passive smoking, residential PM\(_{2.5}\) exposure, history of childhood pneumonia or bronchitis, average annual household income, and medical history (coronary heart disease, hypertension, and diabetes).
\(e\)Total effect of independent variables on dependent variables.
\(f\)Coefficients of independent variables on mediating variables after adjustment for covariates.
\(g\)Coefficients of mediating variables on dependent variables after adjusted for covariates and IVs.
\(h\)Indirect effect of independent variables on dependent variables.
\(i\)Direct effect of independent variables on dependent variables.

### Direct and Indirect Effects of MSDP on COPD in Smokers

The results of the mediation analysis for smokers are presented in Table 4. In this segment, all analyses were adjusted for age, sex, BMI, educational attainment, place of residence, ethnicity, occupation, childhood passive smoking, residential PM\(_{2.5}\) exposure, history of childhood pneumonia or bronchitis, average annual household income, and medical history (coronary heart disease, hypertension, and diabetes). Among smokers, both the total effect and direct effect (\(\beta\)) of MSDP on COPD were .123 and .120, respectively. Additionally, the indirect effect of TD was found to be statistically significant (\(\beta=.003\); \(P=.03\)). The sensitivity analysis revealed similar results (Table S5 in Multimedia Appendix 1).

### Table 4. Mediating effect of TD\(^a\) (mediating variable) on the association between MSDP\(^b\) (independent variable) and COPD\(^c\) (dependent variable) in smokers\(^d\).

<table>
<thead>
<tr>
<th>Path</th>
<th>(\beta) (SE)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSDP on COPD(^e)</td>
<td>.123 (0.032)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>MSDP on TD(^f)</td>
<td>.050 (0.031)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TD on COPD given MSDP(^g)</td>
<td>.059 (0.019)</td>
<td>.004</td>
</tr>
<tr>
<td>Indirect: MSDP on COPD(^h)</td>
<td>.003 (0.002)</td>
<td>.03</td>
</tr>
<tr>
<td>Direct: MSDP on COPD given TD(^i)</td>
<td>.120 (0.031)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(a\)TD: tobacco dependence.
\(b\)MSDP: maternal smoking during pregnancy.
\(c\)COPD: chronic obstructive pulmonary disease.
\(d\)All analyses adjusted for age, sex, BMI, educational attainment, place of residence, ethnicity, occupation, childhood passive smoking, residential PM\(_{2.5}\) exposure, history of childhood pneumonia or bronchitis, average annual household income, and medical history (coronary heart disease, hypertension, and diabetes).
\(e\)Coefficients of independent variables on dependent variables after correction for covariates.
\(f\)Coefficients of independent variables on mediating variables after adjustment for covariates.
\(g\)Coefficients of mediating variables on dependent variables after correction for covariates and IVs.
\(h\)Indirect effect of independent variables on dependent variables.
\(i\)Direct effect of independent variables on dependent variables.
Discussion

Principal Findings

The aim of this study was to investigate whether TD plays a role in the association between MSDP and COPD in offspring. A case-control study was conducted within the Chinese population. It was the first study to observe that MSDP has both a direct impact on COPD and an indirect influence mediated through TD. This mediation effect persists among the smoking population, and sensitivity analysis confirms the stability of these results.

It is well-established that smoking represents the primary risk factor for COPD, and an increasing body of evidence suggests that early life tobacco exposure plays a role in the onset and progression of COPD [26]. MSDP, as highlighted by Jaakkola et al [27], exerts substantial adverse effects on conditions such as asthma, chronic bronchitis, and chronic respiratory symptoms. Nevertheless, limited research has delved into the impact of prenatal tobacco smoke exposure on the development of COPD in later adulthood. This study, however, reveals an association between MSDP and the risk of COPD in offspring. Early life tobacco exposure has enduring consequences on lung function in offspring, with suboptimal intrauterine conditions leading to disturbances in lung development. As a result, affected individuals exhibit diminished lung function at birth, which often persists throughout their lifetime. This, in turn, elevates the risk of subsequent COPD, particularly in childhood wheeze disorders and genetically susceptible individuals [28]. To curb the current upward trend in COPD incidence, it is imperative to mitigate lung development risks by enhancing antenatal and neonatal care and reducing exposure to environmental pollutants, including passive tobacco smoke, both prenatally and postnatally.

In this study, MSDP emerges as a risk factor for TD in offspring, a finding consistent with prior research. Previous studies have established that individuals with a history of MSDP face an elevated risk of TD in adolescence and adulthood [18]. The precise mechanism underlying the connection between MSDP and offspring TD remains elusive. One plausible mechanism could involve the neurotoxic effects of harmful compounds present in tobacco smoke, which can readily traverse the placenta, potentially leading to smoking behavior and dependence in offspring [29]. Another conceivable mechanism is that MSDP is linked to preterm birth [30] and low birth weight [31], both of which could predispose offspring to various health challenges. Furthermore, it is essential for health care professionals to recognize that TD may be linked to early life tobacco exposure. Therefore, children exposed to MSDP should receive comprehensive education on topics such as diet, exercise, and smoking avoidance to mitigate their risk of developing COPD in adulthood.

This study revealed a direct association between TD and COPD. Previous research has identified smoking as one of the primary risk factors for COPD, causing inflammation and damage to the airways and lung tissue, ultimately leading to the development of COPD [32]. Recent studies have further suggested the correlation between TD and COPD [33]. Additionally, other studies suggest that TD may exacerbate the severity and progression of COPD [34]. Among individuals diagnosed with COPD, the presence of TD makes quitting smoking more challenging [35]. Thus, this study underscores the critical importance of addressing TD in the management and mitigation of COPD.

To our knowledge, we have discovered for the first time in this study that TD plays a mediating role in the association between MSDP and COPD. However, the specific reasons underlying this mediating mechanism remain unclear. There are several possible mechanisms to consider. First, genetic factors may be implicated, as MSDP has the potential to induce persistent epigenetic changes until adolescence [36]. These genetic alterations may contribute to the development of TD [37]. Consequently, TD can exacerbate challenges in smoking cessation and the progression of COPD [38]. Second, environmental factors may also be involved. Mothers who smoke during pregnancy may continue smoking during infancy, creating an early growth environment that promotes the risk of TD in offspring during lung development [39,40], subsequently leading to COPD. Our results highlight that TD serves as a partial mediator in the connection between MSDP and offspring COPD. Consequently, it becomes imperative to consider TD when addressing MSDP and smoking cessation efforts in offspring. Prior studies have advocated smoking cessation strategies for TD smokers, incorporating the use of medications such as varenicline, nicotine replacement therapy products, and bupropion [41]. Hence, clinicians should be cognizant of the necessity for appropriate smoking cessation interventions among individuals displaying signs of TD, ultimately contributing to the reduction of COPD risk.

Limitations

This study bears significant implications as it substantiates the partial mediating role of TD in the link between MSDP and COPD, underlining the importance of developing interventions targeting TD to mitigate the incidence of COPD in offspring. Nonetheless, it is essential to acknowledge the study’s limitations. First, the data concerning MSDP relied on questionnaire responses, introducing a potential source of recall bias. Due to the presence of recall bias, we cannot entirely rule out the possibility of uncertainty or erroneous memory in participants when reporting maternal smoking behavior during pregnancy. To enhance the accuracy of capturing MSDP information, future studies may consider using prospective cohort designs. Second, the absence of genetic data represents a significant limitation in this study. Despite the inclusion of various participant covariates, the intricate role of genetic factors in influencing COPD was not considered due to data constraints. Future research endeavors could overcome this limitation by integrating comprehensive genetic analyses into the study design. Finally, our classification of nonsmokers as nonnicotine-dependent during the overall analysis might introduce some bias. To fortify the robustness of our findings, we conducted a parallel subgroup analysis among the smoking population that consistently demonstrated the robustness of our results.
Conclusions
MSDP exerts an adverse influence on the risk of COPD in offspring, and TD plays a partially mediating role in this association. These findings underscore the potential for clinicians to mitigate the impact of MSDP on COPD in offspring by addressing TD. Consequently, early intervention strategies aimed at reducing TD become imperative in the endeavor to mitigate the risk of COPD.

Acknowledgments
The authors thank the participants of the China Pulmonary Health study. For continuous support, assistance, and cooperation, the authors thank Xinran Zhang and Xiaoying Gu (China-Japan Friendship Hospital); Ping Li, Mengyu Cheng, Wen Han, Hu Liu, Pengfei Wang, Jing Li, and Jing Wu (Shanxi Bethune Hospital Shanxi Academy of Medical Sciences); Yang Li and Ting Liu (First Affiliated Hospital of Xi’an Jiaotong University); Liekou Ma (Fufeng People’s Hospital); Luoping Yuan (Luohan Hospital); Jianbo Liu (Jingyang Hospital); Hubin Xi (Shanxi People’s Hospital); Baoping Wu (Changning Township Hospital of Wugong); Zhifang Liu (Meixian People’s Hospital); Cheng Zhang, Hong Yu, Weijia Liu, Ruiming Wu, Li Zhao, Yanjun Jin, Lu Zhang, Mengning Zhen, Ping Lu, and Ling Li (Guizhou Provincial People’s Hospital); Bin Wu and Weimin Yao (Affiliated Hospital of Guangdong Medical University); Yanjie Yang, Li Manning, Qi Chen, Ying Gong, Mingfei Zhang, and Sulan Wei (Shanghai Zhongshan Hospital); Zhihua Chen, Gang Huang, Niya Zhou, Bin Shen, Wen Hua, Bin Zhang, Youlian Yu, and Juan Xiong (Second Affiliated Hospital of Zhejiang University); Ting Yang, Yongchun Shen, Dianlian Li, Hongyu Long, Zenglin Liao, Xiaolu Li, Yanqiu Wu, Xiang Tong, Xiaying Peng, Bo Wang, Zhixin Qiu, Jian Luo, Lanlan Zhang, Shuang Zhao, Xingyu Xiong, Yinyin Yang, Yalun Li, Yanqi He, Faming Jiang, Ting Wang, Jiajia Dong, Jing An, Linwei Li, Lian Liu, and Yonggang Zhang (West China Hospital); Liqun Jia (First Affiliated Hospital of Zhengzhou University); Caishuang Peng (Chongqing Cancer Hospital); Qianjing Hu (People’s Hospital of Yubei District of Chongqing); Shujun Guo and Xiaying Peng (Sichuan Province People’s Hospital); Min Li (First Affiliated Hospital of Kunming Medical University); Lingli Guo (North China University of Science and Technology); Xue Zhang (Luoyang Orthopedic Hospital of Henan Province); Wen Du (West China Teaching Hospital); Yinyin Yang (Chengdu Second People’s Hospital); Lin Li (Mianyang Central Hospital); Jingyu Quan, Baosen Pang, Min Zhu, Xiaohong Chang, Jun Zhang, Baomei Wu, Ping Xin, Xiaoxia Huang, Zhiyuan An, Shuilian Chu, Quyuan Liu, Yanrui Jia, Jie Xie, Ying Cui, Jing Zhao, Chunyan Zhang, Jingyu Yang, and Xu Wu (Beijing Institute of Respiratory Medicine and Beijing Chao-Yang Hospital); Ruiyan Lin, Jie Song, Zhenyang Xu, and Xiaohui Li (Beijing Luhe Hospital); Rui Wu, Yanling Ding, Ming Lu, Jing Zhang, Lina Sun, Chengcheng Liao, Yun Sun, Yixuan Luo, Fan Lin, Yu Bai, and Meng Wang (Peking University Third Hospital); Lei Wang and Rong Gao (Beijing University of Traditional Chinese Medicine); Lingbo Sun and Xiaoliang Xie (Peking University Health Science Center Primary Care Center); and Pengjun Zhang, Hongsheng Zhang, Di Chai, and Xiaomeng Li (Beijing Chao-Yang Hospital). The study was funded by the Chinese Academy of Medical Sciences (CAMS) Initiative for Innovative Medicine (CAMS 2021-I2M-1-010) and the Special Research Foundation for Public Welfare of Health, Ministry of Health of China (grant 201002008).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
JL, DX, KH, TY, JX, LY, J Zhao, Xiangyan Zhang, CB, JK, PR, HS, FW, YC, TS, GS, Y Lin, GH, SW, J Zhu, JH, and CW conceived and designed the study. CW supervised the study. JL did the statistical analysis. All authors contributed to data collection and interpretation. JL, DX, and CW drafted the report. All authors revised the report and approved the final version before submission.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary tables.

References

https://publichealth.jmir.org/2024/1/e53170


Abbreviations

COPD: chronic obstructive pulmonary disease

CPH: China Pulmonary Health

FEV1: forced expiratory volume in 1 second

FTND: Fagerstrom Test for Nicotine Dependence

FVC: forced vital capacity

MSDP: maternal smoking during pregnancy

PM2.5: fine particulate matter

TD: tobacco dependence
Mediating Effect of Tobacco Dependence on the Association Between Maternal Smoking During Pregnancy and Chronic Obstructive Pulmonary Disease: Case-Control Study

JMIR Public Health Surveill 2024;10:e53170

URL: https://publichealth.jmir.org/2024/1/e53170
doi:10.2196/53170
PMID:38386387
Behavioral Intention of Receiving Monkeypox Vaccination and Undergoing Monkeypox Testing and the Associated Factors Among Young Men Who Have Sex With Men in China: Large Cross-Sectional Study

Sitong Luo\textsuperscript{1,2}, PhD; Kedi Jiao\textsuperscript{1}, PhD; Yuhang Zhang\textsuperscript{1}, MEd; Yutong Xu\textsuperscript{1}, MSc; Jingtao Zhou\textsuperscript{1}, MSc; Siwen Huang\textsuperscript{1}, MPH; Yan Li\textsuperscript{3}, PhD; Yongkang Xiao\textsuperscript{4}, PhD; Wei Ma\textsuperscript{5}, PhD; Lin He\textsuperscript{6}, MSc; Xianlong Ren\textsuperscript{2}, PhD; Zhen Dai\textsuperscript{8}, MSc; Jiaruo Sun\textsuperscript{1}, BMgt; Qingyu Li\textsuperscript{1}, BM; Feng Cheng\textsuperscript{1,2}, PhD; Wannian Liang\textsuperscript{1,2}, PhD

\textsuperscript{1}Vanke School of Public Health, Tsinghua University, Beijing, China
\textsuperscript{2}Institute for Healthy China, Tsinghua University, Beijing, China
\textsuperscript{3}Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China
\textsuperscript{4}Department of Acute Infectious Diseases Control and Prevention, Anhui Provincial Center for Disease Control and Prevention, Hefei, China
\textsuperscript{5}Department of Epidemiology, School of Public Health, Shandong University, Jinan, China
\textsuperscript{6}Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
\textsuperscript{7}Department of AIDS/STD Control and Prevention, Beijing Center for Disease Control and Prevention, Beijing, China
\textsuperscript{8}Department of AIDS/STD Control and Prevention, Chengdu Center for Disease Control and Prevention, Chengdu, China

Corresponding Author:
Sitong Luo, PhD
Vanke School of Public Health
Tsinghua University
30 Shuangqing Rd, Hai Dian Qu
Beijing, 100084
China
Phone: 86 010 62791154
Email: sitongluo@tsinghua.edu.cn

Abstract

Background: The worldwide human monkeypox (mpox) outbreak in 2022 mainly affected men who have sex with men (MSM). In China, young men who have sex with men (YMSM) were at a potential high risk of mpox infection due to their sexual activeness and the eased COVID-19 restrictions at the end of 2022.

Objective: This study aimed to investigate the behavioral intention of receiving mpox vaccination and undergoing mpox testing in 4 different scenarios and explore their associations with background and behavioral theory–related factors among Chinese YMSM.

Methods: An online cross-sectional survey was conducted among YMSM aged 18-29 years from 6 representative provinces of China in September 2022. Participants recruited (recruitment rate=2918/4342, 67.2\%) were asked to self-administer an anonymous questionnaire designed based on prior knowledge about mpox and classic health behavior theories. Data on the participants’ background, mpox knowledge and cognition, mpox vaccination and testing cognition, and the behavioral intention of receiving mpox vaccination and undergoing mpox testing were collected. Descriptive analysis and univariate and multivariate linear regressions were performed. Geodetector was used to measure the stratified heterogeneity of behavioral intention.

Results: A total of 2493 YMSM with a mean age of 24.6 (SD 2.9) years were included. The prevalence of having a behavioral intention of receiving mpox vaccination ranged from 66.2\% to 88.4\% by scenario, varying in epidemic status and cost. The prevalence of having an mpox testing intention was above 90\% in all scenarios regardless of the presence of symptoms and the cost. The positive factors related to vaccination intention included mpox knowledge (b\textsubscript{a}=0.060, 95\% CI 0.016-0.103), perceived susceptibility of mpox (b\textsubscript{a}=0.091, 95\% CI 0.035-0.146), perceived severity of mpox (b\textsubscript{a}=0.230, 95\% CI 0.164-0.296), emotional distress caused by mpox (b\textsubscript{a}=0.270, 95\% CI 0.160-0.380), perceived benefits of mpox vaccination (b\textsubscript{a}=0.455, 95\% CI 0.411-0.498), self-efficacy of mpox vaccination (b\textsubscript{a}=0.586, 95\% CI 0.504-0.668), and having 1 male sex partner (b\textsubscript{a}=0.452, 95\% CI 0.098-0.806).
Introduction

Since May 2022, rapidly growing numbers of human monkeypox (moxp) cases have been reported from countries where the disease was not initially endemic [1,2]. The World Health Organization (WHO) declared the worldwide moxp outbreak a public health emergency of international concern on July 23 [3]. By September 28, 2023, a total of 90,618 confirmed cases were reported across 115 countries, with 1484 confirmed cases in China. Among cases with known data on sexual orientation, 83.2% patients were identified as men who have sex with men (MSM) worldwide [4]. In China, more than 90% cases were identified as MSM [5]. One potential reason for MSM being the most affected population is that moxp transmission can result from close contact with the skin lesions of an infected person [1] and MSM communities usually have highly interconnected sexual networks [6].

China has the largest MSM population in the world, with an estimated size of 8.3 million [7], and among them, approximately 1.9 million are young people attending universities [8]. The first moxp case in mainland China was reported on September 16, 2022, and the patient was an MSM. The risk of an moxp outbreak became a public concern due to the large MSM population and the eased COVID-19 restrictions (eg, travel restrictions, social distancing) [9]. WHO recommended that the state parties with recently imported cases (eg, travel restrictions, social distancing) [9]. By September 28, 2023, a total of 90,618 confirmed cases were reported across 115 countries, with 1484 confirmed cases in China. Among cases with known data on sexual orientation, 83.2% patients were identified as men who have sex with men (MSM) worldwide [4]. In China, more than 90% cases were identified as MSM [5]. One potential reason for MSM being the most affected population is that moxp transmission can result from close contact with the skin lesions of an infected person [1] and MSM communities usually have highly interconnected sexual networks [6].

Case detection and targeted immunization are powerful measures to fight the moxp outbreak [10-12]. Timely moxp testing can reduce moxp transmission, avoid delays in moxp treatment, and decrease the cost of health care [11-15]. Vaccination, such as that against smallpox, has been demonstrated to be highly cost-effective in preventing moxp infection [16]. Several classic theories postulate factors causing health-related behaviors and the underlying mechanisms. For example, the knowledge attitude practice (KAP) model [17] argues that knowledge of a disease and the related preventive/risk behaviors influences an individual’s attitude toward the behaviors and, in turn, affects the behaviors. The Health Belief Model (HBM) [18] proposes that people are more likely to take preventative actions if they perceive the threat of a health risk to be serious, feel they are personally susceptible, and perceive fewer costs than benefits of engaging in it. The theory of planned behavior (TPB) [19] says that attitudes, social norms, and perceived control of behaviors are predictors of behavioral intention. According to these theories and previous studies on other sexually associated infectious diseases, several factors may affect the behavioral intention of receiving moxp vaccination and undergoing moxp testing among MSM, including general factors (eg, sociodemographic characteristics, risk behaviors, disease knowledge, perceptions of health behaviors) and MSM-specific factors (eg, stigma) [20,21]. Although a large body of literature has examined preventive behaviors of diseases associated with MSM (eg, HIV/AIDS, sexually transmitted diseases [STDs], human papillomavirus [HPV]), the situation of moxp might be different, as it is a newly emerging disease in MSM communities. To timely inform the moxp prevention and control strategies in China, a comprehensive investigation on the behavioral intention of adopting moxp prevention measures and associated factors is warranted.

This study aimed to investigate the levels of the behavioral intention of receiving moxp vaccination and undergoing moxp testing among Chinese young men who have sex with men (YMSM) aged 18-29 years and to explore the associated factors. This age group was selected as it became the population with a high risk of moxp infection, especially after COVID-19 travel restrictions were eased in China. YMSM are the most socially active subgroup within MSM communities [22], have relatively more opportunities to meet people from abroad due to study or business reasons, and have not received smallpox vaccination, as China ended this immunization program in 1981 [23]. The hypotheses of potential factors were guided by the KAP model [17], the HBM [18], and the TPB [19].

Methods

Study Design

An online cross-sectional survey was conducted in September 2022 with participants from 6 provincial regions of China, which represented the country geographically and socioeconomically (based on the per capita gross domestic product [GDP] of 2020) to some extent [7,24]. These 6 regions were Beijing (the capital;
per capita GDP=Chinese yuan [CNY] 164,889 [US $22,932]), Zhejiang (east; per capita GDP=CNY 100,620 [US $13,994]), Sichuan (west; per capita GDP=CNY 58,126 [US $8084]), Guangdong (south; per capita GDP=CNY 88,521 [US $12,311]), Shandong (north; per capita GDP=CNY 72,151 [US $10,034]), and Anhui (central; per capita GDP=CNY 63,426 [US $8821]). Figure 1 shows the geographic location, per capita GDP, population size, and sample size of the study areas.

Figure 1. Map of the geographic location, per capita GDP, population size, and sample size of the 6 study sites in China. Note: The per capita GDP and population size of each study site were based on the data released by the National Bureau of Statistics of China in 2020. GDP: gross domestic product.

Participant Recruitment and Data Collection

The inclusion criteria were male participants, age 18-29 years, having ever had anal or oral sex with men, and ability to fill an electronic questionnaire and provide oral informed consent. The exclusion criteria were psychiatric disorders based on the recruiters’ report and refusal to participate.

The facility-based sampling method was used to recruit participants from each province with the assistance of local MSM-oriented community-based organizations (CBOs). The selected CBOs have rich experience in providing health services (eg, HIV/sexually transmitted infection [STI] testing) and conducting public health studies in MSM communities. First, CBO staff were trained as recruiters and asked to disseminate the invitation message to potentially eligible YMSM in their networks via commonly used messaging apps (eg, WeChat). YMSM who were interested in the study were then screened for eligibility and fully informed about the study. Next, they were sent a quick response (QR) code and a 1-time password to access the electronic anonymous questionnaire. The self-administered questionnaire took about 10-15 minutes to complete.

The recruitment rate of the study (number of returned questionnaires/number of invitations sent) was 67.2% (2918/4342). In each province, the recruitment rate ranged from 44.5% to 97.2%. The main reasons for nonparticipation (n=1424, 32.8%) reported by the recruiters were not meeting the inclusion criteria, lack of time, and privacy concerns. The 2918 participants completed the questionnaire, and 425 (14.6%) were excluded from analysis due to a failure in the quality check. The final sample size was 2493 (85.4%) participants.

Ethical Considerations

The study was approved by the Institutional Review Board of Tsinghua University, China (reference number: 20220140). Each participant was asked to carefully read the informed consent form before starting the survey, in which the objective, procedure, potential risks and benefits, and voluntary nature of the study were clearly stated. Oral informed consent was waived to protect the participants’ privacy. The questionnaire was fully anonymous and self-administered by the participants through a secure privacy-compliant platform (Wenjuanxing survey platform [25]). Each participant whose submitted questionnaire passed the quality check receive Chinese yuan [CNY] 15 (about US $2.10) via the survey platform for compensation of their time and effort. To ensure confidentiality, further deidentified data were used for analysis.

Measures

Behavioral Intention of Receiving Mpox Vaccination and Undergoing Mpox Testing

All the measures are described in detail in Multimedia Appendix 1. Regarding behavioral intention, the participants were asked to rate their likelihood (1=very unlikely to 5=very likely) of receiving mpox vaccination in 4 scenarios. A summative score was calculated, with a higher score indicating a higher level of intention (range 4–20, α=.90). Similarly, the participants were asked to rate their likelihood of undergoing an mpox test in 4
scenarios, with a higher summative score indicating a higher level of intention (range 4-20, $\alpha=0.94$).

**Mpx Knowledge**

Mpx knowledge was measured with a single yes/no question (“Have you heard of mpx?”) and 10 knowledge questions. A composite score of the knowledge scale was constructed by counting the number of correct responses to the questions; those who had not heard of mpx were assigned a score of 0 (score range 0-10). A higher score indicated a higher level of knowledge ($\alpha$ for the 10 knowledge questions=0.77).

**Mpx Cognition**

The perceived susceptibility of mpx was measured with 3 items, with responses ranging from 1 for very unlikely to 5 for very likely. A higher total score implied a higher level of perceived susceptibility (range 3-15, $\alpha=0.83$). The perceived severity of mpx was measured with 2 items, with responses ranging from 1 for very mild to 5 for very strong. A higher summative score indicated a higher level of perceived severity (range 2-10, $\alpha=0.90$). Emotional distress (eg, panic, anxiety) caused by the mpx outbreak was measured with a single item, with the response ranging from 1 for very mild to 5 for very strong.

**Cognition of Mpx Vaccination and Testing**

Participants were asked about the perceived benefits, barriers, and self-efficacy of mpx vaccination/testing, with responses ranging from 1 for strongly disagree to 5 for strongly agree. The perceived benefits of mpx vaccination were assessed with 4 items (range 4-20, $\alpha=0.90$), and the perceived benefits of mpx testing were measured with 3 items (range 3-15, $\alpha=0.91$), with a higher summative score indicating more perceived benefits. The perceived barriers to mpx vaccination were measured with 4 items (range 4-20, $\alpha=0.81$), and the perceived barriers to mpx testing were measured with 5 items (range 5-25, $\alpha=0.82$), with a higher total score indicating more perceived barriers. The self-efficacy of mpx vaccination was measured with 2 items, with a higher summative score indicating a higher level of self-efficacy (range 2-10, $\alpha=0.82$). The self-efficacy of mpx testing was similarly measured (range 2-10, $\alpha=0.85$).

**Background Information**

Background information was collected, including (1) sociodemographic characteristics, (2) mpx risk behaviors in the past 6 months (eg, overseas travel history, number of male sex partners, frequency of in-person gatherings with MSM, use of any psychoactive substance before/during sexual intercourse [chemsex]), (3) diagnosis of HIV and STDs, and (4) presence of mpx-related symptoms in the past 2 weeks.

**Statistical Analysis**

First, descriptive analyses were conducted, with means and SDs being reported for continuous variables and frequencies and percentages being reported for categorical variables. Second, the q-statistic Geodetector, a spatial variance analysis method that explains nonlinear associations, was used to measure and attribute the stratified heterogeneity [26]. The factor detector was applied to detect the spatial association between explanatory variables (ie, study sites, sociodemographic characteristics, and mpx risk behaviors) and response variables (ie, behavioral intention of receiving mpx vaccination and undergoing mpx testing). The q-statistic and P value of Geodetector were reported. Third, simple regression analyses were performed to explore the crude associations between the background/behavioral theory–related factors and behavioral intention. Fourth, multivariate linear regressions were performed, involving all the background and theoretical factors as independent variables. The point estimate and 95% CIs were reported. SPSS Statistics 24.0 (IBM Corporation) was used. Geodetector analysis was conducted using R version 4.3.1 (R Foundation for Statistical Computing).

**Results**

**Background Characteristics**

The participants’ mean age was 24.6 (SD 2.9) years. Most of the participants were Han Chinese (n=2392, 95.9%), employed (n=1755, 70.4%), and unmarried (n=2380, 95.5%); had completed college education or above (n=2086, 83.7%), reported a monthly income under CNY 6000 (US $836; n=1530, 61.3%), and self-identified as homosexual (n=2023, 81.1%). In the past 6 months, 51 (2%) reported an overseas travel history, 1336 (53.6%) had in-person gatherings with MSM for at least once, 1249 (50.6%) had more than 1 male sex partner, and 541 (21.7%) reported having had chemsex with MSM. A total of 2383 (95.6%) participants had ever undergone an HIV test, and 279 (11.2%) self-reported as HIV positive. In addition, 354 (14.2%) participants reported a history of STIs. Furthermore, 1923 (77.1%) participants reported they did not have any mpx-like symptoms in the past 2 weeks, and 113 (4.5%) reported they had close contact with people who showed mpx symptoms (Table 1).
Table 1. Distributions of background characteristics of YMSM\(^a\) from 6 provinces of China in September 2022 (N=2493).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>24.6 (2.9)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Han</td>
<td>2392 (95.9)</td>
</tr>
<tr>
<td>Others</td>
<td>101 (4.1)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>611 (24.5)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>127 (5.1)</td>
</tr>
<tr>
<td>Employed</td>
<td>1755 (70.4)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>High school or below</td>
<td>407 (14.3)</td>
</tr>
<tr>
<td>College and above</td>
<td>2086 (83.7)</td>
</tr>
<tr>
<td><strong>Monthly income (CNY(^b)), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>≤2000 (US $279)</td>
<td>487 (19.5)</td>
</tr>
<tr>
<td>2001-6000 (US $279-$836)</td>
<td>1043 (41.8)</td>
</tr>
<tr>
<td>6001-10,000 (US $836-$1393)</td>
<td>652 (26.2)</td>
</tr>
<tr>
<td>&gt;10,000 (US $1393)</td>
<td>311 (12.5)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Married/living with a partner</td>
<td>113 (4.5)</td>
</tr>
<tr>
<td>Unmarried/divorced/separated/widowed</td>
<td>2380 (95.5)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Homosexual</td>
<td>2023 (81.1)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>382 (15.3)</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>8 (0.3)</td>
</tr>
<tr>
<td>Not sure/other</td>
<td>80 (3.2)</td>
</tr>
<tr>
<td><strong>Overseas travel history, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>51 (2.0)</td>
</tr>
<tr>
<td>No</td>
<td>2442 (98.0)</td>
</tr>
<tr>
<td><strong>Number of male sex partners, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>384 (15.5)</td>
</tr>
<tr>
<td>1</td>
<td>836 (33.9)</td>
</tr>
<tr>
<td>≥2</td>
<td>1249 (50.6)</td>
</tr>
<tr>
<td><strong>Frequency of in-person gatherings with MSM(^c), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1157 (46.4)</td>
</tr>
<tr>
<td>1-2 times</td>
<td>710 (28.5)</td>
</tr>
<tr>
<td>3-5 times</td>
<td>332 (13.3)</td>
</tr>
<tr>
<td>1-3 times per month</td>
<td>223 (8.9)</td>
</tr>
<tr>
<td>1-4 times per week</td>
<td>59 (2.4)</td>
</tr>
<tr>
<td>&gt;4 times per week</td>
<td>12 (0.5)</td>
</tr>
<tr>
<td><strong>Having chemsex with MSM, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>541 (21.7)</td>
</tr>
<tr>
<td>No</td>
<td>1952 (78.3)</td>
</tr>
</tbody>
</table>
### Characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HIV infection, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>279 (11.2)</td>
</tr>
<tr>
<td>Negative/unknown/not clear</td>
<td>2104 (84.4)</td>
</tr>
<tr>
<td>Never tested</td>
<td>110 (4.4)</td>
</tr>
<tr>
<td><strong>History of STIs(^d), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>354 (14.2)</td>
</tr>
<tr>
<td>No</td>
<td>2139 (85.8)</td>
</tr>
<tr>
<td>Having any mpox(^e) symptoms, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>570 (22.9)</td>
</tr>
<tr>
<td>No</td>
<td>1923 (77.1)</td>
</tr>
<tr>
<td>Having close contact with people who showed mpox symptoms, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>113 (4.5)</td>
</tr>
<tr>
<td>No/not sure</td>
<td>2380 (95.5)</td>
</tr>
</tbody>
</table>

\(^a\)YMSM: young men who have sex with men.  
\(^b\)CNY: Chinese yuan. An exchange rate of CNY 1=US $0.14 has been applied.  
\(^c\)MSM: men who have sex with men.  
\(^d\)STI: sexually transmitted infection.  
\(^e\)mpox: monkeypox.

### Mpox Knowledge and Cognition

In total, 2310 (92.7%) participants had heard of mpox, and the proportions of choosing the correct answers for the 10 knowledge questions among the whole sample ranged from 13.4% (n=334) to 81.7% (n=2037), with a mean knowledge score of 5.7 (SD 2.7). Regarding perceived susceptibility (mean 6.1, SD 2.3), 91 (3.7%), 157 (6.3%), and 425 (17%) participants reported a high likelihood (likely or very likely) of being infected, of being in close contact with patients with mpox, and of an mpox outbreak in China, respectively. For perceived severity, 2035 (81.6%) and 2073 (83.2%) participants perceived a strong/very strong severity of mpox regarding their health and life, respectively (mean 8.4, SD 1.9). In addition, 605 (24.3%) participants reported being emotionally distressed due to mpox (mean 2.7, SD 1.2). See Table 2 and Figure 2.
Table 2. Distribution of mpox\(^a\)-related knowledge and cognition of YMSM\(^b\) from 6 provinces of China in September 2022 (N=2493).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heard of mpox (frequency of choosing yes), n (%)</td>
<td>2310 (92.7)</td>
</tr>
<tr>
<td>Mpox knowledge (frequency of choosing the right answer), n (%)</td>
<td></td>
</tr>
<tr>
<td>Mpox virus can be transmitted to humans through close contact with an infected person.</td>
<td>2037 (81.7)</td>
</tr>
<tr>
<td>Mpox virus can be transmitted to humans through close contact with infected animals.</td>
<td>1860 (74.6)</td>
</tr>
<tr>
<td>Mpox virus can be transmitted to humans through close contact with contaminated objects.</td>
<td>1607 (64.5)</td>
</tr>
<tr>
<td>There are effective drugs against mpox virus in the world now.</td>
<td>1235 (49.5)</td>
</tr>
<tr>
<td>There are vaccines against mpox in China.</td>
<td>1241 (49.8)</td>
</tr>
<tr>
<td>After being infected with mpox virus, symptoms such as fever, rash, and swollen lymph nodes may occur.</td>
<td>1914 (76.8)</td>
</tr>
<tr>
<td>After a patient with mpox is cured, they may have scars for life once the skin rash goes away.</td>
<td>1535 (61.6)</td>
</tr>
<tr>
<td>Most patients with mpox can recover by themselves.</td>
<td>334 (13.4)</td>
</tr>
<tr>
<td>Symptoms are usually observed immediately once someone is infected with mpox.</td>
<td>1310 (52.5)</td>
</tr>
<tr>
<td>The smallpox vaccination can reduce the risk of mpox infection.</td>
<td>1141 (45.8)</td>
</tr>
<tr>
<td>Total mpox knowledge score ranging from 0 to 10, mean (SD)</td>
<td>5.7 (2.7)</td>
</tr>
<tr>
<td>Perceived susceptibility (frequency of choosing likely and very likely), n (%)</td>
<td></td>
</tr>
<tr>
<td>Likelihood of being infected with mpox</td>
<td>91 (3.7)</td>
</tr>
<tr>
<td>Likelihood of having close contact with patients with mpox</td>
<td>157 (6.3)</td>
</tr>
<tr>
<td>Likelihood of an mpox outbreak in China</td>
<td>425 (17.0)</td>
</tr>
<tr>
<td>Total perceived susceptibility score ranging from 3 to 15, mean (SD)</td>
<td>6.1 (2.3)</td>
</tr>
<tr>
<td>Perceived severity (frequency of choosing strong and very strong), n (%)</td>
<td></td>
</tr>
<tr>
<td>Perceived negative impact of mpox infection on your health</td>
<td>2035 (81.6)</td>
</tr>
<tr>
<td>Perceived negative impact of mpox infection on your life</td>
<td>2073 (83.2)</td>
</tr>
<tr>
<td>Total perceived severity score ranging from 2 to 10, mean (SD)</td>
<td>8.4 (1.9)</td>
</tr>
<tr>
<td>Emotional distress caused by mpox (frequency of choosing strong and very strong), n (%)</td>
<td></td>
</tr>
<tr>
<td>Emotional distress caused by the mpox epidemic (eg, panic, anxiety)</td>
<td>605 (24.3)</td>
</tr>
<tr>
<td>Total emotional distress score ranging from 1 to 5, mean (SD)</td>
<td>2.7 (1.2)</td>
</tr>
</tbody>
</table>

\(^a\)mpox: monkeypox.

\(^b\)YMSM: young men who have sex with men.
Figure 2. Level of cognition of mpox, mpox vaccination, and mpox testing among YMSM from 6 provinces in China in September 2022 (N=2493). mpox: monkeypox; YMSM: young men who have sex with men.

Cognition of Mpxo Vaccination and Testing
For the perceived benefits of mpox vaccination (mean 16.6, SD 2.9), the participants agreed/strongly agreed with the following 4 benefits presented: protecting oneself (n=1948, 78.1%), protecting others (n=1997, 80.1%), protecting the country (n=2187, 87.7%), and helping them live normally (n=1954, 78.4%). For the perceived barriers to mpox vaccination (mean 13.8, SD 3.5), the participants agreed/strongly agreed that the following 4 barriers are present: high cost of vaccination (n=1316, 52.8%), identity leakage (n=1002, 40.2%), inconvenience (n=1230, 49.3%), and side effects (n=1598, 64.1%). For self-efficacy (mean 7.5, SD 1.5), 1717 (68.9%) participants were confident in taking mpox vaccination and 1490 (59.8%) thought it would be easy to do so (Table 3 and Figure 2B).
Table 3. Behavioral intention and cognition related to mpox\(^a\) vaccination among YMSM\(^b\) from 6 provinces of China in September 2022 (N=2493).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intention of receiving mpox vaccination by scenario (frequency of choosing likely and very likely), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>When there was no local case reported</td>
<td>1651 (66.2)</td>
</tr>
<tr>
<td>When there was no local case reported and the vaccine was free</td>
<td>1971 (79.1)</td>
</tr>
<tr>
<td>When there were local cases reported</td>
<td>2112 (84.7)</td>
</tr>
<tr>
<td>When there were local cases reported and the vaccine was free</td>
<td>2204 (88.4)</td>
</tr>
<tr>
<td>Total intention score ranging from 4 to 20, mean (SD)</td>
<td>16.8 (3.6)</td>
</tr>
<tr>
<td><strong>Perceived benefits of mpox vaccination (frequency of choosing agree and strongly agree), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>It can protect me from mpox infection.</td>
<td>1948 (78.1)</td>
</tr>
<tr>
<td>It can protect those around me from mpox infection.</td>
<td>1997 (80.1)</td>
</tr>
<tr>
<td>It can assist the country in preventing and controlling the mpox epidemic.</td>
<td>2187 (87.7)</td>
</tr>
<tr>
<td>It can make me to work, study, and live normally.</td>
<td>1954 (78.4)</td>
</tr>
<tr>
<td>Total perceived benefits score ranging from 4 to 20, mean (SD)</td>
<td>16.6 (2.9)</td>
</tr>
<tr>
<td><strong>Perceived barriers to mpox vaccination (frequency of choosing agree and strongly agree), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>I am worried that the vaccination fee might be high.</td>
<td>1316 (52.8)</td>
</tr>
<tr>
<td>I am worried that my MSM(^c) identity might be revealed due to vaccination.</td>
<td>1002 (40.2)</td>
</tr>
<tr>
<td>I am worried that receiving vaccination might be inconvenient.</td>
<td>1230 (49.3)</td>
</tr>
<tr>
<td>I am worried that the vaccine might have side effects.</td>
<td>1598 (64.1)</td>
</tr>
<tr>
<td>Total perceived barriers score ranging from 4 to 20, mean (SD)</td>
<td>13.8 (3.5)</td>
</tr>
<tr>
<td><strong>Self-efficacy of mpox vaccination (frequency of choosing agree and strongly agree), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>I am confident in taking mpox vaccination.</td>
<td>1717 (68.9)</td>
</tr>
<tr>
<td>I think it is easy for me to take mpox vaccination if I want.</td>
<td>1490 (59.8)</td>
</tr>
<tr>
<td>Total self-efficacy score ranging from 2 to 10, mean (SD)</td>
<td>7.5 (1.5)</td>
</tr>
</tbody>
</table>

\(^a\)mpox: monkeypox.
\(^b\)YMSM: young men who have sex with men.
\(^c\)MSM: men who have sex with men.

For the perceived benefits of mpox testing (mean 13.4, SD 1.9), the participants agreed/strongly agreed with the following 3 benefits presented: timely finding out about the infection (n=2300, 92.3%), protecting others (n=2249, 90.2%), and protecting the country (n=2293, 92%). Regarding perceived barriers (mean 17.8, SD 4.4), the participants agreed/strongly agreed that the following 4 barriers are present: high cost (n=1495, 60%), inconvenience (n=1324, 53.1%), positive testing results (n=1370, 55%), identity leakage (n=1091, 43.8%), and leakage of testing results (n=1543, 61.9%). For self-efficacy (mean 7.9, SD 1.6), 1863 (74.7%) participants were confident in obtaining a mpox test and 1627 (65.3%) thought it would be easy to do so (Table 4 and Figure 2C).
Table 4. Behavioral intention and cognition related to mpox\textsuperscript{a} testing among YMSM\textsuperscript{b} from 6 provinces of China in September 2022 (N=2493).  

<table>
<thead>
<tr>
<th>Variables</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intention of getting mpox testing by scenario (frequency of choosing likely and very likely), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>When I had mpox symptoms</td>
<td>2302 (92.3)</td>
</tr>
<tr>
<td>When I had mpox symptoms and testing was free</td>
<td>2324 (93.2)</td>
</tr>
<tr>
<td>When I had close contact with patients with mpox</td>
<td>2320 (93.1)</td>
</tr>
<tr>
<td>When I had close contact with patients with mpox and testing was free</td>
<td>2342 (93.9)</td>
</tr>
<tr>
<td>Total intention score ranging from 4 to 20, mean (SD)</td>
<td>18.5 (2.5)</td>
</tr>
<tr>
<td><strong>Perceived benefits of mpox testing (frequency of choosing agree and strongly agree), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>It can help me know in time whether I have an mpox infection.</td>
<td>2300 (92.3)</td>
</tr>
<tr>
<td>It can protect people around me from mpox infection.</td>
<td>2249 (90.2)</td>
</tr>
<tr>
<td>It can assist the country in preventing and managing mpox epidemics.</td>
<td>2293 (92.0)</td>
</tr>
<tr>
<td>Total perceived benefits score ranging from 3 to 15, mean (SD)</td>
<td>13.4 (1.9)</td>
</tr>
<tr>
<td><strong>Perceived barriers to mpox testing (frequency of choosing agree and strongly agree), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>I am worried that the testing fee might be high.</td>
<td>1495 (60.0)</td>
</tr>
<tr>
<td>I am worried that my MSM\textsuperscript{c} identity might be revealed due to testing.</td>
<td>1091 (43.8)</td>
</tr>
<tr>
<td>I am worried that undergoing testing might be inconvenient.</td>
<td>1324 (53.1)</td>
</tr>
<tr>
<td>I am worried that my testing results might be positive.</td>
<td>1370 (55.0)</td>
</tr>
<tr>
<td>I am worried that my testing results might be revealed to others (eg, family members, colleagues, classmates).</td>
<td>1543 (61.9)</td>
</tr>
<tr>
<td>Total perceived barriers score ranging from 5 to 25, mean (SD)</td>
<td>17.8 (4.4)</td>
</tr>
<tr>
<td><strong>Self-efficacy of mpox testing (frequency of choosing agree and strongly agree), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>I am confident in taking an mpox test.</td>
<td>1863 (74.7)</td>
</tr>
<tr>
<td>I think it is easy for me to take an mpox test if I want.</td>
<td>1627 (65.3)</td>
</tr>
<tr>
<td>Total self-efficacy score ranging from 2 to 10, mean (SD)</td>
<td>7.9 (1.6)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}mpox: monkeypox.  
\textsuperscript{b}YMSM: young men who have sex with men.  
\textsuperscript{c}MSM: men who have sex with men.

**Behavioral Intention of Receiving Mpox Vaccination and Undergoing Mpox Testing**

The prevalence of having a behavioral intention of receiving mpox vaccination varied in different scenarios (n=1651, 66.2\%, when there was no local case reported; n=1971, 79.1\%, when there was no local case reported and the vaccine was free; n=2112, 84.7\%, when there were local cases reported; and n=2204, 88.4\%, when there were local cases reported and the vaccine was free; mean 16.8, SD 3.6). See Table 3 and Figure 3A.
Figure 3. Behavioral intention of receiving mpox vaccination and testing in different scenarios among YMSM from 6 provinces of China in September 2022 (N=2493). mpox: monkeypox; YMSM: young men who have sex with men.

The prevalence of having a behavioral intention of undergoing mpox testing was above 90% in different scenarios (n=2302, 92.3%, when they had mpox symptoms; n=2324, 93.2%, when they had mpox symptoms and testing was free; n=2320, 93.1%, when they had close contact with patients with mpox; and n=2342, 93.9%, when they had close contact with patients with mpox and testing was free; mean 18.5, SD 2.5). See Table 4 and Figure 3B.

Stratified Heterogeneity of Behavioral Intention of Receiving Mpox Vaccination and Undergoing Mpox Testing

As presented in Table 5, no heterogeneity was observed in the behavioral intention of receiving mpox vaccination and undergoing mpox testing across study sites and socioeconomic characteristics. However, heterogeneity was observed in the behavioral intention of receiving mpox vaccination in a number of male sexual partners (q=0.005, P=.04). In addition, heterogeneity was observed in the behavioral intention of undergoing mpox testing in in-person gatherings with MSM (q=0.002, P<.001).
<table>
<thead>
<tr>
<th>Strata</th>
<th>Behavioral intention of receiving mpox vaccination (score)</th>
<th>Behavioral intention of undergoing mpox testing (score)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>q-statistic (P value)</td>
</tr>
<tr>
<td>Study sites&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Province A</td>
<td>16.79 (3.55)</td>
<td>0.002 (.39)</td>
</tr>
<tr>
<td>Province B</td>
<td>16.93 (3.16)</td>
<td>—</td>
</tr>
<tr>
<td>Province C</td>
<td>16.88 (3.50)</td>
<td>—</td>
</tr>
<tr>
<td>Province D</td>
<td>16.85 (3.70)</td>
<td>—</td>
</tr>
<tr>
<td>Province E</td>
<td>16.43 (3.70)</td>
<td>—</td>
</tr>
<tr>
<td>Province F</td>
<td>16.68 (3.89)</td>
<td>—</td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>16.85 (3.55)</td>
<td>—</td>
</tr>
<tr>
<td>25-29</td>
<td>16.68 (3.62)</td>
<td>—</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Han</td>
<td>16.74 (3.59)</td>
<td>—</td>
</tr>
<tr>
<td>Others</td>
<td>17.19 (3.49)</td>
<td>—</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student/unemployed</td>
<td>16.75 (3.67)</td>
<td>—</td>
</tr>
<tr>
<td>Employed</td>
<td>16.77 (3.55)</td>
<td>—</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or below</td>
<td>16.66 (3.71)</td>
<td>—</td>
</tr>
<tr>
<td>College and above</td>
<td>16.78 (3.57)</td>
<td>—</td>
</tr>
<tr>
<td>Monthly income (CNY&lt;sup&gt;f&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6000 (≤US $836)</td>
<td>16.80 (3.56)</td>
<td>—</td>
</tr>
<tr>
<td>&gt;6000 (&gt;US $836)</td>
<td>16.71 (3.64)</td>
<td>—</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/living with a partner</td>
<td>16.63 (3.71)</td>
<td>—</td>
</tr>
<tr>
<td>Unmarried/divorced/separated/widowed</td>
<td>16.77 (3.58)</td>
<td>—</td>
</tr>
<tr>
<td>Overseas travel history</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>16.84 (3.91)</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>16.76 (3.58)</td>
<td>—</td>
</tr>
<tr>
<td>Number of male sexual partners</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>16.16 (3.85)</td>
<td>—</td>
</tr>
<tr>
<td>1</td>
<td>16.89 (3.57)</td>
<td>—</td>
</tr>
<tr>
<td>≥2</td>
<td>16.88 (3.49)</td>
<td>—</td>
</tr>
<tr>
<td>In-person gatherings with MSM&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>16.63 (3.79)</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>16.88 (3.41)</td>
<td>—</td>
</tr>
<tr>
<td>HIV infection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>17.37 (3.39)</td>
<td>—</td>
</tr>
<tr>
<td>Negative/unknown/not clear/never tested</td>
<td>16.69 (3.61)</td>
<td>—</td>
</tr>
</tbody>
</table>
Factors Associated with the Behavioral Intention of Receiving Mpox Vaccination and Undergoing Mpox Testing

Table 6 presents the results of univariate and multivariate regression analyses of the behavioral intention of receiving mpox vaccination. The multivariate regression model showed that vaccination intention was significantly and positively associated with having 1 male sex partner in the past 6 months \( (b_a=0.452, 95\% \ CI \ 0.098-0.806) \), mpox knowledge \( (b_a=0.060, 95\% \ CI \ 0.016-0.103) \), perceived susceptibility \( (b_a=0.091, 95\% \ CI \ 0.035-0.146) \), perceived severity \( (b_a=0.230, 95\% \ CI \ 0.164-0.296) \), emotional distress \( (b_a=0.270, 95\% \ CI \ 0.160-0.380) \), perceived benefits \( (b_a=0.455, 95\% \ CI \ 0.411-0.498) \), and self-efficacy \( (b_a=0.586, 95\% \ CI \ 0.504-0.668) \). Perceived barriers were negatively associated with the vaccination intention \( (b_a=-0.056, 95\% \ CI \ -0.090 \ to \ -0.022) \).
### Table 6. Univariate and multivariate regression analyses of the behavioral intention of receiving mpox\(^a\) vaccination among YMSM\(^b\) from 6 provinces of China in September 2022 (N=2493).

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Univariate regression (b) (95% CI)</th>
<th>Multivariate regression (b) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>(-0.032 (-0.081 to 0.018))</td>
<td>(-0.040 (-0.090 to 0.010))</td>
</tr>
<tr>
<td>Ethnicity (others vs Han)</td>
<td>(0.443 (-0.272 to 1.158))</td>
<td>(0.107 (-0.476 to 0.691))</td>
</tr>
<tr>
<td>Employment status (employed vs student/unemployed)</td>
<td>(0.019 (-0.289 to 0.328))</td>
<td>(0.028 (-0.295 to 0.352))</td>
</tr>
<tr>
<td>Education level (college and above vs high school or below)</td>
<td>(0.119 (-0.262 to 0.500))</td>
<td>(0.015 (-0.311 to 0.341))</td>
</tr>
<tr>
<td>Monthly income (&gt;CNY(^c) 6000 vs ≤CNY 6000 [US $836])</td>
<td>(-0.089 (-0.379 to 0.200))</td>
<td>(-0.167 (-0.444 to 0.111))</td>
</tr>
<tr>
<td>Marital status (unmarried/divorced/separated/widowed vs married/living with a partner)</td>
<td>(0.141 (-0.537 to 0.819))</td>
<td>(-0.123 (-0.684 to 0.438))</td>
</tr>
<tr>
<td>Overseas travel history (yes vs no)</td>
<td>(0.082 (-0.914 to 1.078))</td>
<td>(0.216 (-0.593 to 1.024))</td>
</tr>
</tbody>
</table>

#### Number of male sex partners in the past 6 months (reference: 0)

- **1**
  - \(0.730 (0.298 to 1.162)^d\)
  - \(0.452 (0.098 to 0.806)^f\)
- **≥2**
  - \(0.726 (0.317 to 1.135)^d\)
  - \(0.200 (-0.149 to 0.550)\)

- Having in-person gatherings in the past 6 months (yes vs no)
  - \(0.243 (-0.039 to 0.526)\)
  - \(0.058 (-0.179 to 0.295)\)

- Having chemsex with male sex partners in the past 6 months (yes vs no)
  - \(0.409 (0.067 to 0.751)^e\)
  - \(0.235 (-0.059 to 0.529)\)

- HIV infection (yes vs other)
  - \(0.683 (0.236 to 1.129)^d\)
  - \(0.321 (-0.066 to 0.708)\)

- History of STIs\(^f\) (yes vs other)
  - \(0.434 (0.031 to 0.838)^e\)
  - \(0.111 (-0.239 to 0.462)\)

- Having any mpox symptoms in the past 2 weeks (yes vs no)
  - \(0.039 (-0.297 to 0.375)\)
  - \(-0.008 (-0.295 to 0.280)\)

- Having close contact with people showing mpox symptoms in the past 2 weeks (yes vs other)
  - \(-0.067 (-0.744 to 0.611)\)
  - \(-0.191 (-0.762 to 0.381)\)

- Mpox knowledge
  - \(0.156 (0.105 to 0.208)^f\)
  - \(0.060 (0.016 to 0.103)^d\)

- Perceived susceptibility of mpox
  - \(0.183 (0.122 to 0.244)^f\)
  - \(0.091 (0.035 to 0.146)^d\)

- Perceived severity of mpox
  - \(0.491 (0.417 to 0.565)^f\)
  - \(0.230 (0.164 to 0.296)^f\)

- Emotional distress caused by mpox
  - \(0.571 (0.455 to 0.686)^f\)
  - \(0.270 (0.160 to 0.380)^f\)

- Perceived benefits of mpox vaccination
  - \(0.622 (0.580 to 0.664)^f\)
  - \(0.455 (0.411 to 0.498)^f\)

- Perceived barriers to mpox vaccination
  - \(0.051 (0.010 to 0.092)^e\)
  - \(-0.056 (-0.090 to -0.022)^d\)

- Self-efficacy of receiving mpox vaccination
  - \(1.002 (0.919 to 1.085)^f\)
  - \(0.586 (0.504 to 0.668)^f\)

---

\(a\)mpox: monkeypox.

\(b\)YMSM: young men who have sex with men.

\(c\)CNY: Chinese yuan. An exchange rate of CNY 1=US $0.14 has been applied.

\(d\)\(P\)<.01.

\(e\)\(P\)<.05.

\(f\)STI: sexually transmitted infection.

\(g\)\(P\)<.001.

Table 7 presents the results of univariate and multivariate regression analyses of the behavioral intention of undergoing mpox testing. The testing intention was significantly and positively associated with having 1 male sex partner in the past 6 months (\(b_a=0.290, 95\% \text{ CI} 0.070-0.510\)), having in-person gatherings with MSM (\(b_a=0.219, 95\% \text{ CI} 0.072-0.366\)), perceived severity (\(b_a=0.283, 95\% \text{ CI} 0.241-0.325\)), perceived benefits (\(b_a=0.679, 95\% \text{ CI} 0.636-0.721\)), and self-efficacy (\(b_a=0.195, 95\% \text{ CI} 0.146, 0.245\)). Emotional distress was negatively associated with testing intention (\(b_a=-0.069, 95\% \text{ CI} -0.137 \text{ to } -0.001\)).
Table 7. Univariate and multivariate regression analyses of the behavioral intention of undergoing mpox testing among YMSM from 6 provinces of China in September 2022 (N=2493).

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Univariate regression $b_a$ (95% CI)</th>
<th>Multivariate regression $b_a$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>–0.004 (–0.035 to 0.027)</td>
<td>0.012 (–0.022 to 0.046)</td>
</tr>
<tr>
<td>Ethnicity (others vs Han)</td>
<td>0.093 (–0.269 to 0.454)</td>
<td>0.158 (–0.056 to 0.372)</td>
</tr>
<tr>
<td>Employment status (employed vs student/unemployed)</td>
<td>0.141 (–0.059 to 0.341)</td>
<td>0.103 (–0.161 to 0.368)</td>
</tr>
<tr>
<td>Education level (college and above vs high school or below)</td>
<td>0.141 (–0.059 to 0.341)</td>
<td>0.103 (–0.099 to 0.303)</td>
</tr>
<tr>
<td>Monthly income (≤CNY 6000 vs &gt;CNY 6000 [US $836])</td>
<td>0.102 (–0.311 to 0.484)</td>
<td>0.042 (–0.428 to 0.513)</td>
</tr>
<tr>
<td>Marital status (unmarried/divorced/separated/widowed vs married/living with a partner)</td>
<td>0.136 (–0.021 to 0.484)</td>
<td>0.219 (–0.472 to 0.909)</td>
</tr>
<tr>
<td>Overseas travel history (yes vs no)</td>
<td>0.219 (0.070 to 1.000)</td>
<td>0.061 (–0.028 to 0.026)</td>
</tr>
<tr>
<td>Number of male sex partners in the past 6 months (reference: 0)</td>
<td>0.998</td>
<td>0.219 (0.072 to 0.366)</td>
</tr>
<tr>
<td>≥2</td>
<td>0.247 (0.051 to 0.443)</td>
<td>0.208 (–0.029 to 0.445)</td>
</tr>
<tr>
<td>Having in-person gatherings in the past 6 months (yes vs no)</td>
<td>0.14 (–0.038 to 0.317)</td>
<td>0.06 (–0.021 to 0.025)</td>
</tr>
<tr>
<td>Having chemsex with male sex partners in the past 6 months (yes vs no)</td>
<td>0.256 (0.024 to 0.489)</td>
<td>0.208 (–0.029 to 0.445)</td>
</tr>
<tr>
<td>HIV infection (yes vs other)</td>
<td>0.479 (0.170 to 0.789)</td>
<td>0.290 (0.070 to 0.510)</td>
</tr>
<tr>
<td>History of STIs (yes vs other)</td>
<td>0.035 (–0.018 to 0.252)</td>
<td>0.240 (–0.012 to 0.486)</td>
</tr>
<tr>
<td>Having any mpox symptoms in the past 2 weeks (yes vs no)</td>
<td>0.14 (–0.038 to 0.317)</td>
<td>0.256 (0.024 to 0.489)</td>
</tr>
<tr>
<td>Having close contact with people showing mpox symptoms in the past 2 weeks (yes vs other)</td>
<td>0.140 (–0.099 to 0.380)</td>
<td>0.479 (0.170 to 0.789)</td>
</tr>
<tr>
<td>Mpox knowledge</td>
<td>0.081 (0.045 to 0.117)</td>
<td>0.081 (0.045 to 0.117)</td>
</tr>
<tr>
<td>Perceived susceptibility of mpox</td>
<td>–0.053 (–0.096 to –0.011)</td>
<td>–0.053 (–0.096 to –0.011)</td>
</tr>
<tr>
<td>Perceived severity of mpox</td>
<td>0.511 (0.482 to 0.579)</td>
<td>0.511 (0.482 to 0.579)</td>
</tr>
<tr>
<td>Emotional distress caused by mpox</td>
<td>0.283 (0.241 to 0.325)</td>
<td>0.283 (0.241 to 0.325)</td>
</tr>
<tr>
<td>Perceived benefits of mpox vaccination</td>
<td>0.679 (0.636 to 0.721)</td>
<td>0.679 (0.636 to 0.721)</td>
</tr>
<tr>
<td>Perceived barriers to mpox vaccination</td>
<td>0.011 (–0.007 to 0.027)</td>
<td>0.011 (–0.007 to 0.027)</td>
</tr>
<tr>
<td>Self-efficacy of receiving mpox vaccination</td>
<td>0.195 (0.146 to 0.245)</td>
<td>0.195 (0.146 to 0.245)</td>
</tr>
</tbody>
</table>

a mpox: monkeypox.
b YMSM: young men who have sex with men.
c CNY: Chinese yuan. An exchange rate of CNY 1=US $0.14 has been applied.
d P<.001.
e P<.05.
f P<.01.
g STI: sexually transmitted infection.

Discussion

Principal Findings

This was among the earliest studies in China examining the levels and factors of the behavioral intention of receiving mpox vaccination and undergoing mpox testing among YMSM. After the COVID-19 travel restrictions were eased in China at the end of 2022, this specific age population became the group with a high risk of mpox infection due to its social activeness. The study findings shed some light on the preparedness for mpox prevention and control in the country.

Comparison With Prior Work

Overall, about two-thirds of the surveyed YMSM intended to receive the mpox vaccination if there was no case present locally, and the prevalence increased to 84.7% if local cases were reported. The latter prevalence was higher than that in 2 other countries with an mpox outbreak. A recent study in the Netherlands showed that 70% of MSM (56.6% aged <45 years)
had an intention of getting vaccinated [27]. Another study in France found that 33.6% of MSM (80.6% between 30 and 59 years old) declared their hesitancy to get vaccinated against mpox [28]. Age may be a potential reason for the relatively high prevalence, as YMSM tend to be more open to newly emerging health services (eg, pre-exposure prophylaxis [PrEP], HPV vaccines) [29-31]. The finding implies that it is important to keep MSM informed about the most up-to-date mpox statistics. It may assist them in decision-making about adopting preventive behaviors. Furthermore, the study found the prevalence of vaccination intention increased to some extent if the vaccination was free. This indicates that cost may be a barrier to mpox vaccination. The study showed about half of the participants expressed a concern about the potential high cost of vaccination, which is consistent with previous research [32]. Therefore, free or subsidized immunization for the high-risk population is recommended to be considered in the future national response.

In this study, over 90% of the participants reported an intention of undergoing an mpox test regardless of the presence of symptoms and the cost. A possible explanation for the high prevalence is that the participants may be knowledgeable about the benefits of testing. As shown in the study, most participants agreed/strongly agreed with the stated benefits of testing. The mass HIV testing campaigns in the past decades may have contributed to these perceptions. Our data showed that only 4.4% of the participants had never been tested for HIV before. Furthermore, the COVID-19 battle since 2020 may have improved people's health literacy, which may, in turn, have positively affected their attitudes and practices of other health behaviors [33-35]. Lessons learned from the previous HIV testing campaigns targeting MSM and the COVID-19 testing campaign can be drawn by the future promotion of mpox testing in China [3,35]. For instance, it is imperative to remove structural barriers to testing (eg, provide anonymous tests) and offer culturally congruent services. Additionally, mpox testing is recommended to be provided in community settings, CBOs, and sexual health clinics trusted by MSM populations.

Guided by the classic health behavioral theories, we identified several factors associated with the behavioral intention of receiving mpox vaccination and undergoing mpox testing. First, according to the KAP model, knowledge is the foremost factor affecting the practice of preventive behaviors [17]. Previous studies have demonstrated that disease knowledge is positively associated with HIV testing and COVID-19–preventive behaviors [36,37]. Multivariate regression indicated a moderate association between mpox knowledge and vaccination intention and a statistically insignificant association between mpox knowledge and testing intention. The findings suggest that disease knowledge may play a role in motivating the intention of adopting preventive behaviors, but the strength and magnitude of the effect may be limited. Increasing mpox knowledge is important but far from enough in the future response to mpox in China.

The study found that 3 HBM constructs (perceived severity of mpox, perceived benefits of mpox vaccination/testing, and self-efficacy of mpox vaccination/testing) are independently and positively associated with vaccination and testing intentions. The findings are consistent with previous evidence on other diseases, such as COVID-19 and H1N1 [38,39]. Meanwhile, 2 other HBM constructs, perceived susceptibility of mpox and perceived barriers to the corresponding service, were also significantly associated with the vaccination intention but not with the testing intention. A possible reason is that the level of testing intention was particularly high in the study and the influence of the 2 factors were limited. Overall, these findings evidence that the HBM is a useful framework to predict health behaviors. Some related components are recommended for future mpox interventions, such as timely updating and disseminating mpox statistics, implementing targeted risk communication, educating the community about the benefits and ways of access to testing and vaccination, ensuring the privacy and confidentiality of services, and debunking and refuting misinformation promptly [40,41].

An interesting finding of this study was that the emotional distress caused by mpox is positively associated with the vaccination intention but negatively associated with the testing intention. It is understandable that YMSM who have experienced more emotional distress due to mpox might be more worried about mpox infection and thus intend to receive immunization. This is in line with previous research showing that people with more emotional distress caused by COVID-19 were more likely to receive COVID-19 vaccination [42]. In contrast, it is speculated that some YMSM were emotionally distressed because they were afraid that the testing result would be positive or the result would be revealed to others. Such fears consequently hindered their willingness to undergo testing. Our data showed that 55% of the participants were worried that their testing results might be positive, while 61.9% were worried that the testing results might be leaked to others. A previous HIV study also indicated that emotional distress is a major barrier to seeking HIV-testing services [43]. The opposite results remind us that it is necessary to monitor the mental status in the affected community during an emergency and provide timely advice on positive coping strategies.

Stratified heterogeneity analysis revealed variations in the behavioral intention among people with or without mpox risk behaviors. Certain mpox risk behaviors presented significant associations with the intention of receiving mpox vaccination and undergoing mpox testing. Those who had 1 male sex partner in the past 6 months were more intended to uptake both vaccination and testing services compared to those with none or multiple partners. It is understandable that MSM without sex partners recently might think they have no risk of infection and thus have no need to receive the vaccination or undergo testing, while MSM with multiple partners might possess a lower level of health literacy and risk perceptions [44]. The finding warns us that MSM with multiple sexual partners may have a lower intention of adopting preventive behaviors against mpox and should be given a priority focus in the future fight against mpox. In addition, attending in-person gatherings with MSM and traveling overseas in the past 6 months were both positively associated with the testing intention. The findings are expected because they are known risk factors of mpox infection [45,46]. It is worth noting that YMSM are at high risk of not only mpox but also HIV/AIDS. Evidence shows that mpox is an opportunistic infection of HIV/AIDS [47]. China has a
well-established comprehensive public health system for HIV/AIDS health promotion, testing, treatment, and management, which is considered highly effective for HIV/AIDS high-risk groups. Therefore, it is highly recommended to integrate mpox-related health services and prevention actions into the existing HIV/AIDS prevention and control system. For example, it may be feasible and cost-effective to provide mpox vaccination and testing services at the current HIV testing and counseling sites that are often operated by MSM-friendly CBOs and clinics. It is also important to empower MSM communities and highly involve them, especially the key persons in the communities, in targeted health communication, health counseling, and testing and treatment referrals. This integrated approach is key to addressing the interconnected health challenges faced by YMSM.

Limitations
The study has a few limitations. First, selection bias might exist due to the facility-based sampling and refusal to participate. YMSM who could be approached in this study might be more likely to interact positively with CBOs, receive health education information more frequently, and have a higher level of health literacy. Driven by self-selection, YMSM who were invited but refused to participate might tend to have more concerns about this sensitive topic and their privacy. These systematic difference biases potentially resulted in an overestimation of the behavioral intention related to mpox vaccination and testing. However, as MSM groups are hidden populations, the convenient sampling method is a common way for recruitment. In addition, data were collected from August to September 2022, when China was still under strict COVID-19 restrictions. The recruitment rate may have also been impacted by the pandemic. Second, due to self-reporting, social desirability bias might exist and lead to an overestimation of the behavioral intention. Third, the study was conducted in 6 provincial regions of China and focused on YMSM aged 18-29 years. The generalization of the study findings to broader populations should be made with caution. Fourth, the cross-sectional design allowed a limited ability of causal inferences. Longitudinal studies are warranted.

Conclusion
The majority of the YMSM in this study showed an intention of receiving mpox vaccination, and the intention was somewhat sensitive to the epidemic status and cost. Most of the YMSM had an intention of undergoing an mpox test regardless of scenarios. It is recommended that the future national response to mpox in China prioritize groups with high-risk exposure, raise their mpox knowledge, promote cognition of the disease and preventive measures, improve accessibility and privacy of health services, monitor the mental health status of the groups, and provide advice on positive coping strategies.

Acknowledgments
This study was supported by the Research Fund, Yanke School of Public Health, Tsinghua University (number: 100010001). We thank Mumu and Keke from the Home of the Red Ribbon in Beijing; Huzi from the Qingdao Qingtong AIDS Prevention Volunteer Service Center; Hui Yao from the Hefei Center for Disease Control and Prevention; Maomaoyu from the Hefei Qingwei Public Health Service Center; Xiaodong Wang and Haipeng Zha from the Chengdu Tongle Health Consultation Center; Han Shi, Chen Yang, and ToAlan from the Zhejiang Coastal Charity Volunteer Service Center; Huiyang Xu from the Guangdong Association of STD & AIDS Prevention and Control; Fang Yang from the Guangdong Provincial Center for Disease Control and Prevention; Jason and Shenghong Liang from the Shenzhen 158 Rainbow Volunteer Service Center; Xi He from the Zhuhai Xutong Volunteer Service Center; Jie Lu from the Guangzhou Lingnan Partnership Community Support Center; and all other field-workers for their contributions to participant recruitment and data collection. We also deeply appreciate the time and effort of all the participants.

Data Availability
All data generated and analyzed during this study are included in this published paper and its supplementary information files.

Authors’ Contributions
SL, WL, and FC conceptualized the study and contributed to the methodology. SL supervised the study. KJ, YZ, YX, JZ, SH, JS, QL, YL, YX, WM, LH, XR, and ZD contributed to the questionnaire design and data collection. KJ, YX, and SL conducted the formal analysis. SL, KJ, YX, JZ, and SH contributed to the original draft of the manuscript. SL finalized the manuscript. All authors were responsible for the decision to submit the manuscript for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Measures of the study.
[DOCX File, 18 KB - publichealth_v10i1e47165_app1.docx ]

References


47. O'Shea J, Daskalakis D, Brooks JT. The emergence of mpox as an HIV-related opportunistic infection. Lancet 2023 Apr 15;401(10384):1264 [FREE Full text] [doi: 10.1016/S0140-6736(23)00395-1] [Medline: 37061263]

Abbreviations

CBO: community-based organization
CNY: Chinese Yuan
GDP: gross domestic product
HBM: Health Belief Model
HPV: human papillomavirus
KAP: knowledge attitude practice
mpox: monkeypox
MSM: men who have sex with men
STD: sexually transmitted disease
STI: sexually transmitted infection
TPB: theory of planned behavior
WHO: World Health Organization
YMSM: young men who have sex with men
Sex- and Age-Specific Prevalence of Osteopenia and Osteoporosis: Sampling Survey

Yao Fan¹*, MS; Qun Li²*, BS; Yu Liu³*, MS; Jing Miao⁴, MS; Ting Zhao², MS; Jinxin Cai², AD; Min Liu³, MS; Jun Cao⁵, BS; Haifeng Xu³, MS; Lai Wei⁵, MS; Mengxia Li⁵, PhD; Chong Shen⁵, PhD

¹Division of Clinical Epidemiology, Affiliated Geriatric Hospital of Nanjing Medical University, Nanjing, China
²Department of Nutrition, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China
³Institute for the Prevention and Control of Chronic Non-communicable Diseases, Center for Disease Control and Prevention of Jurong City, Jurong, China
⁴Department of Clinical Nutrition, Affiliated Sir Run Run Hospital of Nanjing Medical University, Nanjing, China
⁵Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, China

* these authors contributed equally

Corresponding Author:
Chong Shen, PhD
Department of Epidemiology
School of Public Health, Nanjing Medical University
No. 101 Longmian Avenue, Jiangning District
Nanjing, 211166
China
Phone: 86 25 86868443
Fax: 86 25 86527613
Email: sc@njmu.edu.cn

Abstract

Background: Osteopenia and osteoporosis are posing a long-term influence on the aging population’s health contributing to a higher risk of mortality, loss of autonomy, hospitalization, and huge health system costs and social burden. Therefore, more pertinent data are needed to demonstrate the current state of osteoporosis.

Objective: This sampling survey seeks to assess the trends in the prevalence of osteopenia and osteoporosis in a Chinese Han population.

Methods: A community-based cross-sectional study involving 16,377 participants used a multistage sampling method. Bone mineral density was measured using the quantitative ultrasonic densitometry. Student t test and Mann-Whitney U test were used to test the difference between normally and nonnormally distributed quantitative variables between male and female participants. A chi-square (χ²) test was used to compare categorized variables. Stratified analysis was conducted to describe the prevalence rates of osteoporosis (T score ≤–2.5) and osteopenia (T score –2.5 to –1.0) across age, sex, calcium intake, and menopause. A direct standardization method was used to calculate the age-standardized prevalence rates of osteoporosis and osteopenia. T-score was further categorized into quartiles (T1-T4) by age- and sex-specified groups.

Results: The prevalence rates of osteopenia and osteoporosis were 40.5% (6633/16,377) and 7.93% (1299/16,377), respectively, and the age-standardized prevalence rates were 27.32% (287,877,129.4/1,053,861,940) and 3.51% (36,974,582.3/1,053,861,940), respectively. There was an increase in osteopenia and osteoporosis prevalence from 21.47% (120/559) to 56.23% (754/1341) and 0.89% (5/559) to 17.23% (231/1341), respectively, as age increased from 18 years to 75 years old. The prevalence rates of osteopenia and osteoporosis were significantly higher in female participants (4238/9645, 43.94% and 1130/9645, 11.72%) than in male participants (2395/6732, 35.58% and 169/6732, 2.51%; P<.001), and in postmenopausal female participants (3638/7493, 48.55% and 1053/7493, 14.05%) than in premenopausal female participants (538/2026, 26.55% and 53/2026, 2.62%; P<.001). In addition, female participants with a history of calcium intake had a lower osteoporosis prevalence rate than female participants without any history of calcium intake in all age groups (P=.004). From low quartile to high quartile of T-Score, the prevalence of diabetes mellitus (752/4037, 18.63%; 779/4029, 19.33%; 769/3894, 19.75%; and 869/3879, 22.4%) and dyslipidemia (2228/4036, 55.2%; 2304/4027, 57.21%; 2306/3891, 59.26%; and 2379/3878, 61.35%) were linearly increased (P<.001), while the prevalence of cancer (112/4037, 2.77%; 110/4029, 2.73%; 103/3894, 2.65%; and 77/3879, 1.99%) was decreased (P=.03).
Conclusions: Our data imply that as people age, osteopenia and osteoporosis are more common in females than in males, particularly in postmenopausal females than in premenopausal females, and bone mineral density significantly affects the prevalence of chronic diseases. These findings offer information that can be applied to intervention programs meant to prevent or lessen the burden of osteoporosis in China.

(JMIR Public Health Surveill 2024;10:e48947)  doi: 10.2196/48947

KEYWORDS

cross-sectional study; osteopenia; osteoporosis; prevalence; quantitative ultrasound

Introduction

Osteoporosis is a degenerative condition that affects the entire body’s bones. It is characterized by decreased bone mineral density (BMD) with low bone mass and deterioration of bone tissue microstructure, which makes bones more fragile and increases the risk of fracture [1]. Osteopenia refers to BMD levels that are below the normal range but not as low as osteoporosis [2]. The prevalence rates of osteopenia and osteoporosis are estimated to be 40.40% and 19.75% globally [3], respectively, and their burden has been increasing for nearly 3 decades. From 1990 to 2019, the disability-adjusted life years (DALYs) and global death of both osteopenia and osteoporosis increased by 93.82% and 111.16% globally [4], respectively. The DALY and death count of both osteopenia and osteoporosis-related fractures increased by 121.07% and 148.65%, respectively during the same period, and China is among the top 5 countries with the highest DALYs number in the osteopenia and osteoporosis-related fractures [4].

The risk factors of osteoporosis include family history, abnormal BMI, unhealthy lifestyles, being a woman, and aging [5]. It is estimated that by 2050, the proportion of the older adult population in China will represent 26.1% of the total Chinese population [6]. With the rapid increase in the proportion of older adults globally and in China, osteoporosis might pose a major threat to public health. Osteoporosis frequently induces fragile fractures of the spine, hip, distal forearm, and proximal humerus even with little strength stress force [7,8]. Generally, the clinical symptoms of osteoporosis appear after fracture occurrence and can result in significant disability and excruciating pain, which interferes with normal activities and frequently lowers the quality of life [9,10]. In addition, osteoporosis and its related fractures also increase the risk of mortality, loss of autonomy, nursing home referral, and hospitalization, leading to long-term disability, public health system costs, and a huge social burden [10,11].

It is estimated that by 2050, the annual osteoporosis-related fractures will increase to 5.99 million in China, with a projected cost of US $25.43 billion, representing a 2.7-fold increase compared to 2010 [12]. With the varying socioeconomic conditions, varied lifestyles, and expanding older population in a sizable developing nation like China, more pertinent data are needed to demonstrate the current state of osteoporosis. In addition, the prevalence of aging-related diseases such as type 2 diabetes mellitus (T2DM), coronary heart disease (CHD), and cardiovascular and cerebrovascular diseases (CCVD) is also increasing. These diseases not only often coexist with osteopenia and osteoporosis in older people but also share some common risk factors such as abnormal BMI, unhealthy lifestyles, and aging [5,13,14]. However, the relationship between osteoporosis and T2DM [15,16], CHD [17,18], and CCVD [19,20] remains controversial across studies. Therefore, to aid in the development of health care strategies for the expanding population, high-quality epidemiological studies are required to periodically estimate the prevalence of osteoporosis. Herein, we aimed to investigate the trends in the prevalence of osteoporosis and osteopenia and evaluate the impact of decreased BMD on common chronic diseases in a 2-stage sampling epidemiological survey.

Methods

Study Design and Participants

The baseline survey of the Jurong cohort study was carried out using a multistage sampling method in 16 towns in Jiangsu Province, in the south of China [21]. Eligible participants included were adult inhabitants (who lived locally for more than 6 months in the past 12 months and aged older than 18 years), with over 80% (13,682/17,102) coming from rural areas and having resided locally for more than 6 months in the past year. Individuals not included in this study were those having acute or severe medical conditions and those who were unwilling to take part in the study survey. The recruitment of the study population was done in 2 stages: the first stage included 11,150 participants from October to November 2015 and the second stage included 5952 participants from November to December 2018. Because this study aimed to observe the prevalence of chronic diseases including osteoporosis, stroke, and CHD, the 2-stage sampling was undertaken to have an efficient sample size for the need to estimate risk factors. The participants with missing information (n=6) and without measures of BMD (n=719) were excluded, and finally, 16,377 participants were included in the study (Figure 1).
Questionnaire

The questionnaire included demographic information, smoking and drinking habits, nutritional status, physical activity (PA), menopausal status, and history of chronic diseases (Multimedia Appendix 1). PA encompasses dynamic behavior and static behavior. Dynamic behaviors include PAs at work, transportation, and leisure time, while static behaviors include total static behavior and spare time static behavior. Using the International Physical Activity Questionnaire as a reference, the level of PA was quantified by metabolic equivalent (MET). The average daily activity duration was calculated by dividing the time of participating in dynamic behavior by 7 times the daily activity duration. Physical activity index (PAI), the ratio of an individual’s total energy consumption and basic metabolic energy consumption in one day, was calculated as follows:

\[
PAI = \frac{X \times Y}{7 \times 24}
\]

where \(X\) is the intensity of an activity and \(Y\) is the time for this activity (minutes).

BMI was calculated as the ratio of the weight (kg) to squared height (m\(^2\)). Participants with a BMI of 18.5-24, 24-28, and >28 kg/m\(^2\) were defined as normal weight, overweight, and obese, respectively [22]. The history of calcium and drug intake and the history of fracture or fall were self-reported by participants. To ensure the accuracy of medication information, participants who were taking drugs were required to bring the package of their medicines.

Menopause is defined as the absence of menstruation for 12 months after the last menstruation or self-reported removal of the uterus or ovaries surgically. The history of CHD, stroke, and cancer were also self-reported by participants. CCVD was defined as a history of CHD, stroke, or both. Hypertension was defined as an average systolic blood pressure/diastolic blood pressure level ≥140/90 mm Hg or self-reported by participants or those taking antihypertensive medicine. Diabetes mellitus (DM) was defined as an average fasting blood glucose ≥7.0 mmol/L, or as self-reported by participants, or those taking hypoglycemic agents or insulin injections. Dyslipidemia was defined as total cholesterol (TC) ≥5.2 mmol/L or low-density lipoprotein cholesterol ≥3.4 mmol/L or high-density lipoprotein cholesterol <1.0 mmol/L or triglyceride (TG) ≥1.7 mmol/L or self-reported participants about the history of dyslipidemia according to the Guidelines for the Prevention and Treatment of Dyslipidemia in Chinese Adults (Revised edition, 2016).

BMD Measurement

BMD measurement was performed using CM-200. The BMD values of the left side of the heel were measured while seated. Before measurement, the age and sex of participants were entered into the quantitative ultrasound system (QUS), and the instrument automatically converted them to a T-score. The BMD was grouped into 3 based on the updated World Health...
Organization–recommended criteria: T-score ≤ −2.5 was defined as osteoporosis, −2.5 to −1.0 was defined as osteopenia, and T-score > −1.0 was defined as normal BMD [23].

Biochemistry Measurements
After fasting over 8 hours, venous blood was drawn to measure the serum level of high-density lipoprotein cholesterol, TC, TG, and the plasma level of glucose. Low-density lipoprotein (μmol/L) = TC – (high-density lipoprotein + TG/2.2). Blood samples were tested within 4 hours using a Cobas 6000 analyzer (Roche Diagnostics GmbH).

Statistical Analysis
The normally distributed continuous variables were described by mean (SD). The nonnormally distributed variables were described by the median and IQR. Student t test and Mann-Whitney U test were used to test the difference of variables between male and female participants. A chi-square (χ²) test was used to compare categorized variables. Stratified analysis was conducted to describe the prevalence rates of osteoporosis and osteopenia across age, sex, and calcium intake. Based on the 2010 national census data for the Chinese population, we used a direct standardization method to calculate the age-standardized prevalence rate of osteoporosis and osteopenia. Logistic regression analyses were performed, and the association was estimated with an odds ratio (OR) and 95% confidence interval (CI). T-score was categorized into quartiles (T1-T4) by age- and sex-specified groups. A 2-tailed P value less than .05 was considered to indicate statistical significance.

All statistical analyses were performed using SPSS Statistics (version 24.0; IBM Corp).

Ethical Considerations
Nanjing Medical University’s Research Ethics Committee approved the study (2015077). The study was performed following the Declaration of Helsinki. The study’s scope, objectives, and advantages were thoroughly communicated to people who were invited to participate. All survey participants gave written informed consent. The collected information was kept confidential and used solely for research purposes.

Results

General Characteristics of Participants
Of the 16,377 participants included in this study, 6732 (41.1%) were male participants with an average age of 61.51 (10.84) years, and 9645 (58.9%) were female participants with an average age of 59.48 (SD 10.93) years. Participants with chronic diseases were identified, including 1287 (7.9%) with fracture or fall, 9558 (58.4%) with hypertension, 3113 (19%) with DM, 9523 (58.2%) with dyslipidemia, 267 (1.6%) with CHD, 890 (5.4%) with strokes, and 439 (2.7%) with cancer. In total, 8709 (53.4%) individuals self-reported a history of taking drugs in the past year, and among those, 766 (4.7%) people had taken calcium tablets. About 77.7% (7493/9645) of the female participants were postmenopausal, with 3.3% (247/7493) having their uterus or ovaries removed surgically (Table 1).
Table 1. Characteristics of the study participants.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total population (N=16,377)</th>
<th>Male participants (n=6732)</th>
<th>Female participants (n=9645)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>60.32 (10.94)</td>
<td>61.51 (10.84)</td>
<td>59.48 (10.93)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Illiteracy or primary</td>
<td>9819 (60)</td>
<td>3148 (46.8)</td>
<td>6671 (69.3)</td>
<td></td>
</tr>
<tr>
<td>Junior high school</td>
<td>4838 (29.6)</td>
<td>2514 (37.4)</td>
<td>2324 (24.1)</td>
<td></td>
</tr>
<tr>
<td>Senior high school or above</td>
<td>1699 (10.4)</td>
<td>1062 (15.8)</td>
<td>637 (6.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Unmarried</td>
<td>246 (1.5)</td>
<td>205 (3)</td>
<td>41 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>14,134 (86.3)</td>
<td>5972 (88.7)</td>
<td>8162 (84.6)</td>
<td></td>
</tr>
<tr>
<td>Divorced or widowed</td>
<td>1997 (12.2)</td>
<td>555 (8.2)</td>
<td>1442 (15)</td>
<td></td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Manual worker</td>
<td>11,983 (73.2)</td>
<td>5470 (83.3)</td>
<td>6513 (68.1)</td>
<td></td>
</tr>
<tr>
<td>Technician</td>
<td>323 (2)</td>
<td>202 (3.1)</td>
<td>121 (1.3)</td>
<td></td>
</tr>
<tr>
<td>Businessman</td>
<td>551 (3.4)</td>
<td>346 (5.3)</td>
<td>205 (2.1)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3272 (20)</td>
<td>550 (8.4)</td>
<td>2722 (28.5)</td>
<td></td>
</tr>
<tr>
<td>Income (CNY(^a), median (IQR)</td>
<td>2500 (1250-5000)</td>
<td>2500 (1250-4583)</td>
<td>2917 (1250-5000)</td>
<td>.13</td>
</tr>
<tr>
<td>Smoking, n (%)</td>
<td>3651 (22.3)</td>
<td>3522 (52.3)</td>
<td>129 (1.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Drinking, n (%)</td>
<td>4542 (27.7)</td>
<td>3673 (54.6)</td>
<td>869 (9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Physical activity index, median (IQR)</td>
<td>4 (1.28-16.00)</td>
<td>3.36 (1.18-16.00)</td>
<td>5.03 (1.37-14.67)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>History of fracture or fall, n (%)</td>
<td>1287 (7.9)</td>
<td>504 (7.5)</td>
<td>783 (8.1)</td>
<td>.14</td>
</tr>
<tr>
<td>BMI (kg/m(^2)), mean (SD)</td>
<td>24.99 (3.38)</td>
<td>24.74 (3.26)</td>
<td>25.17 (3.44)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>History of taking calcium, n (%)</td>
<td>766 (4.7)</td>
<td>197 (2.9)</td>
<td>569 (5.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>History of taking drugs, n (%)</td>
<td>8709 (53.4)</td>
<td>3605 (53.7)</td>
<td>5104 (53.4)</td>
<td>.61</td>
</tr>
<tr>
<td>Postmenopausal in female participants, n (%)</td>
<td>N/A(^b)</td>
<td>N/A</td>
<td>7493 (77.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>FBG(^c) (mmol/L), mean (SD)</td>
<td>6.21 (1.86)</td>
<td>6.23 (1.89)</td>
<td>6.20 (1.84)</td>
<td>.36</td>
</tr>
<tr>
<td>TC(^d) (mmol/L), mean (SD)</td>
<td>5.06 (0.93)</td>
<td>4.95 (0.91)</td>
<td>5.13 (0.93)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>TG(^e) (mmol/L), mean (SD)</td>
<td>1.66 (1.29)</td>
<td>1.63 (1.42)</td>
<td>1.68 (1.19)</td>
<td>.02</td>
</tr>
<tr>
<td>HDL-C(^f) (mmol/L), mean (SD)</td>
<td>1.52 (0.42)</td>
<td>1.52 (0.47)</td>
<td>1.52 (0.38)</td>
<td>.67</td>
</tr>
<tr>
<td>LDL-C(^f) (mmol/L), mean (SD)</td>
<td>2.83 (0.79)</td>
<td>2.74 (0.78)</td>
<td>2.88 (0.80)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>HTN(^b), n (%)</td>
<td>9558 (58.4)</td>
<td>4220 (62.7)</td>
<td>5338 (55.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>T2DM(^i), n (%)</td>
<td>3113 (19)</td>
<td>1259 (18.7)</td>
<td>1854 (19.2)</td>
<td>.14</td>
</tr>
<tr>
<td>Dyslipidemia, n (%)</td>
<td>9523 (58.2)</td>
<td>3651 (54.2)</td>
<td>5872 (60.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CHD(^j), n (%)</td>
<td>267 (1.6)</td>
<td>133 (2)</td>
<td>134 (1.4)</td>
<td>.004</td>
</tr>
<tr>
<td>Stroke, n (%)</td>
<td>890 (5.4)</td>
<td>389 (5.8)</td>
<td>501 (5.2)</td>
<td>.11</td>
</tr>
<tr>
<td>Cancer, n (%)</td>
<td>439 (2.7)</td>
<td>227 (3.4)</td>
<td>212 (2.2)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\) A currency exchange rate of 1 CNY= US $0.14 is applicable.

\(^b\) N/A: not applicable.

\(^c\) FBG: fasting blood glucose.

\(^d\) TC: total cholesterol.

\(^e\) TG: triglycerides.

\(^f\) HDL-C: high-density lipoprotein cholesterol.
Male participants had higher prevalence of hypertension (4220/6732, 62.7%), CHD (133/6732, 2%), and cancer (227/6732, 3.4%) than female participants (5338/9645, 55.3%; 134/9645, 1.4%; and 212/9645, 2.2%) at a significant level of $P<.05$ but had a lower prevalence of dyslipidemia (3651/6730, 54.2%) than female participants (5872/9640, 60.9%; $P<.05$). As expected, male participants had a higher proportion of smoking (3522/6732, 52.3%) and drinking (3673/6732, 54.6%) than female participants (129/9644, 1.3%; and 869/9643, 9%; $P<.001$). Male participants had lower levels of PAI (median 3.36, IQR 1.18-16.00), BMI (mean 24.74, SD 3.26 kg/m$^2$), TC (mean 4.95, SD 0.91 mmol/L), and TG (mean 1.63, SD 1.42 mmol/L) than female participants (median 5.03, IQR 1.37-14.67; mean 25.17, SD 3.44 kg/m$^2$; mean 5.13, SD 0.93 mmol/L; and mean 1.68, SD 1.19 mmol/L; $P<.05$).

**Prevalence of Osteopenia and Osteoporosis Among Age Groups**

The prevalence rates of osteopenia and osteoporosis were 40.5% (6633/16,377) and 7.93% (1299/16,377), respectively. The age-standardized prevalence rates of osteopenia and osteoporosis were 27.32% (287,877,129.4/1,053,861,940) and 3.51% (36,974,582.3/1,053,861,940), respectively. In the population aged 50 years and older, the age-standardized prevalence rates of osteopenia and osteoporosis were 42.34% (142,938,163.8/337,624,151) and 8.96% (30,253,250.83/337,624,151), respectively. There was an increase in osteopenia prevalence rate from 21.47% (120/559) in the 18-year-old group to 56.23% (754/1341) in the 75-year-old group (Figure 2A). Meanwhile, the prevalence rate of osteoporosis increased from 0.89% (5/559) in the 18-year-old group to 17.23% (231/1341) in the 75-year-old group (Figure 2B).

---

4 LDL-C: low-density lipoprotein cholesterol.
5 HTN: hypertension.
6 T2DM: type 2 diabetes mellitus.
7 CHD: coronary heart disease.
Sex-Specific Prevalence Changes of Osteopenia and Osteoporosis With Age

The prevalence rates of osteopenia and osteoporosis among female participants were 4238/9645, 43.94% and 1130/9645, 11.72% were higher than in male participants 2395/6732, 35.58% and 169/6732, 2.51%, respectively; $P<0.001$. In the study population, the age-standardized prevalence rates of osteopenia and osteoporosis were 22.38% (119,148,921.4/532,444,837) and 0.94% (5,022,330.82/532,444,837) for male participants, while it was 30.97% (161,470,215.4/521,417,103) and 5.52% (28,781,912.82/521,417,103) for female participants. In the population aged 50 years and older, the age-standardized prevalence of osteopenia and osteoporosis was 36.1% (60,821,276.55/168,477,281) and 2.65% (4,471,495.81/168,477,281) for male participants and 47.14% (79,727,904.8/169,146,870) and 14.26% (24,113,222.94/169,146,870) for female participants. Among male participants, the observed prevalence rates of osteopenia (Figure 2A) and osteoporosis (Figure 2B) were 20.71% (41/198) and 0% (0/198) in the 18-year-old group and increased to 51.54% (351/681) and 5.43% (37/681) in the 75-year-old group, respectively. Among female participants, the observed prevalence rates of osteopenia (Figure 2A) and osteoporosis (Figure 2B) were 21.88% (79/361) and 1.39% (5/361) in the 18-year-old group and increased to 61.06% (403/660) and 29.39% (194/660) in the 75-year-old group, respectively. The prevalence rates of osteopenia and osteoporosis among age groups are shown in Multimedia Appendix 2.

Prevalence of Osteopenia and Osteoporosis Between Postmenopausal and Premenopausal Female Participants

The prevalence rates of osteopenia and osteoporosis in postmenopausal female participants were higher (3638/7493, 48.88% and 1130/7493, 15.03%) than in premenopausal female participants 2395/6732, 35.58% and 169/6732, 2.51%, respectively; $P<0.001$. In the study population, the age-standardized prevalence rates of osteopenia and osteoporosis were 22.38% (119,148,921.4/532,444,837) and 0.94% (5,022,330.82/532,444,837) for male participants, while it was 30.97% (161,470,215.4/521,417,103) and 5.52% (28,781,912.82/521,417,103) for female participants. In the population aged 50 years and older, the age-standardized prevalence of osteopenia and osteoporosis was 36.1% (60,821,276.55/168,477,281) and 2.65% (4,471,495.81/168,477,281) for male participants and 47.14% (79,727,904.8/169,146,870) and 14.26% (24,113,222.94/169,146,870) for female participants. Among postmenopausal female participants, the observed prevalence rates of osteopenia (Figure 2A) and osteoporosis (Figure 2B) were 20.71% (41/198) and 0% (0/198) in the 18-year-old group and increased to 51.54% (351/681) and 5.43% (37/681) in the 75-year-old group, respectively. Among premenopausal female participants, the observed prevalence rates of osteopenia (Figure 2A) and osteoporosis (Figure 2B) were 21.88% (79/361) and 1.39% (5/361) in the 18-year-old group and increased to 61.06% (403/660) and 29.39% (194/660) in the 75-year-old group, respectively. The prevalence rates of osteopenia and osteoporosis among age groups are shown in Multimedia Appendix 2.
48.55% and 1053/7493, 14.05%) than in premenopausal female participants (538/2026, 26.55% and 53/2026, 2.62%; \( P < .001 \)).

In premenopausal female participants, the observed prevalence of osteopenia was 21.41% (76/355) among the 18-year-old group and increased as the age increased. In postmenopausal female participants, the observed prevalence of osteopenia was 23.81% (5/21) in the 40-year-old group and increased to 61.37% (394/642) in the 75-year-old group (Figure 3A).

**Figure 3.** Prevalence of osteopenia and osteoporosis between postmenopausal and premenopausal female participants. In this cross-sectional study, the age-specific prevalence of osteoporosis and osteopenia was calculated for 7493 Chinese postmenopausal female participants and 2162 premenopausal female participants. The prevalence rates of osteopenia significantly increased with age in premenopausal and postmenopausal female participants, and all \( P \text{trend} \) values were less than .001, the results were statistically significant (A). Prevalence of osteopenia showed a significant statistical difference between premenopausal and postmenopausal female participants in the age group 50 years, \( P \) value was less than .001. The prevalence rate of osteoporosis significantly increased with age in premenopausal and postmenopausal female participants and all \( P \text{trend} \) values were less than .001, the results were statistically significant (B). Prevalence rates of osteoporosis showed a significant statistical difference between premenopausal and postmenopausal female participants in the age group 50 years (\( P < .05 \)).

In premenopausal female participants, the observed prevalence of osteoporosis was 1.25% (5/400) among the 40-year-old group and increased to 14.29% (2/14) among the 60-year-old group. In postmenopausal female participants, the observed prevalence of osteoporosis was 6.56% (17/259) among the 45-year-old group and increased to 29.28% (188/642) in the 75-year-old group (Figure 3B). The prevalence rates of osteopenia and osteoporosis in postmenopausal and premenopausal female participants are shown in Multimedia Appendix 3.

**Prevalence of Osteopenia and Osteoporosis in the Subgroups With and Without a History of Calcium Intake**

Among participants with and without a history of calcium intake, the prevalence rate of osteopenia and osteoporosis was 44.52% (341/766), 40.28% (6276/15,581), 6.27% (48/766), and 8.02% (1250/15,581), respectively. For male participants, the prevalence rate of osteopenia was lower among those taking calcium (55/197, 27.92%) than those not taking calcium (2339/6527 35.84%; \( P = .03 \)). The prevalence rate of osteopenia was higher among female participants taking calcium (286/569, 50.26%) than those not taking calcium (3937/9054, 43.48%; \( P = .04 \); Figure 4A). The prevalence rates of osteoporosis among
the overall study participants taking calcium and not taking calcium were 4.06% (8/197) and 2.47% (161/6527) in male participants and 7.03% (40/569) and 12.03% (1089/9054) in female participants, respectively, with significant differences among female participants ($P=0.004$; Figure 4B). The prevalence rates of osteopenia and osteoporosis in subgroups with and without taking calcium are shown in Multimedia Appendix 4.

**Figure 4.** Prevalence of osteopenia and osteoporosis in the subgroups with and without a history of calcium intake. In this cross-sectional study, the age-specific prevalence of osteoporosis and osteopenia were calculated for 766 Chinese adults taking calcium and separately for 197 male participants and 569 female participants as well as for 15,581 Chinese adults not taking calcium and separately for 6527 male participants and 9054 female participants. The prevalence rate of osteopenia significantly increased with the age in the whole study population, male, and female participants with and without a history of taking calcium (A). All $P$-trend values were less than .001, the results were statistically significant except in male participants with a history of taking calcium. The prevalence rates of osteoporosis significantly increased with the age in the whole study population, male and female participants with and without a history of taking calcium (B). All $P$-trend values were less than .001, the results were statistically significant. The prevalence rates of osteopenia and osteoporosis showed no significant statistical difference between participants with and without a history of taking calcium in all age groups.

**Association Analysis of Taking Calcium With Osteopenia and Osteoporosis**

After adjustment for age, sex, PAI, occupation, history of hypertension, history of T2DM, history of dyslipidemia, history of CHD, history of stroke, and history of cancer, there was no significant association between calcium intake and osteopenia and osteoporosis, and the corresponding ORs (95% CIs) were 1.10 (0.94-1.29) and 0.90 (0.61-1.34), respectively.

Furthermore, sex-stratified analysis showed that calcium intake increased the risk of osteopenia in female participants (OR 1.26, 95% CI 1.04-1.52; Multimedia Appendix 5) after adjusting age, PAI, occupation, history of hypertension, history of T2DM, history of dyslipidemia, history of CHD, history of stroke, and history of cancer.

**Prevalence of Chronic Disease Among Different T-Score Groups**

In the total study population, the prevalence of DM, dyslipidemia, CHD, and CCVD differed among the low, medium-low, medium-high, and high quartiles of the T-score. From the low quartile to the high quartile of the T-score, the prevalence rates of DM (752/4037, 18.63%; 779/4029, 19.33%; 769/3894, 19.75%; and 869/3879, 22.4%) and dyslipidemia (2228/4036, 55.2%; 2304/4027, 57.21%; 2306/3891, 59.26%; and 2379/3878, 61.35%) were linearly increased ($P<.001$), while the prevalence rate of cancer (112/4037, 2.77%; 110/4029, 2.73%; 103/3894, 2.65%; and 77/3879, 1.99%) was decreased ($P=.03$). The prevalence rate of CHD was 1.29% (52/4037), 1.54% (62/4029), 2.16% (84/3894), and 1.99% (77/3879) from low quartile to high quartile. The prevalence rate of CCVD was
Further stratified analysis by BMI showed that in the normal weight group, the prevalence rates of DM were 13.49% (241/1786), 13.03% (214/1642), 13.27% (206/1552), and 16.25% (217/1335) among the lowest, second, third, and highest quartile values of the T-score. In the overweight group, the prevalence rates of dyslipidemia were 64.75% (395/610), 71.33% (510/715), 72.83% (520/714), and 71.6% (580/810), and the prevalence rates of CHD were 0.82% (5/610), 2.09% (15/716), 2.66% (19/714), and 3.46% (28/810) from the lowest quartile to the highest quartile. All the P values were less than .05 (Multimedia Appendix 6).

Discussion

Principal Findings

With a growing older adult population, the public health issue caused by osteoporosis and osteopenia is worthy of attention. However, the prevalence of osteoporosis varied across different studies because of inconsistent measuring instruments [24], bone sites [25], age group, sample size, region [26], and diagnosis criteria. This study investigated the prevalence of osteoporosis and osteopenia in a relatively large sample size Chinese population and explored the impact of chronic common disorders on BMD. With these results, interventions can be carried out for the target population to reduce the burden that osteoporosis poses to public health.

In this study, the age-standardized prevalence rates of osteoporosis and osteopenia were 3.51% (36,974,582.3/1,053,861,940) and 27.32% (287,877,129.4/1,053,861,940), respectively. Similarly, Qiao et al [27] analyzed the prevalence of osteoporosis and osteopenia using the same measuring instruments at the same bone site with this study in rural areas of Henan province in Middle-Eastern China and reported the age-standardized prevalence rates to be 11.76% and 42.09%, respectively. Another study reported that Shanghai had the lowest prevalence rates of osteoporosis and osteopenia compared with other regions in China [28]. A possible reason for the differences in the prevalence of osteopenia and osteoporosis is that each province in China may have different lifestyle-related characteristics that are population specific due to educational, sociocultural, and geographic considerations. For instance, a study has shown that provinces where alcohol is produced, such as Jiangsu, Anhui, Shandong, and Sichuan, had higher alcohol consumption proportions compared with the national level [29]. Meanwhile, lifestyle factors including alcohol are associated with poor bone health [30].

Similarly, Zeng et al [31] measured the BMD values using GE Lunar dual-energy X-ray absorptiometry (DXA) in China and estimated the age-standardized prevalence rate of osteoporosis in male and female participants older than 50 years old to be 6.46% and 29.13%, respectively. Wang et al [32] measured the BMD values using DXA and reported that the prevalence rate of osteoporosis among male participants aged 40 years or older was 5% and 20.6% among female participants of the same age group in China. A previous study has shown that QUS and X-ray density measurement BMD methods (DXA and central quantitative computed tomography) showed the same ability to distinguish between normal populations and osteoporosis [33]. Many epidemiological studies use QUS as a screening tool for osteoporosis because it is easier to expand the sample size, simplify the measuring method, and reduce the cost [34,35].

Other important prospective studies have assessed fracture risk through calcaneal QUS [36,37]. The EPIC-Norfolk prospective population study [37] demonstrated that measuring BMD by QUS at the calcaneus can effectively predict fracture risk in the future. It is worth noting that the QUS T-score may be more adaptable for clinically detecting the risk of osteoporosis in female patients [38]. However, the age-standardized prevalence of osteoporosis at different sites (lumbar spine, femoral neck, and total femur) is also different [31]. Wang et al [39] and Chen et al [40] reported gaps in the prevalence of osteoporosis. Wang et al [28] detected the BMD in the lumbar vertebra and left hip joint by DXA and also confirmed that the prevalence of osteoporosis would be greatly reduced if diagnosed with BMD in every single region only. It is worth noting that the prevalence rates of postmenopausal female participants in this study are close to the results of Wang et al [28], which diagnosed hip BMD. Thus, we speculated that the relative differences in prevalence rates may be due to different measurement regions and a single region. The prevalence rates of osteopenia and osteoporosis in the current population were moderate compared to the previous studies. Importantly, the age-standardized prevalence rates of osteopenia and osteoporosis were higher (79,727,904.8/169,146,870, 47.14% and 24,113,222.94/169,146,870, 14.26%) in female participants aged 50 years and older.

There are significant sex differences in the prevalence and pathogenesis of osteoporosis [41]. Previous studies found that osteoporosis was more prevalent among female than male participants, especially among postmenopausal female participants [27,31]. For menopausal females, the sudden change in hormonal levels can contribute to increased bone loss. Our results are consistent with the previous reports.

A prospective cohort study in an Italian population assessed the association between dietary calcium intake and osteoporosis and a fragility fracture. Their results indicated that low calcium intake was a risk factor for low BMD in Italy [42]. However, in one of the central populations of the Canadian Multicentre Osteoporosis Study, it was found that older female participants taking calcium supplements had no significant BMD preservation [43]. In this study, the prevalence of osteopenia was higher in female participants taking calcium than in those who did not. That may be attributed to the fact that more patients with osteopenia receive diagnosis and treatment advice and take a calcium supplement. Moreover, considering the design of our study is a cross-sectional study, a causal association between calcium supplementation and BMD cannot be stated.
This study indicated that patients with T2DM, dyslipidemia, CHD, and CCVD were associated with higher BMD. In reality, previous observations reported that the association of BMD value with T2DM was inconclusive [15,16]. Recently, Wang et al [28] also found that T2DM was associated with higher BMD. Even some large clinical studies revealed the same results [44,45]. Some researchers have speculated the anabolic effect of insulin on bone tissue [46]. Hyperinsulinemia caused by insulin resistance in T2DM may negatively affect sex hormone–binding globulin, leading to higher free sex hormone levels that may protect patients with T2DM from bone loss [47,48]. However, animal models have indicated that although bone density is greater in diabetes, the bone structure is more fragile, with fractures occurring under a smaller load and the bones exhibiting reduced mechanical indices [49]. Thus, even though the BMD was high in T2DM, it still should pay more attention to patients with T2DM.

As a marker of bone turnover, osteoprotegerin (OPG) can reflect the level of bone metabolism and is used to treat osteoporosis. The animal study results found that OPG-deficient mice develop osteoporosis and premature arterial calcification, suggesting that the OPG system affects vascular calcification [50]. The human study showed that increased OPG levels were associated with the risk of fatal stroke and raised the possibility that the OPG system may be involved in vascular calcification [51]. However, this study was unable to confirm the association between BMD and stroke, although the results indicated the association between BMD and CHD and CCVD.

The findings of our investigation have significant implications for public health. With the increase of Chinese older adults, the prevalence and burden of osteoporosis and osteopenia are expected to increase in the coming years. This study offers information that can be applied to intervention programs meant to prevent or lessen the burden of osteoporosis in China.

**Strengths and Limitations**

This study has a relatively large sample size of a Chinese population involving a fairly broad spectrum of ages taken to boost the statistical power. The study also explored the trend of prevalence of osteopenia and osteoporosis across different age groups and identified high-risk individuals and some protective factors such as early calcium supplementation. However, this study has some limitations. First, a causal association between calcium and osteoporosis cannot be drawn since this was a cross-sectional study. A distinct temporal order in which the cause comes before the result is necessary for causality. It is difficult to discern the direction of the relationship when using cross-sectional research because they can only reveal associations at a single point in time. Although cross-sectional research can yield useful findings, it can only reveal information regarding potential protective or risk factors. Therefore, to support the findings of this investigation, experimental studies should be carried out. Second, the average age of the participants in this study was 60 years, while the number of individuals younger than 40 years old was relatively small. Thus, the estimated prevalence rate of young people may not be sufficient and stable; further studies involving a larger number of participants aged younger than 40 years old might verify the prevalence of osteoporosis and osteopenia for that age group. Third, patients with chronic diseases affecting BMD were self-reported and did not include patients who were bedridden for a long time, which might underestimate the prevalence rates of osteoporosis and osteopenia. Despite this, the necessity of a positive response to promoting bone health and intervention measures for osteopenia and osteoporosis is more strongly suggested.

**Conclusions**

Our findings suggest that as age increases, females had a higher increase in the prevalence of osteopenia and osteoporosis than males. Importantly, postmenopausal females had higher prevalence rates of osteopenia and osteoporosis than premenopausal females. Higher BMD is associated with an increased risk of DM, dyslipidemia, CHD, and CCVD but a decreased risk of cancer. This research will help develop aging coping strategies to promote bone health and prevent osteoporosis.

**Acknowledgments**

We would like to thank Hailong Zhao, Lijun Zhu, Ting Zhang, Jinbo Wen, Hui Zhu, and Zhengmei Fang for quality control and the staff for working on data collecting and sorting. We would like to thank Solim Essomandan Clemence Bafei for helping to edit the English language. This study was funded by grants from the National Key Research and Development Program of China (Grant 2018YFC2000703) the National Natural Science Foundation of China (Grant 81872686 and 81573232), the Social Development Guidance Project of Zhenjiang (FZ2015064 and SH20202050), Livelihood Science and Technology Plan Project of Science and Technology Bureau in Jurong (SF2016896226). The funders had no role in the study design, collection, analysis, or interpretation of the data; writing the manuscript; or the decision to submit the paper for publication.

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

**Authors’ Contributions**

CS initiated, conceived, and supervised the study; YL, M Liu, and HX participated in the implementation of the onsite work; YF, J Cai, and J Cao checked the results of the onsite work; QL, JM, and TZ participated in the data collation; LW and M Li assisted...
with the study and analyses; and YF completed the analyses and led the writing of the paper. All authors read and approved the final manuscript.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Questionnaire.

[PDF File (Adobe PDF File), 963 KB - publichealth_v10i1e48947_app1.pdf]

**Multimedia Appendix 2**

Prevalence rates of osteopenia and osteoporosis among age-groups.

[DOCX File, 17 KB - publichealth_v10i1e48947_app2.docx]

**Multimedia Appendix 3**

Prevalence rates of osteopenia and osteoporosis in postmenopausal and premenopausal female participants.

[DOCX File, 16 KB - publichealth_v10i1e48947_app3.docx]

**Multimedia Appendix 4**

Prevalence rates of osteopenia and osteoporosis in subgroups with and without taking calcium.

[DOCX File, 21 KB - publichealth_v10i1e48947_app4.docx]

**Multimedia Appendix 5**

Association analysis of history of taking calcium with osteopenia and osteoporosis.

[DOCX File, 17 KB - publichealth_v10i1e48947_app5.docx]

**Multimedia Appendix 6**

Prevalence of chronic disease among different T-score levels.

[DOCX File, 25 KB - publichealth_v10i1e48947_app6.docx]

**Multimedia Appendix 7**

Prevalence of chronic disease among different T-score levels.

[PNG File, 164 KB - publichealth_v10i1e48947_app7.png]

**References**


Abbreviations

BMD: bone mineral density
CCVD: cardiovascular and cerebrovascular disease
CHD: coronary heart disease
DALY: disability-adjusted life-year
DXA: dual-energy X-ray absorptiometry
MET: metabolic equivalent
OPG: osteoprotegerin
OR: odds ratio
PA: physical activity
PAI: physical activity index
QUS: quantitative ultrasound system
T2DM: type 2 diabetes mellitus
TC: total cholesterol
TG: triglyceride

© Yao Fan, Qun Li, Yu Liu, Jing Miao, Ting Zhao, Jinxin Cai, Min Liu, Jun Cao, Haiyong Xu, Lai Wei, Mengxia Li, Chong Shen. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 05.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Association Between Frequency of Muscle-Strengthening Exercise and Depression Symptoms Among Middle and High School Students: Cross-Sectional Survey Study

Hao Wang¹, MPH; Huaidong Du²,³, PhD; Yunqi Guan¹, MPH; Jieming Zhong¹, MPH; Na Li¹, MPH; Jin Pan¹, MPH; Min Yu¹, MPH

¹Department of Noncommunicable Disease Control and Prevention, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
²Medical Research Council Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom
³Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Min Yu, MPH
Department of Noncommunicable Disease Control and Prevention
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Road
Binjiang Distric
Hangzhou, 310051
China
Phone: 86 571 87115005
Email: zjcdcmyu@163.com

Abstract

Background: Existing literature on the association between the frequency of muscle-strengthening exercise (MSE) and depression among adolescents is limited and contradictory.

Objective: This study aimed to elucidate the association of MSE frequency with depression symptoms among middle and high school students in China.

Methods: A total of 27,070 students in grades 7-12 from 376 middle and high schools were surveyed using an anonymous self-administered questionnaire between April and June 2022. Information on engaging in MSE was self-reported, and depression symptoms were assessed using the Patient Health Questionnaire-9 (PHQ-9). Poisson regression was used to examine the association between MSE frequency and depression symptoms.

Results: Among the 27,006 eligible students, 51.6% (n=13,933) were boys, and the mean age was 15.6 (SD 1.7) years. The overall prevalence of meeting MSE recommendations (ie, engaging in MSE ≥3 days/week) was 34.6% (95% CI 32.6%-36.6%; n=9145); the prevalence was higher in boys (43.8%, 95% CI 41.8%-45.8%; 6067/13,933) than in girls (24.3%, 95% CI 22%-26.6%; 3078/13,073; P<.001). A total of 5882 (21.8%) students reported having depression symptoms. After adjustment for sociodemographic status, lifestyle factors, academic performance, and experience of physical fighting, compared to students who did not engage in MSE, the prevalence ratios (PRs) for depression symptoms were 0.98 (95% CI 0.97-0.99) for those engaging in MSE once a week, 0.95 (95% CI 0.93-0.97) for 2 days/week, 0.93 (95% CI 0.90-0.96) for 3 days/week, 0.90 (95% CI 0.87-0.94) for 4 days/week, 0.88 (95% CI 0.84-0.93) for 5 days/week, 0.86 (95% CI 0.81-0.92) for 6 days/week, and 0.84 (95% CI 0.78-0.90) for 7 days/week, respectively.

Conclusions: The overall prevalence of meeting MSE recommendations among Chinese adolescents is low. The frequency of MSE was inversely associated with depression symptoms.

(JMIR Public Health Surveill 2024;10:e50996) doi:10.2196/50996

KEYWORDS

depression symptoms; muscle-strengthening exercise; adolescents; cross-sectional study

https://publichealth.jmir.org/2024/1/e50996
Introduction

Depression disorders account for a large and increasing health burden among adolescents aged 10 to 19 years worldwide [1]. The estimated number of disability-adjusted life years among adolescents aged 10 to 19 years diagnosed with depression disorders increased worldwide from 3.4 million to 4.3 million between 1990 and 2019 [2], and in 2019, there were 22.9 million adolescents aged 10 to 19 years with depression disorders worldwide, of whom 1.5 million were in China [2]. Depression among adolescents may contribute to many negative consequences, including functional impairment, poor cognitive development, poor academic performance, and suicide [3,4]. Depression is also associated with a wide variety of chronic physical disorders, including arthritis, asthma, cancer, cardiovascular disease, diabetes, obesity, hypertension, chronic respiratory disorders, and dementia, acting via multiple mechanisms that are not yet completely clear [5].

Muscle-strengthening exercise (MSE) is defined as physical exercise that increases skeletal muscle strength, power, endurance, and mass (eg, strength training, resistance training, or muscular strength and endurance exercises) [6]. Although MSE is now included in many national public health guidelines, in comparison to physical activity guidance on aerobic activities, there is still a lack of emphasis and guidance on MSE throughout public health policies in various countries [7-9]. Prevalence rates for meeting MSE recommendations (ie, engaging in MSE ≥3 days/week) among adolescents vary across countries and regions, as do secular trends. For instance, the prevalence rate of adhering to MSE guidelines is 44.9% in the United States [10], 53.7% in Canada [11], and 39.3% in China [12]. The prevalence in US adolescents decreased from 55.6% in 2011 to 49.5% in 2019 [13], while among Korean adolescents it slightly increased from 20.1% in 2009 to 21.9% in 2019 [14]. More recently, emerging epidemiological and clinical evidence has demonstrated that engagement in MSE could reduce the risk of chronic diseases, including obesity [14,15], hypertension [16], osteoporosis [17], and metabolic syndrome [18]; it can also enhance cardiometabolic health [19] and improve physical fitness [20].

Previous studies documented that MSE was inversely associated with depression. However, the majority of this research focused on adults [21-24], and little is known about adolescents. Furthermore, existing literature on the association of MSE frequency with depression among adolescents is contradictory. While one study indicated that meeting MSE recommendations was inversely associated with depression [25], a null association was found among US high school students and young adults aged 20-39 years [26,27]. Hence, this study was designed to evaluate the association between MSE frequency and depression symptoms among school students in China.

Methods

Study Design

This cross-sectional survey applied a multistage cluster sampling method to recruit participants. In stage 1, 30 counties or districts were sampled randomly from all 90 counties and districts in Zhejiang Province. In stage 2, 11 middle school classes, 6 academic high school classes, and 6 vocational high school classes were selected randomly within each chosen county or district. In stage 3, all students in the selected classes were invited to participate in the study. Students in the selected classes who also returned signed informed consent forms were included in the analysis unless they had serious health problems or illnesses that would restrict them from participating, including intellectual disabilities or language disorders.

A self-administered anonymous questionnaire was filled in by students in a classroom setting without school teachers’ supervision. The field survey was implemented by trained staff from county centers for disease control and prevention (CDC) using standardized procedures.

Outcome Variables

The severity of depressive symptoms was assessed using the Patient Health Questionnaire-9 (PHQ-9) [28]. The PHQ-9, widely used among adolescents [29-32], is a brief scale designed to screen for symptoms of major depressive disorder within the past 2 weeks based on the codes of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Participants were asked to score the following 9 items [33]: (1) “little interest or pleasure in doing things,” (2) “feeling down, depressed, or hopeless,” (3) “trouble falling or staying asleep, or sleeping too much,” (4) “feeling tired or having little energy,” (5) “poor appetite or overeating,” (6) “feeling bad about yourself,” (7) “trouble concentrating,” (8) “moving or speaking so slowly that other people could notice,” and (9) “thoughts that you would be better off dead or of hurting yourself.” Scores for each item ranged from 0 (not at all) to 3 (nearly every day); therefore, the total combined score ranged between 0 and 27. Following the current recommendation [28], depression was defined as having a total PHQ-9 score of no less than 10. A previous study documented high specificity (85%) and high sensitivity (88%) of the PHQ-9 scale in detecting depression using the cutoff value of 10 [34]. In addition, depression was divided into 4 groups according to the sum of the PHQ-9 score (5-9: mild depression; 10-14: moderate depression; 15-19: moderately severe depression; and 20-27: severe depression) [35].

Exposure Variables

The frequency of engaging in MSE was evaluated through the question “During the past 7 days, on how many days did you do exercises to strengthen or tone your muscles, such as push-ups, sit-ups, or weight lifting?” Response options included “none,” “1 day,” “2 days,” “3 days,” “4 days,” “5 days,” “6 days,” and “7 days.” This item, with an acceptable reliability for children and adolescents (κ coefficient >0.55) [36], has been widely used for health behavior surveillance worldwide [10,11]. Meeting MSE recommendations was defined as engaging in MSE at least 3 days in the past 7 days, which is in accordance with World Health Organization (WHO) guidelines [6].

Ethical Considerations

The study was approved by the Ethics Committee of the Zhejiang Provincial CDC (grant number 2022-007-01). The survey was anonymous, and participants did not need to write down their name on the questionnaire. Every student who...
finished the questionnaire received a gift as compensation (a pencil box). A written consent form was provided by all students and their guardians 2 weeks prior to the survey.

**Statistical Analysis**

Continuous variables are presented as the mean (SD). Categorical variables are presented as the percentage (95% CI). Weighted prevalences were calculated. A modified Poisson regression, which is considered to be more robust than the traditionally used logistic regression [37], was used to examine the association between MSE frequency and depression symptoms [38]. Potential confounding factors comprised sociodemographic status, behavioral lifestyle factors, academic performance, and experience of physical fighting. Prevalence ratios (PRs) were calculated using two regression models: model 1 was adjusted for age (≤13 years, 14-15 years, or ≥16 years), sex (boy or girl), region (urban or rural), and type of school (middle school, academic high school, or vocational high school); model 2 was additionally adjusted for paternal and maternal education level (middle school or below, high school, or college or above), parental marital status (married or other), household income (very poor/poor, fair, or rich/very rich), cigarette smoking (yes or no), alcohol drinking (yes or no), physical activity (none, 1-2 days/week, 3-5 days/week, or 6-7 days/week), academic performance (excellent, middle, or poor), and being involved in physical fighting (yes or no). In addition, both exposure (ie, frequency of engaging in MSE) and outcome (ie, depression symptom score) variables were considered to be continuous variables, and multiple linear regression analyses were used to ascertain the association of days of engaging in MSE with depression symptom score. In sensitivity analyses, additional adjustment for experience of being bullied was performed. All statistical analyses were performed using SAS (version 9.4; SAS Institute). The statistical significance level was set at a P value <.05 using a 2-sided test.

**Results**

**Descriptive Statistics**

Overall, 28,043 students from 376 schools were invited and a total of 27,070 students participated in the survey (114 refused to participate and 859 were absent from school on the survey day), resulting in a response rate of 96.5%. Out of 27,070 students, a total of 64 were excluded because of either an incomplete questionnaire (n=40), missing at least 1 of the 9 items in the PHQ-9 questionnaire (n=17), or missing information on engaging in MSE (n=7). Ultimately, 27,006 students, consisting of 13,933 boys and 13,073 girls, were included in the current analyses (Figure S1 in Multimedia Appendix 1). The mean age was 15.6 (SD 1.7) years. A total of 21.8% (n=5882) of students reported experiencing depression symptoms. In addition, the prevalence of mild, moderate, moderately severe, and severe depression symptoms was 39.9% (95% CI 38.8%-40.9%; 10,902/27,006), 14.4% (95% CI 13.7%-15%; 3746/27,006), 5.3% (95% CI 5%-5.7%; 1407/27,006), and 2.7% (95% CI 2.4%-3%; 729/27,006), respectively (Table S1 in Multimedia Appendix 1).

Compared to students who did not perform MSE, students engaging in MSE were more likely to be young, male, from an urban area, attend middle school, live in an intact family, have parents educated to the college level or above, come from a high-income family, be physically active, achieve excellent academic performance, smoke cigarettes, and drink alcohol (Table 1).
Table 1. Participant characteristics by frequency of muscle-strengthening exercise (N=27,006). Percentages are weighted.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Frequency of muscle-strengthening exercise (days/week) among participants</th>
<th>P value for trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics</td>
<td>0 (n=10,703)</td>
<td>1 (n=3474)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>16.1 (1.7)</td>
<td>15.6 (1.7)</td>
</tr>
<tr>
<td>Girls, n (%)</td>
<td>6560 (60.1)</td>
<td>1802 (51.1)</td>
</tr>
<tr>
<td>Rural, n (%)</td>
<td>6636 (66.3)</td>
<td>2073 (65.4)</td>
</tr>
<tr>
<td>Middle school, n (%)</td>
<td>3701 (39.5)</td>
<td>1603 (50)</td>
</tr>
<tr>
<td>Living in intact family, n (%)</td>
<td>9182 (86.3)</td>
<td>3047 (88.3)</td>
</tr>
<tr>
<td>Father with college education or above, n (%)</td>
<td>1687 (16.3)</td>
<td>624 (18.7)</td>
</tr>
<tr>
<td>Mother with college education or above, n (%)</td>
<td>1571 (15.4)</td>
<td>597 (17.6)</td>
</tr>
<tr>
<td>Family with high income, n (%)</td>
<td>725 (7)</td>
<td>266 (8.4)</td>
</tr>
<tr>
<td>Physically active ≥6 d/wk, n (%)</td>
<td>791 (7.4)</td>
<td>306 (8.4)</td>
</tr>
<tr>
<td>Excellent academic performance, n (%)</td>
<td>1818 (16.8)</td>
<td>675 (19.3)</td>
</tr>
<tr>
<td>Cigarette smoking, n (%)</td>
<td>370 (3.4)</td>
<td>135 (3.8)</td>
</tr>
<tr>
<td>Alcohol drinking, n (%)</td>
<td>1668 (15.2)</td>
<td>569 (15.8)</td>
</tr>
<tr>
<td>Physical fighting, n (%)</td>
<td>1078 (10.5)</td>
<td>477 (14.2)</td>
</tr>
</tbody>
</table>

Prevalence of Meeting MSE Recommendations

The prevalence of meeting MSE recommendations was 34.6% (95% CI 32.6%-36.6%; 9145/27,006) overall and 41.4% (95% CI 37.4%-45.5%), 42.2% (95% CI 38.3%-46.2%), and 25.2% (95% CI 23.3%-27.1%) for students aged ≤13 years, 14-15 years, and ≥16 years, respectively (P<.001). The prevalence was higher among boys (43.8%, 95% CI 41.8%-45.8%) than among girls (24.3%, 95% CI 22%-26.6%; P<.001). In addition, the prevalence among students attending middle school, academic high school, and vocational high school was 43.6% (95% CI 40.3%-46.9%), 25.5% (95% CI 23.2%-27.8%), and 24.1% (95% CI 21.3%-27%), respectively (P<.001) (Table 2).

Table 2. Weighted prevalence of meeting muscle-strengthening exercise recommendations by different characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants meeting recommendations/total participants (n/N)</th>
<th>Prevalence (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age range (years)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≤13</td>
<td>2326/5594</td>
<td>41.4% (37.4%-45.5%)</td>
<td></td>
</tr>
<tr>
<td>14-15</td>
<td>3617/8575</td>
<td>42.2% (38.3%-46.2%)</td>
<td></td>
</tr>
<tr>
<td>≥16</td>
<td>3202/12,837</td>
<td>25.2% (23.3%-27.1%)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Boys</td>
<td>6067/13,933</td>
<td>43.8% (41.8%-45.8%)</td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>3078/13,073</td>
<td>24.3% (22%-26.6%)</td>
<td>.14</td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>3767/10,788</td>
<td>37% (32.9%-41.1%)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>5378/16,218</td>
<td>33.3% (30.8%-35.8%)</td>
<td></td>
</tr>
<tr>
<td>Type of school</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Middle school</td>
<td>5618/12,762</td>
<td>43.6% (40.3%-46.9%)</td>
<td></td>
</tr>
<tr>
<td>Academic high school</td>
<td>1845/7373</td>
<td>25.5% (23.2%-27.8%)</td>
<td></td>
</tr>
<tr>
<td>Vocational high school</td>
<td>1682/6871</td>
<td>24.1% (21.3%-27%)</td>
<td></td>
</tr>
</tbody>
</table>
Association Between Frequency of Engaging in MSE and Depression Symptoms

After adjusting for sociodemographic factors, behavioral lifestyle, academic performance, and physical fighting, engaging in MSE was inversely associated with the prevalence of depression in an apparent exposure-response manner. Compared to those who did not take part in MSE, the PRs for depression symptoms were 0.98 (95% CI 0.97-0.99), 0.95 (95% CI 0.93-0.97), 0.93 (95% CI 0.90-0.96), 0.90 (95% CI 0.87-0.94), 0.88 (95% CI 0.84-0.93), 0.86 (95% CI 0.81-0.92), and 0.84 (95% CI 0.78-0.90), respectively, for those performing MSE for 1 to 7 days/week (Table 3). In multiple linear regression analyses, the adjusted $\beta$ coefficient of the association between days of performing MSE and depression symptoms scores was $-0.05$ (95% CI $-0.08$ to $-0.02$). An increment of 1 day of MSE was associated with a 0.05 decrease in PHQ-9 score among the adolescents (Table S2 in Multimedia Appendix 1).

Table 3. Adjusted prevalence ratios (PRs) of depression symptoms associated with frequency of muscle-strengthening exercise among students.

<table>
<thead>
<tr>
<th>Frequency of muscle-strengthening exercise, (days/week)</th>
<th>$P$ value for trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Model 1: adjusted for age, gender, region, and type of school.
Model 2: adjusted for age, gender, region, type of school, parental education level, parental marital status, family income, cigarette smoking, alcohol drinking, physical activity, academic performance, and physical fighting.

**Sensitivity Analyses**

The sensitivity analysis (ie, additional adjustment for experience of being bullied) did not essentially change the association of engaging in MSE with depression symptoms (Table S3 in Multimedia Appendix 1).

**Discussion**

**Principal Findings**

To our knowledge, this is the first study examining the association between MSE frequency and depression in a provincially representative sample of Chinese middle and high school students. Our findings show that performing MSE was associated with a lower prevalence of depression symptoms among adolescents. In addition, our data also provide the latest information on the prevalence of meeting the MSE recommendations among Chinese adolescents.

**Prevalence of Meeting MSE Recommendations**

In this study, the overall prevalence of meeting MSE recommendations was 34.6%, much lower than in other studies performed in high-income countries and a previous study conducted in China. For instance, one nationally representative survey conducted in 2021 among 17,232 US students reported that 44.9% of high school students exercised $\geq 3$ days/week to strengthen or tone their muscles [10]. Another study conducted in 2013 and 2014 of 35,297 grade 9-12 Canadian students...
reported that 53.7% of students performed MSE ≥3 days per week [11]. One nationally representative survey implemented in 2019 of 56,636 Chinese middle and high school students found that 43.4% of middle school students and 28.1% of high school students met the WHO MSE recommendations [12]. Physical activity guidelines for Chinese people published in 2021 recommend that children and adolescents aged 6-17 years should engage in moderate- to vigorous-intensity physical activity at least 60 minutes daily and engage in MSE at least 3 days a week [39]. The Chinese government has formulated many regulations to ensure that there are at least 3 physical education (PE) classes per week for middle school students and at least 2 PE classes per week for high school students [40]. However, each 45-minute PE class in China comprises learning and training for basic sports skills, confrontational competitions, and muscle stretching and relaxing [40], and MSE was not specifically emphasized for PE classes. Hence, a closer focus should be placed on MSE, and consideration should be given to incorporating it into PE classes as an important component in the future.

Correlates of meeting the MSE recommendations observed in this study included household income, parental education level, and overall physical activity level. This is similar to findings from a report from Xin et al [12] showing that students who were from low-income households, had poorly-educated parents, and were physically inactive were less likely to meet the MSE recommendations. Consistent with this previous study [12], the prevalence of meeting MSE recommendations was lower among high school students than among middle school students. Also in line with previous studies [12,13], the prevalence of meeting MSE recommendations was lower among girls than among boys. In contrast to the study by Miller et al [41], which found a null significant association between cigarette smoking and meeting the MSE recommendations among 8383 adolescents aged 12-17 years, the more frequently students engaged in MSE, the more likely they were to smoke and drink in this study.

**Association Between Frequency of Engaging in MSE and Depression Symptoms**

Although an accumulating body of research has focused on the association of engaging in MSE with depression symptoms, the results remain conflicting and inconclusive. In this study, MSE frequency was inversely associated with depression symptoms among adolescents, which is consistent with previous findings. For example, in a cross-sectional study of 411,080 US adults aged ≥18 years without any moderate-to-vigorous intensity physical activity, engaging in MSE was significantly and inversely associated with the prevalence of clinically diagnosed depression, with a PR of 0.89 (95% CI 0.85-0.93) for MSE 3 times/week and 0.89 (95% CI 0.85-0.93) for MSE ≥4 times/week in comparison to those did not engage in MSE [21]. Another cross-sectional study of 601 Irish participants aged ≥18 years observed that higher frequency and intensity of MSE protected against depression [42]. However, results from a nationally representative cross-sectional study of 13,677 US high school students in grades 9-12 indicated no statistical significance in the association of engaging in MSE ≥3 days/week with depression symptoms [26]. Another study of 2088 US participants aged 20-39 years, with data retrieved from the National Health and Nutrition Examination Survey, demonstrated that it was physical activity, not MSE, that was independently associated with depressive symptoms [27].

Interestingly, several cohort studies have documented that handgrip strength was inversely associated with the incidence of depression [43-45]. For example, a UK Biobank prospective cohort study of 162,167 participants with a median follow-up of 10 years indicated that compared with participants in the highest tertile of handgrip strength, those in the medium and lowest tertiles had an 11% (hazard ratio [HR] 1.11, 95% CI 1.04-1.19) and 24% (HR 1.24, 95% CI 1.16-1.33) higher risk of depression, respectively [43]. Handgrip strength is often recognized as an indicator of overall muscle strength [46].

The underpinning mechanisms through which engaging in MSE could reduce the development of depression symptoms are largely unclear. First, engaging in MSE could improve muscle strength, which has been recognized as a health indicator and constituent of physical fitness, and enhance physical function and quality of life [20,47]. Second, exercise could increase feelings of self-efficacy and self-esteem, which may contribute to preventing mental health disorders [47,48].

The findings of this study are of practical public health importance and provide vital evidence that may inform the assessment of MSE and prevention of adolescent depression. First, only one-third of middle and high school students met the MSE recommendations, suggesting that insufficient MSE prevails among students of this age, and more comprehensive and effective efforts are needed to address these issues in China. These might include raising social awareness of the beneficial effect of engaging in MSE, incorporating MSE into PE classes, encouraging students to engage in MSE regularly, and evaluating the effects of MSE on physical and mental health among adolescents. Second, the American Heart Association updated a statement in September 2023 to clearly state that MSE is a safe, effective, and essential component of the overall physical activity regimen for cardiovascular disease risk reduction [49]. This study fills a gap in knowledge of the association of MSE frequency and depression symptoms among Chinese adolescents.

**Strengths and Limitations**

The strengths of this study included a large sample size, a high response rate, and a standardized procedure. The study also had several limitations. First, because the study was cross-sectional in design, engaging in MSE and depression symptoms cannot be temporally ordered. It is highly possible that bidirectional associations exist between MSE and depression, because adolescents who have depressive symptoms may be less likely to engage in MSE. Large prospective studies using genetic tools (eg, Mendelian randomization methods) may help with clarifying the association structure between these 2 factors. Second, we only collected the number of days of engaging in MSE in the past week and did not collect detailed information on intensity and duration of MSE. Third, all data were self-reported, and the findings presented may be susceptible to recall or social desirability biases.
Conclusions
In summary, our study sheds light on the association between MSE frequency and depression symptoms among middle and high school students in China. We found that the prevalence of meeting MSE recommendations was low, and that frequency of MSE was inversely associated with depression symptoms among students, although the exact direction of the association could not be determined. Once approved, it would be appropriate to incorporate MSE into methods for targeted prevention of adolescent depression.

Acknowledgments
We express our gratitude to all the students, parents, teachers, and local officials for their participation, assistance, and cooperation. This work was supported by Program of Zhejiang Federation of Humanities and Social Sciences (grant 2023B059) and Program of Zhejiang Youth Research (grant ZQ2023093).

Data Availability
The data sets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions
HW designed the study, drafted the manuscript, and analyzed the data. NL, YG, and JP collected the data. JZ and MY were involved in data interpretation. HD provided important comments on the manuscript and revised the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental figures and tables.
[DOCX File, 42 KB - publichealth_v10i1e50996_app1.docx ]

References
2. GBD Results. Institute For Health Metrics and Evaluation. URL: https://vizhub.healthdata.org/gbd-results/ [accessed 2024-04-03]


**Abbreviations**

- CDC: county center for disease control and prevention
- HR: hazard ratio
- MSE: muscle-strengthening exercise
- PE: physical education
- PHQ-9: Patient Health Questionnaire-9
PR: prevalence ratio
WHO: World Health Organization

© Hao Wang, Huaidong Du, Yunqi Guan, Jieming Zhong, Na Li, Jin Pan, Min Yu. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Health Literacy and Health Care System Confidence as Determinants of Attitudes to Vaccines in France: Representative Cross-Sectional Study

Georges Khoury¹, MPH; Jeremy K Ward², PhD; Julien Mancini³, Prof Dr, PhD; Amandine Gagneux-Brunon⁴, MD; Liem Binh Luong Nguyen¹, MD, PhD

¹CIC Cochin Pasteur, Hôpital Cochin, Assistance Publique—Hôpitaux de Paris, Paris, France
²Cermes3, CNRS, INSERM, Université Paris Cité, Paris, France
³Cancer, Biomedicine & Society Group, APHM, Hop Timone, Public Health Department (BIOSTIC), INSERM, IRD, ISSPAM, SESSTIM, Aix-Marseille University, Marseille, France
⁴Centre International de Recherche en Infectiologie, Team GIMAP, Inserm, U1111, CNRS, UMR530, CHU de Saint-Etienne, Saint-Etienne, France

Corresponding Author:
Georges Khoury, MPH
CIC Cochin Pasteur
Hôpital Cochin
Assistance Publique—Hôpitaux de Paris
19 Avenue de la Bourdonnais
Paris, 75007
France
Phone: 33 0610581889
Email: georxgeskhoury2000@hotmail.com

Abstract

Background: Health literacy involves individuals’ knowledge, personal skills, and confidence to take action to evaluate and appraise health-related information and improve their health or that of their community.

Objective: This study aimed to analyze the association between health literacy and attitude toward vaccines, adjusted with other factors.

Methods: We used the SLAVACO Wave 3, a survey conducted in December 2021 among a sample of 2022 individuals, representative of the French adult population. We investigated factors associated with the attitude toward vaccines using respondents’ different sociodemographic data, health literacy levels, and the health care system confidence levels using a multinomial logistic regression analysis.

Results: Among the participants, 440.4 (21.8%) were classified as “distrustful of vaccines in general,” 729.2 (36.1%) were “selectively hesitant,” and 852.4 (42.2%) were “nonhesitant.” In our model, the level of health literacy was not statistically different between the “distrustful of vaccines in general” and the “selectively hesitant” (P=.48), but it was associated with being a “nonhesitant” (adjusted odds ratio [aOR] 1.86, 95% CI 1.25-2.76). The confidence in the health care system was a strong predictor for a “nonhesitant” attitude toward vaccines (aOR 12.4, 95% CI 7.97-19.2). We found a positive correlation of 0.34 (P<.001) between health literacy and confidence in the health care system, but the interaction term between health literacy and health care system confidence was not significant in our model.

Conclusions: Health literacy was associated with a “nonhesitant” attitude toward vaccines. The findings demonstrated that health literacy and confidence in the health care system are modestly correlated. Therefore, to tackle the subject of vaccine hesitancy, the main focus should be on increasing the population’s confidence and on increasing their health literacy levels or providing vaccine information addressing the needs of less literate citizens.

(JMIR Public Health Surveill 2024;10:e45837) doi:10.2196/45837

KEYWORDS

vaccine hesitancy; health literacy; trust; attitude toward vaccines; public health; vaccination; COVID-19; adult; sociodemographic factor
Introduction

Vaccination is an essential component of primary health care, preventing over 20 potential fatal infectious diseases [1]. Vaccine hesitancy is defined as the refusal or hesitation in accepting vaccines when immunization services are available [2]. It has been designated by the World Health Organization [3] as one of the top 10 challenges in public health, underlining the need to research, understand, and tackle this global issue. It is dynamic and complex, affected by external factors such as time, location, and type of vaccine but also by individual factors such as complacency, convenience, and confidence [4,5].

Vaccine hesitancy is particularly strong in France, which is one of the most vaccine-hesitant countries in the world [6]. At the same time, France has reached a high level of vaccination coverage against COVID-19 [7]. During the COVID-19 vaccination campaign, many factors have been found to be strongly correlated with attitudes to vaccination in general and to vaccination against COVID-19, in particular, age, gender, complacency, political opinions, perception of one’s health status, trust in health authorities and professionals, and health literacy [8-10]. The ability of people to understand vaccine-related information as well as their understanding of medicine have focused much of the debates on the rise of vaccine hesitancy [11]. These observations resonate with the concept of “health literacy.” This term has emerged in the 1970s and refers to the motivation and skills used by individuals to access, understand, evaluate, and apply health information for their own health and that of their family and community; to make judgments and decisions about health care, prevention, and health promotion; and to maintain and promote quality of life throughout [12,13]. Health literacy includes abilities to evaluate and criticize health-related information; it involves a level of knowledge, personal skills, and confidence to take action to improve one’s personal health and the health of one’s community by changing one’s lifestyle and living conditions [14,15]. Low health literacy is directly associated with preventable undesirable health outcomes, including poorer general health, mortality, and inadequate decisions for preventive measures [16,17]. It has also been associated with a higher risk of being “hesitant” rather than “provaccination” [18].

Because health literacy bears heavily on people’s ability to access medical information and health care providers as well as their treatment of this information, it stands to reason that it constitutes a significant determinant of attitudes to vaccinations. However, very few papers have empirically investigated this relationship, which is far from straightforward [19]. Indeed, health literacy interplays with the components of the “3C” model, as described earlier.

Most studies assess functional health literacy mainly using tools like S-TOFHLA (Short Test of Functional Health Literacy in Adults), thereby focusing only on the ability to understand medical information [20,21]. This is restrictive compared with more recent definitions of health literacy, which include the ability to discuss and ultimately make decisions to promote one’s health, lacking the use of rigorously validated scales as we will in our research. This study aims to contribute comprehensively to the understanding of the relationship between health literacy and vaccine hesitancy. In this paper, we will not only investigate the role of health literacy but also the importance of trust in the health care system and their combined impact on vaccine hesitancy. By examining these factors simultaneously, we aim to offer a more holistic perspective on the phenomenon of vaccine hesitancy, ultimately contributing to more effective public health interventions and strategies.

Methods

Recruitment Procedure and Study Sample

Between December 2 and December 17, 2021, invitations to participate in the study were sent to 25,800 French adults, randomly selected from an digital panel of more than 700,000 respondents (provided by Bilendi [22]). The self-administered web-based questionnaires, lasting approximately 15 minutes, were completed during this period. Participants received remuneration in the form of points for completing these surveys. Ultimately, participants could exchange the accumulated points for gift cards, further acknowledging their invaluable contributions to the survey. We then used the quota sampling method to finally obtain a study sample of 2022 respondents matching the French mainland adult population in terms of age (18-24, 25-34, 35-49 50-64, 75+ years), sex (male and female), occupation (farmers, craftsmen, executives, intermediate professions, employees, workers, retirees, and other inactives), population size of the area of residence (<2000, 2000-20,000, 20,000-100,000, and >100,000 inhabitants), and region (Alsace, Aquitaine, Auvergne, Burgundy, Brittany, Center, Île-de-France, Languedoc, Nord-Pas-de-Calais, Normandy, Pays de la Loire, and Provence-Alpes-Côte d’Azur). A weighting procedure was applied to further match the sample to these characteristics when the quota was not perfectly met. To do so, we used the raking ratio method using the SAS Calmar macro developed in France by the National Institute of Statistics and Economic Studies.

Data Collected and Outcome

After the participants’ consent was obtained, we collected information on respondents’ sociodemographic characteristics such as age, sex, educational attainment, and income, as well as their attitudes and practices on a number of issues including vaccination, politics, alternative medicine, and trust in various institutions. Our main outcome is the attitude of the French adult population toward vaccines, using a widely recognized typology for assessing attitudes toward vaccines in France [23,24]. We made a typology in 3 categories based on the answers to 5 questions with the same format, asking whether the responders were in favor of (1) vaccines in general, (2) the flu vaccine, (3) the hepatitis B virus (HBV) vaccine, (4) the human papillomavirus (HPV) vaccine, and (5) the measles vaccine. The categories were built as follows:

- Category 1 (nonhesitant): All respondents who answered favorably to all 5 questions (ie, individuals who are favorable for vaccines in general and 100% accepting all types of vaccines).
- Category 2 (provaccine but selectively hesitant): All respondents who answered “favorable” to the question on...
vaccines in general but who answered “unfavorable” or “I do not know” to at least 1 of the 4 questions on specific vaccines. These represent the people who are favorable to vaccines in general but have doubts or reservations about specific vaccines.

• Category 3 (distrustful of vaccines in general): All respondents who answered “unfavorable” to vaccines in general.

Assessment of Health Literacy Level and Health Care System Confidence Level

We calculated 2 scores: a health literacy score (HLS_{19-Q12}) and a confidence score. The health literacy score was based on a series of 12 questions with a 4-point Likert scale on the ability of each participant to understand, evaluate, and make health decisions in their everyday life using a 48-point score (Multimedia Appendix 1). A score between 4 and 48 was obtained summing the 12 responses, with a higher score indicating a higher health literacy level. We divided the population into quartiles for the analysis: “very high” for a score >36, “high” for a score between 32 and 36, “low” for a score between 26 and 31, and “very low” for a score <26. The health care system confidence score was based on a series of 5 questions about the confidence in science, government agencies that monitor health and environmental risks, government, physicians, and drug manufacturers using a 20-point score with a 4-point scale (Multimedia Appendix 2). A higher score indicates a higher confidence level. The confidence levels score was categorized as follows based on a quartile division: “confident” for a score ≥19, “somewhat confident” for a score between 17 and 19, “somewhat not confident” for a score between 13 and 16, and “not confident” for a score <13. To estimate the internal validity of our scores, we calculated the Cronbach coefficient, in which a result of >0.9 for the health literacy score and the confidence score was considered as a strong internal validity criterion.

Statistical Analysis

Participants’ baseline characteristics were described for each level of vaccine hesitancy. We used chi-square test with Rao and Scott’s second-order correction in cross-tabulations, a univariate regression model to select statistically significant variables by a forward stepwise selection method (entry threshold $P<.2$), and a multinomial logistic regression to investigate factors associated with the attitude of the French adult population toward vaccines using respondents’ different background data, health literacy levels, and confidence levels. Our model was adjusted on gender, age, level of education, income, health care system confidence level, and health literacy level. Because of weighting, frequency counts including decimal points, odds ratios, and adjusted odds ratio (aOR) were reported with 95% CI. We also tested the first-order interaction between health literacy and confidence level to uncover hidden or conditional relationships in data that would not be apparent when looking at individual variables in isolation. A further sensitivity analysis was conducted based on our model excluding health care system confidence.

Pearson correlation tests were used to test the relationship between health literacy and health care system confidence to assess whether there is an association or linear relationship between them, which also provides information about the strength and direction of the association. This test will help us identify whether these 2 factors are positively, negatively, or not significantly associated with each other.

Ethical Considerations

Informed consent to participate in this study (including analyses presented in this paper) was collected before the completion of the questionnaire. Researchers did not have access to identifying data on participants following standard practices for web-based surveys. Participants received points after completing the surveys. Ultimately, participants could exchange all the points they received as gift cards. The methodology of the study was reviewed and approved by the ethical committee of the INSERM (IRB00003888, #21-770).

Results

Characteristics of the Population

In our study population, 208.1 (10.3%) were young adults, 1276.3 (63.1%) were 25-64 years of age, and 537.4 (26.6%) were 64 years and older with a sex ratio of 91 male participants per 100 female participants. Regarding the level of education, 1322.5 (65.4%) achieved secondary education, and 743.4 (36.7%) earned less than €2000 per month (a currency exchange rate of €1=US $1.183 is applicable). Baseline characteristics of the study population, overall and by the level of vaccine hesitancy, are described in Table 1.
## Table 1. Baseline demographics of study participants (SLAVACO Wave 3, 2021).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N=2022), n (%)</th>
<th>Distrustful of vaccines in general (n=440.4), n (%)</th>
<th>Selectively hesitant (n=729.2), n (%)</th>
<th>Nonhesitant (n=852.4), n (%)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>18-24</td>
<td>208.1 (10.3)</td>
<td>56.2 (12.7)</td>
<td>56.84 (7.8)</td>
<td>95.2 (11.2)</td>
<td>.001</td>
</tr>
<tr>
<td>25-34</td>
<td>295.9 (14.6)</td>
<td>87.6 (19.9)</td>
<td>98.1 (13.4)</td>
<td>110.3 (12.9)</td>
<td>.001</td>
</tr>
<tr>
<td>35-49</td>
<td>487.5 (24.1)</td>
<td>132.8 (30.2)</td>
<td>156.7 (21.5)</td>
<td>198.0 (23.2)</td>
<td>.001</td>
</tr>
<tr>
<td>50-64</td>
<td>492.9 (24.4)</td>
<td>94.6 (21.5)</td>
<td>198.9 (27.3)</td>
<td>199.5 (23.4)</td>
<td>.001</td>
</tr>
<tr>
<td>65-74</td>
<td>290.3 (14.4)</td>
<td>41.6 (9.4)</td>
<td>117.9 (16.2)</td>
<td>130.8 (15.3)</td>
<td>.001</td>
</tr>
<tr>
<td>75+</td>
<td>247.1 (12.2)</td>
<td>27.7 (6.3)</td>
<td>100.8 (13.8)</td>
<td>118.6 (13.9)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>962.5 (47.6)</td>
<td>193.5 (43.9)</td>
<td>319.0 (43.7)</td>
<td>450.1 (52.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female</td>
<td>1059.5 (52.4)</td>
<td>246.9 (56.1)</td>
<td>410.2 (56.3)</td>
<td>402.3 (47.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Diploma</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No degree</td>
<td>699.5 (34.6)</td>
<td>197.8 (44.9)</td>
<td>242.5 (33.3)</td>
<td>259.2 (30.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Baccalaureate</td>
<td>519.3 (25.7)</td>
<td>112.9 (25.6)</td>
<td>196.2 (26.9)</td>
<td>210.3 (24.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Bachelor</td>
<td>569.3 (28.2)</td>
<td>102.7 (23.3)</td>
<td>200.4 (27.5)</td>
<td>266.2 (31.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>233.9 (11.6)</td>
<td>27.0 (6.1)</td>
<td>90.1 (12.4)</td>
<td>116.8 (13.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Monthly salary (€)</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Less than €1000</td>
<td>189.2 (9.4)</td>
<td>46.4 (10.5)</td>
<td>65.8 (9)</td>
<td>77.0 (9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>€1000-€1500</td>
<td>250.6 (12.4)</td>
<td>75.6 (17.2)</td>
<td>75.9 (10.4)</td>
<td>99.1 (11.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>€1500-€2000</td>
<td>303.3 (15)</td>
<td>75.6 (17.2)</td>
<td>104.2 (14.3)</td>
<td>123.5 (14.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>€2000-€3000</td>
<td>461.7 (22.8)</td>
<td>88.7 (20.1)</td>
<td>197.4 (27.1)</td>
<td>175.7 (20.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>€3000-€4000</td>
<td>330.9 (16.4)</td>
<td>52.3 (11.9)</td>
<td>134.6 (18.5)</td>
<td>141.1 (16.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>More than €4000</td>
<td>249.3 (12.3)</td>
<td>30.6 (6.9)</td>
<td>74.1 (10.2)</td>
<td>144.7 (17)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Health literacy level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Very high (36+)</td>
<td>532.8 (26.3)</td>
<td>87.9 (20)</td>
<td>169.4 (23.2)</td>
<td>275.4 (32.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High (32-36)</td>
<td>457.1 (22.6)</td>
<td>77.4 (17.6)</td>
<td>166.6 (22.8)</td>
<td>213.1 (25)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Low (26-31)</td>
<td>519.4 (25.7)</td>
<td>108.5 (24.6)</td>
<td>197.7 (27.1)</td>
<td>213.2 (25)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Very low (&lt;26)</td>
<td>512.7 (25.3)</td>
<td>166.6 (37.8)</td>
<td>195.5 (26.8)</td>
<td>150.6 (17.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Health care system confidence level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Confident (19+)</td>
<td>571.6 (28.3)</td>
<td>41.7 (9.5)</td>
<td>183.2 (25.1)</td>
<td>346.7 (40.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Somewhat confident (17-19)</td>
<td>322.3 (15.9)</td>
<td>42.8 (9.7)</td>
<td>120.0 (16.4)</td>
<td>159.5 (18.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Somewhat not confident (13-16)</td>
<td>684.6 (33.9)</td>
<td>170.6 (38.7)</td>
<td>259.9 (35.6)</td>
<td>254.1 (29.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not confident (&lt;13)</td>
<td>443.5 (21.9)</td>
<td>185.2 (42.1)</td>
<td>166.1 (22.8)</td>
<td>92.2 (10.8)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>Chi-square test with Rao and Scott’s second-order correction.

<sup>b</sup>A currency exchange rate of €1=US $1.183 is applicable.

### Pattern of Vaccine Hesitancy in the Population

Among the 2022 participants, 852.4 (42.2%) were “nonhesitant,” 729.2 (36.1%) were “selectively hesitant” (among which, n=678.2, 93% were not favorable to 1 or 2 vaccines), and 440.4 (21.8%) were “distrustful of vaccines in general.” The “nonhesitant” group was older with a higher male or female sex ratio. Regarding health literacy score, the “nonhesitant” group had the highest proportion 275.4 (32.3%) with a “very high” score, whereas the “distrustful of vaccines in general” group had the highest proportion 166.6 (37.8%) with a “very low” score. Regarding confidence scores, the “nonhesitant” group had the highest proportion 346.7 (40.7%) of “confident” (>19/24), whereas the “distrustful of vaccines in general” group had the highest proportion 185.2 (42.1%) of “not confident.”
Of note, a large majority of “distrustful of vaccines in general” individuals 355.8 (80.8%) belong to either “somewhat not confident” or “not confident” categories.

Factors Associated With Vaccine Hesitancy

We examined our multinomial logistic regression model factors associated with attitude toward vaccination taking the “distrustful of vaccines in general” group as a reference (Table 2). The sex ratio of male and female participants was not significantly different with the “selectively hesitant” group but was higher among the “nonhesitant” with an aOR of 1.43 (95% CI 1.09-1.88). Age was higher in the “selectively hesitant” population, and the aOR varied from 2.72 (95% CI 1.72-4.29) for the 50- to 64-year age group to 4.01 (95% CI 1.94-8.27) for the 75+-year age group. In the “nonhesitant” group, aOR passed from 1.93 (95% CI 1.16-3.20) for the 50- to 64-year age category to 3.28 (95% CI 1.57-6.86) for the 75+-year category. No association was found for the younger age group. The level of education was also associated with vaccine hesitancy: participants having a postgraduate degree yielded an aOR of 2.63 (95% CI 1.58-4.40) in the “selectively hesitant” group and an aOR of 2.21 (95% CI 1.33-3.68) in the “nonhesitant” group, taking the “distrustful of vaccines in general” as reference. Confidence was a strong predictor of the attitude toward vaccination in both categories, with a higher confidence score leading to a higher aOR. In the “selectively hesitant” category, aOR varied from 1.60 (95% CI 1.16-2.20) for “somewhat not confident” level to 4.13 (95% CI 2.64-6.46) for the “confident” level. In the “nonhesitant” group, aOR varied from 2.77 (95% CI 1.91-4.01) for the “somewhat not confident” level to 12.4 (95% CI 7.97-19.2) for the “confident” level. While the health literacy levels did not appear statistically significant in the “selectively hesitant” group, a statistical difference was observed in the “nonhesitant” group compared to the “distrustful of vaccines in general” group: aOR increased from 1.50 (95% CI 1.04-2.16) for the “low” level of health literacy to 1.69 (95% CI 1.14-2.50) for the “high” level of health literacy and eventually 1.86 (95% CI 1.25-2.76) for the “very high” level of health literacy.
Table 2. Multinomial logistic regression results: exploring determinants of vaccine attitudes and hesitancy in France (SLAVACO Wave 3, 2021; N=2022).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Selectively hesitant</th>
<th>Nonhesitant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aOR (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.38 (0.20-0.74)</td>
<td>.005 b</td>
</tr>
<tr>
<td></td>
<td>0.19 (0.09-0.38)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Male</td>
<td>0.99 (0.73-1.36)</td>
<td>&gt;.88</td>
</tr>
<tr>
<td></td>
<td>1.43 (1.09-1.88)</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>25-34</td>
<td>1.20 (0.73-1.97)</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>0.98 (0.56-1.71)</td>
<td>&gt;.90</td>
</tr>
<tr>
<td>35-49</td>
<td>1.40 (0.85-2.30)</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>1.26 (0.76-2.12)</td>
<td>.39</td>
</tr>
<tr>
<td>50-64</td>
<td>2.72 (1.72-4.29)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>1.93 (1.16-3.20)</td>
<td>.01</td>
</tr>
<tr>
<td>65-74</td>
<td>3.05 (1.73-5.38)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>2.25 (1.23-4.14)</td>
<td>.01</td>
</tr>
<tr>
<td>75+</td>
<td>4.01 (1.94-8.27)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>3.28 (1.57-6.86)</td>
<td>.002</td>
</tr>
<tr>
<td><strong>Degree</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No degree</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Baccalaureate</td>
<td>1.59 (1.10-2.31)</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>1.56 (1.08-2.24)</td>
<td>.02</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>1.67 (1.18-2.37)</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>1.82 (1.27-2.61)</td>
<td>.001</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>2.63 (1.58-4.40)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>2.21 (1.33-3.68)</td>
<td>.003</td>
</tr>
<tr>
<td><strong>Monthly salary (€)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than €1000</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>€1000-€1500</td>
<td>0.62 (0.34-1.13)</td>
<td>.12</td>
</tr>
<tr>
<td>€1500-€2000</td>
<td>0.76 (0.45-1.31)</td>
<td>.29</td>
</tr>
<tr>
<td>€2000-€3000</td>
<td>1.04 (0.61-1.75)</td>
<td>.87</td>
</tr>
<tr>
<td>€3000-€4000</td>
<td>1.09 (0.61-1.87)</td>
<td>.83</td>
</tr>
<tr>
<td>More than €4000</td>
<td>0.83 (0.43-1.61)</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td>1.22 (0.65-2.32)</td>
<td>.52</td>
</tr>
<tr>
<td><strong>Health care system confidence level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not confident (&lt;13)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Somewhat not confident (13-16)</td>
<td>1.60 (1.16-2.20)</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>2.77 (1.91-4.01)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Somewhat confident (17-19)</td>
<td>2.88 (1.77-4.71)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>6.51 (4.13-10.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Confident (19+)</td>
<td>4.13 (2.64-6.46)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>12.4 (7.97-19.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Health literacy level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low (&lt;26)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Low (26-31)</td>
<td>1.22 (0.84-1.77)</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td>1.50 (1.04-2.16)</td>
<td>.03</td>
</tr>
<tr>
<td>High (32-36)</td>
<td>1.27 (0.85-1.90)</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>1.69 (1.14-2.50)</td>
<td>.009</td>
</tr>
<tr>
<td>Very high (36+)</td>
<td>1.13 (0.77-1.67)</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>1.86 (1.25-2.76)</td>
<td>.003</td>
</tr>
</tbody>
</table>

aOR: adjusted odds ratio.
bValues in italics format are statistically significant with a P<.05.
cA currency exchange rate of €1=US $1.183 is applicable.

The robustness of these results was tested by conducting a multivariable regression model for each type of vaccine included in the questionnaire (vaccine in general, flu vaccine, HPV vaccine, HBV vaccine, and measles vaccine) with the same covariables and found that for each vaccine, health care system confidence was highly associated with a positive attitude toward vaccine (P<.001), but health literacy was no longer associated with a positive attitude toward each vaccine (P>.05) with no interaction term between these 2 variables (P>.10). The results of the Pearson correlation coefficient showed a modest positive
correlation of 0.34 (P<.001) between health literacy and the health care system confidence.

To explore these findings, our multivariable regression model for each of the vaccination types listed in the questionnaire (vaccine in general, flu vaccine, HPV vaccine, HBV vaccine, and measles vaccine) showed that in all of the regression models, vaccine acceptance for each of the vaccines mentioned was strongly associated with the confidence level but not with health literacy with no interaction between health literacy and confidence (P value of interaction >.1). We found a modest correlation (Pearson correlation coefficient=0.34) between health literacy and confidence. Another multinomial logistic regression was performed by removing the health care system confidence indicator from our model (Table 3). In this model, the effect of health literacy on vaccine hesitancy appeared to be significant (P<.001) with higher aOR.

Table 3. Sensitivity analysis on multinomial logistic regression model: exploring determinants of vaccine attitudes and hesitancy in France (SLAVACO Wave 3, 2021; N=2022) with the exclusion of the health care system confidence variable.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Selectively hesitant</th>
<th></th>
<th>Nonhesitant</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>aOR b (95% CI)</td>
<td>P value</td>
<td>aOR (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td></td>
<td>0.63 (0.34-1.16)</td>
<td>.13</td>
<td>0.54 (0.31-0.95)</td>
<td>.03 b</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Male</td>
<td>0.98 (0.77-1.26)</td>
<td>.87</td>
<td>1.43 (1.11-1.83)</td>
<td>.006</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>25-34</td>
<td>0.98 (0.60-1.60)</td>
<td>&gt;.90</td>
<td>0.63 (0.42-1.08)</td>
<td>.10</td>
</tr>
<tr>
<td>35-49</td>
<td>1.18 (0.75-1.84)</td>
<td>.54</td>
<td>0.94 (0.60-1.49)</td>
<td>.78</td>
</tr>
<tr>
<td>50-64</td>
<td>2.33 (1.48-3.66)</td>
<td>&lt;.001</td>
<td>1.49 (0.99-2.23)</td>
<td>.05</td>
</tr>
<tr>
<td>65-74</td>
<td>2.87 (1.55-5.29)</td>
<td>.001</td>
<td>2.03 (1.15-3.61)</td>
<td>.02</td>
</tr>
<tr>
<td>&gt;=75</td>
<td>3.60 (2.01-6.48)</td>
<td>&lt;.001</td>
<td>2.76 (1.54-4.95)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Degree</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No degree</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Baccalaureate</td>
<td>1.53 (1.04-2.25)</td>
<td>.03</td>
<td>1.44 (1.02-2.05)</td>
<td>.04</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>1.73 (1.21-2.46)</td>
<td>.003</td>
<td>1.92 (1.34-2.75)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Postgraduate</td>
<td>2.83 (1.58-5.08)</td>
<td>&lt;.001</td>
<td>2.54 (1.45-4.42)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Monthly salary (€)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than €1000</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>€1000-€1500</td>
<td>0.61 (0.34-1.08)</td>
<td>.08</td>
<td>0.70 (0.38-1.27)</td>
<td>.20</td>
</tr>
<tr>
<td>€1500-€2000</td>
<td>0.75 (0.45-1.26)</td>
<td>.32</td>
<td>0.82 (0.47-1.44)</td>
<td>.51</td>
</tr>
<tr>
<td>€2000-€3000</td>
<td>1.09 (0.68-1.76)</td>
<td>.68</td>
<td>0.85 (0.51-1.41)</td>
<td>.54</td>
</tr>
<tr>
<td>€3000-€4000</td>
<td>1.11 (0.65-1.91)</td>
<td>.70</td>
<td>0.96 (0.58-1.62)</td>
<td>.88</td>
</tr>
<tr>
<td>More than €4000</td>
<td>0.87 (0.43-1.78)</td>
<td>.66</td>
<td>1.36 (0.90-2.66)</td>
<td>.41</td>
</tr>
<tr>
<td><strong>Health literacy level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very low (&lt;26)</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Low (26-31)</td>
<td>1.42 (1.01-1.99)</td>
<td>.04</td>
<td>1.98 (1.40-2.81)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High (32-36)</td>
<td>1.74 (1.16-2.60)</td>
<td>.008</td>
<td>2.92 (1.97-4.35)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Very high (36+)</td>
<td>1.57 (1.10-2.24)</td>
<td>.01</td>
<td>3.33 (2.35-4.73)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aOR: adjusted odds ratio.
bValues in italics format are statistically significant with a P<.05.
cA currency exchange rate of €1=US $1.183 is applicable.
Discussion

Principal Findings

This study confirms a relationship between vaccine hesitancy and health literacy in this study sample, representative of the French population, in which only 852.4 (42.2%) are “nonhesitant.” In a regression model looking at factors associated with vaccine hesitancy, we found that higher age, male sex, health care system confidence, health literacy, and educational attainment were independently associated with less vaccine hesitancy. Health care system confidence was notably a strong predictor: a person with the highest level of confidence is 12 times less likely to be “distrustful of vaccines in general” and 4 times less likely to be “selectively hesitant.”

The level of vaccine hesitancy in this study sample is consistent with other studies conducted in France [23,24]. Our regression model confirms the “confidence, complacency, constrain” (so-called “3C”) model developed by the Strategic Advisory Group of Experts on Immunization Working Group [2]. However, confidence is multileveled: it includes confidence in health care professionals but also in the health care system, and the national state. This study specifically addressed confidence in the health care system. While primary care providers are often among the most trusted, confidence in the state and policy makers can be much more difficult to build or regain. In this study, we found a widespread mistrust of government and the pharmaceutical industry that logically translates into mistrust of vaccines (n=355.8, 80% of the “distrustful of vaccines in general” participants belonged to the nonconfident categories).

We also looked at the role of health literacy in vaccine confidence. A recent systematic review that looked at the association of health literacy with intention to vaccinate and vaccination status found high heterogeneity in the assessment of health literacy, inconsistent results, or a weak association when analyzing 21 papers [19]. Studies have hypothesized a mediating role of health literacy on the relationship between health care system distrust and vaccine hesitancy [25], while in other context, researchers studied, for example, the potential mediating effect of trust in one’s physician on the association of health literacy and medication adherence [26]. This could reflect the fact that health literacy is often associated with other properties that are also likely to affect attitudes to vaccines such as mistrust of actors in vaccination policies, but it could also reflect the paradoxical effect that health literacy is likely to have on attitudes to vaccines. Indeed, health literacy can both help to navigate the health care system and gain access to medical professionals who provide evidence-based information, but it can also favor “doing your own research” and coming across vaccine-critical information. This potentially paradoxical effect of the ability and propensity to look for information regarding health has been well documented by work on the relationship between healthism—the tendency for individuals to exercise control over their own choices in order to maximize their life expectancy—and vaccine hesitancy [27]. Our results contribute to these reflections. We found that a higher health literacy level is associated with being “nonhesitant,” whereas it was not statistically significant in selective hesitancy when controlling for confidence in health care actors.

The relationship between confidence in the health care system and health literacy was also examined. We found that these 2 factors were independent in our model (no interactions found) but modestly correlated. The correlation between the 2 variables seems to be less strong in France than in other countries. For instance, a cross-sectional survey conducted in Taiwan to measure the association between health literacy and trust in physicians and health care systems with a population of 2199 adults showed that respondents’ level of confidence in medical professionals and the health care system was higher among those with better health literacy [28]. After adjusting for respondents’ sociodemographic factors, health literacy remained significantly and positively associated with confidence. Our results therefore suggest that literacy in itself might be less important in France compared to trust. This could reflect the fact that distrust of public actors is particularly strong in France. Comparative work could shed light on national differences in the relationship between trust in health care actors and health literacy. The relationship between the 2 is likely to be affected—among other factors—by the organization of health care and how social inequalities are reflected in health. The absence of interaction effect between health literacy and trust in health care actors is also interesting. We could expect that a higher level of literacy combined with a higher distrust of health care actors would favor the tendency to find and appreciate vaccine-critical sources. It is possible that these tendencies are counterbalanced by a greater propensity among the more literate to identify the public signals denouncing the antivaccine rhetoric as antiscientific, which tend to be pervasive in the mainstream media, especially in France [29,30]. One possible avenue of research in future work could be to explore the relationship between health literacy and information-seeking practices on social media. Inspiration could come from research on how political attitudes interact with social media use and with educational attainment in the inception of vaccine attitudes (for reviews of the literature, see [31,32]).

We found that the male sex was associated with the “nonhesitant” group (aOR 1.43, 95% CI 1.09-1.88). These findings confirm that male sex had more often positive attitudes toward vaccines, whereas women are more likely to be hesitant. These findings are in line with previous studies on vaccine hesitancy [33-35].

Level of education and educational attainment were found to be an independent factor on vaccine acceptance after adjusting for health literacy and health care system confidence. Finally, in our model, the level of income did not appear to be statistically significant with vaccine acceptability. Despite not achieving statistical significance in the multinomial regression model, it is worth noting that the estimated aORs for vaccine hesitancy showed a trend to enhance vaccine acceptance with a salary in the range of €2000-4000 per month for the “selectively hesitant” group and for a salary of more than €4000 per month for the “nonhesitant” group. These results confirm prior study results, which indicates that families in the lowest income categories in Brazil, which account for one-third of the

https://publichealth.jmir.org/2024/1/e45837

JMRI Public Health Surveill 2024 | vol. 10 | e45837 | p.1258
(page number not for citation purposes)
population, are reported to be the most vaccine hesitant [36]. Given that those in lower income brackets (income <€1500 per month) make up about 15% of the French population in 2018 [37] and showed the lowest vaccination acceptance rates (23.05%) and the greatest vaccine hesitancy rates (33.04%) [38].

Limitations
This study also has several limitations: health literacy is complex. Our questionnaires using close-ended questions may not capture the whole dimensions of this concept. For example, our questionnaire could not separate health literacy from scientific literacy. The same limitation applies for confidence. These limits could only be resolved using qualitative methods that could examine how health literacy and confidence translate in reality with vaccine providers. The second limitation of this study was the recruitment procedure, which uses quota sampling and can lead to more biases than random sampling. However, the frequency of vaccine hesitancy appeared to be consistent with the frequency previously reported in the literature.

Strengths
Finally, this study has several strengths: the study population was representative of the French population, which has one of the highest vaccine hesitancy rates; we used a validated tool to assess vaccine hesitancy; and the large sample size of this study allows to examine various factors that have been separately associated with vaccine hesitancy in previous studies as well as different vaccines.

Conclusions
Interventions aiming to reduce vaccine hesitancy should prioritize confidence-building measures, which are difficult. Improving health literacy may also help, and public health interventions demonstrated their ability to increase health literacy. We believe that there is a possibility that improving health literacy can also lead to enhanced confidence in the health care system in the long run. Public health campaigns and educational programs should enhance health literacy, while health care providers engage in open, empathetic vaccine discussions. Community leaders and organizations play a role in promoting vaccination. Further research on the relationship between confidence and health literacy is essential. Transparency, timely updates, and public-private partnerships are vital. Tailored interventions should address gender disparities and involve continuous monitoring. Fostering understanding, trust, and tailored strategies can enhance vaccine acceptance and coverage in France.

Acknowledgments
This work was supported by grants from the Agence Nationale de la Recherche (ANR-20-COVI-0035-01) and the Agence Nationale de la Recherche Scientifique—Maladie Infectieuses Emergentes (Comité ad-hoc de pilotage national des essais thérapeutiques et autres recherches sur la COVID-19 [CAPNET] project 0344). The funding sources had no role in the design of the study, analysis of the data, or writing of the paper. This material is the authors’ own original work, which has not been previously published elsewhere.

Data Availability
The data sets analyzed during this study are available in the Harvard Dataverse repository [39].

Conflicts of Interest
None declared.

Multimedia Appendix 1
Health literacy questionnaire (HLS19-Q12).
[DOCX File, 14 KB - publichealth_v10i1e45837_app1.docx ]

Multimedia Appendix 2
Health care system confidence questionnaire.
[DOCX File, 13 KB - publichealth_v10i1e45837_app2.docx ]

References


Abbreviations

- **aOR:** adjusted odds ratio
- **HBV:** hepatitis B virus
- **HPV:** human papillomavirus
- **S-TOFHLA:** Short Test of Functional Health Literacy in Adults

©Georges Khoury, Jeremy K Ward, Julien Mancini, Amandine Gagneux-Brunon, Liem Binh Luong Nguyen. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 07.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Multimorbidity and its Associated Factors in Korean Shift Workers: Population-Based Cross-Sectional Study

Hye Chong Hong¹, MSN, PhD; Young Man Kim²,³,⁴, MSN, PhD

¹Department of Nursing, Chung-Ang University, Seoul, Republic of Korea
²College of Nursing, Jeonbuk National University, Jeonju, Republic of Korea
³Research Institute of Nursing Science, Jeonbuk National University, Jeonju, Republic of Korea
⁴Biomedical Research Institute, Jeonbuk National University Hospital, Jeonju, Republic of Korea

Corresponding Author:
Young Man Kim, MSN, PhD
College of Nursing
Jeonbuk National University
567 Baekje-daero
Deokjin-gu
Jeonju, 54896
Republic of Korea
Phone: 82 10 3498 5078
Fax: 82 63 270 3127
Email: ymk@jbnu.ac.kr

Abstract

Background: Multimorbidity is a crucial factor that influences premature death rates, poor health, depression, quality of life, and use of health care. Approximately one-fifth of the global workforce is involved in shift work, which is associated with increased risk for several chronic diseases and multimorbidity. About 12% to 14% of wage workers in Korea are shift workers. However, the prevalence of multimorbidity and its associated factors in Korean shift workers are rarely reported.

Objective: This study aimed to assess multimorbidity prevalence, examine the factors associated with multimorbidity, and identify multimorbidity patterns among shift workers in Korea.

Methods: This study is a population-based cross-sectional study using Korea National Health and Nutrition Examination Survey data from 2016 to 2020. The study included 1704 (weighted n=2,697,228) Korean shift workers aged 19 years and older. Multimorbidity was defined as participants having 2 or more chronic diseases. Demographic and job-related variables, including regular work status, average working hours per week, and shift work type, as well as health behaviors, including BMI, smoking status, alcohol use, physical activity, and sleep duration, were included in the analysis. A survey-corrected logistic regression analysis was performed to identify factors influencing multimorbidity among the workers, and multimorbidity patterns were identified with a network analysis.

Results: The overall prevalence of multimorbidity was 13.7% (302/1704). Logistic regression indicated that age, income, regular work, and obesity were significant factors influencing multimorbidity. Network analysis results revealed that chronic diseases clustered into three groups: (1) cardiometabolic multimorbidity (hypertension, dyslipidemia, diabetes, coronary heart disease, and stroke), (2) musculoskeletal multimorbidity (arthritis and osteoporosis), and (3) unclassified diseases (depression, chronic liver disease, thyroid disease, asthma, cancer, and chronic kidney disease).

Conclusions: The findings revealed that several socioeconomic and behavioral factors were associated with multimorbidity among shift workers, indicating the need for policy development related to work schedule modification. Further organization-level screening and intervention programs are needed to prevent and manage multimorbidity among shift workers. We also recommend longitudinal studies to confirm the effects of job-related factors and health behaviors on multimorbidity among shift workers in the future.

(JMIR Public Health Surveill 2024;10:e55014) doi:10.2196/55014
KEYWORDS
chronic disease; multimorbidity; shift work schedule; shift workers; population-based study; Korea; network analysis; logistic regression; cross-sectional study; public health

Introduction

Shift work includes any work schedule that is outside of the conventional 7 AM to 6 PM working hours. Shift work is prevalent and inevitable in some workforces, including healthcare, law enforcement, and manufacturing [1]. Globally, a total of approximately one-fifth of the workforce is involved in shift work [2]. While shift work allows workplace flexibility and may provide economic benefits, it may also be associated with adverse chronic health outcomes [3].

Shift work is known to disrupt circadian rhythms and affect sleep patterns, hormone secretion, and other biological processes [4]. Furthermore, these disruptions have been associated with chronic diseases among shift workers, such as metabolic disorders, diabetes, cardiovascular disease, stroke, cancer, and depression [5,6]. However, most epidemiological studies and systematic reviews examining the relationships between shift work and chronic diseases only examined the effect of shift work on a single chronic disease, despite the potential for chronic disease comorbidities. Few studies have examined the coexistence of chronic disease and shift work. For example, Yang et al [7] examined shift work and the risk of cardiometabolic multimorbidity among patients with hypertension and found that shift work was associated with cardiometabolic multimorbidity. In the Korean population, shift work was found to be associated with mental health problems, such as depression and suicide ideation in electronics workers [8], chronic kidney disease in female manual laborers [9], and metabolic syndrome in female workers [10].

Multimorbidity refers to the existence of 2 or more chronic diseases in an individual [11,12]. Unlike comorbidity, which refers to the combined effects of chronic diseases related to a primary chronic disease, multimorbidity examines all chronic diseases simultaneously, which means that no single condition is more important than any other [13]. Multimorbidity is person centered and does not assign priority to a single condition [14]. Worldwide, approximately 37% of the general population has multimorbidity [15], and it is associated with premature death [16], poor health [17,18], depression [19], poor quality of life [20], and increased use of health care [21]. Additionally, age, gender, educational level, smoking, and obesity were associated with multimorbidity in previous research among adults aged 50 years and older [22]. Multimorbidity has become increasingly important as changing health behaviors, such as physical activity and obesity, are a core focus in multimorbidity prevention. Screening and behavioral changes, as well as developing intervention programs, may be important for prevention among people with multimorbidity.

About 12% to 14% of wage workers in Korea are shift workers, and weekly working hours vary from 50 hours to 58 hours depending upon the shift types, which is significantly higher than for day workers [23]. Shift work and long working hours are well-known risk factors for several chronic diseases.

However, the prevalence of multimorbidity and its associated factors in Korean shift workers are rarely reported. Identifying and understanding the prevalence of multimorbidity may yield important information for focused care of shift workers and help develop policies related to shift work schedules in Korea, as well as interventions needed to prevent and manage shift workers’ multimorbidity. We additionally performed a network analysis to determine multimorbidity patterns, as certain chronic diseases are likely to co-occur because their pathophysiological pathways are similar [24]. Understanding the patterns of these co-occurring chronic diseases may be beneficial as it may provide vital information for clinicians and policy makers to develop and implement intervention programs for specific groups with similar multimorbidity. Therefore, the purposes of this study are to (1) assess the prevalence of multimorbidity, (2) examine the factors associated with multimorbidity among shift workers in Korea, and (3) identify patterns of multimorbidity.

Methods

This study follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [25].

Design

This is a population-based, cross-sectional study using Korea National Health and Nutrition Examination Survey (KNHANES) data from 2016 to 2020.

Data Source

KNHANES uses a nationally representative sample based on a complex, stratified, multistage cluster sampling method that includes geographical area, gender, and age to provide a representation of the Korean population that is freely available to the public. KNHANES uses a series of cross-sectional national surveys that have been conducted by the Korean Centers for Disease Control and Prevention. These consist of a health survey, a health examination, and a nutrition survey [26].

Study Sample

The participants of this study were shift workers aged 19 years and older. Shift workers were defined as those who work at night or are nonday workers. This included both evening work, night work, day/night regular shift work, and irregular shift work. The exclusion criteria were as follows: (1) those who were younger than 19 years, (2) those who were nonworkers, and (3) those with missing data on any of the chronic diseases. The detailed sample selection flow is shown in Multimedia Appendix 1, Figure S1.

Ethical Considerations

This study was exempted from ethical review and approval by the institutional review board of Chung-Ang University, HCH’s institution (1041078-20230518-HR-139), as this study was a secondary analysis of preexisting data. The primary data (from
Measures

Multimorbidity
We defined multimorbidity as participants having 2 or more chronic conditions simultaneously. Based on previous studies [27-29], 13 common worldwide chronic diseases were included. These included hypertension, dyslipidemia, diabetes, arthritis, cancer, asthma, depression, osteoporosis, thyroid disease, coronary heart disease (CHD), chronic liver disease (CLD), stroke, and chronic kidney disease (CKD). The presence of disease was determined by self-reporting by the participants as to whether they had ever been diagnosed with each disease by a physician. The multimorbidity group was operationally defined as shift workers with 2 or more of the 13 diseases. The nonmultimorbidity group was defined as healthy participants or participants with only 1 chronic condition among the 13 diseases.

Demographic and Job-Related Variables
Sex, age, household income, education, and marital status were considered as demographic characteristics. Household income was divided into quartiles, which were calculated annually to evenly distribute the population into 4 groups by sex and age using the monthly average equivalized income. Education was divided into 4 groups: elementary school or less, middle school, high school, and university and above. Marital status was categorized as married or single. Job-related variables included regularity of work and average working hours per week. The type of shift work was divided into evening work, night work, regular shift work, and irregular shift work. Regular work was defined as that conducted by permanent workers or full-time employees, while nonregular work included work conducted by contract workers, contractors, and part-time workers.

Health Behaviors
This study used multiple health behavior variables, including BMI, smoking status, alcohol use, physical activity, and poor sleep duration. BMI was categorized as underweight (<18.5 kg/m^2), normal (≥18.5 to <23 kg/m^2), overweight (≥23 to <25 kg/m^2), and obese (≥25 kg/m^2) based on the World Health Organization cutoffs for Asia-Pacific countries [30]. Smoking status was categorized as nonsmoker, ex-smoker, and current smoker. Alcohol use was categorized into low-risk drinking, at-risk drinking, alcohol abuse, and alcohol dependence using the Korean version of the Alcohol Use Disorders Identification Test–Concise (AUDIT-C) [31]. Physical activity was measured by aerobic physical activity rate. If participants exercised for at least 2 hours and 30 minutes of moderate-intensity physical activity per week, or at least 1 hour and 15 minutes of high-intensity physical activity, or a mixture of moderate and high-intensity physical activities (1 min of high-intensity activity equals 2 min of moderate-intensity activity), then it was counted as “yes” for aerobic physical activity. Poor sleep duration was defined as either less than 7 hours or more than 9 hours per day; therefore, 7 to 9 hours of sleep per day was considered a good sleep duration [32,33].

Data Analysis
Data were analyzed using Stata (version 16.1; StataCorp LLC) and JASP (version 0.17.1; Eric-Jan Wagenmakers). Additionally, the KNHANES analytic guidelines were followed to adjust the data for complex sampling designs to estimate the population-level statistics for Korea. We conducted a complete case analysis without any imputation following the analytic guidelines [26]. All statistical tests were based on point estimation using a 2-sided P value (<.05) and interval estimation using 95% CIs. Specific analyses was conducted as follows: first, descriptive statistics were used, including weighted N, N, weighted percentage, weighted mean, and SE to present the population characteristics and their multimorbidity characteristics. Second, univariate analyses using the survey-corrected Rao-Scott χ² test and adjusted Wald test were conducted to reveal the differing characteristics between the multimorbidity and the nonmultimorbidity shift worker groups. Third, survey-corrected logistic regression analyses were used to identify factors influencing multimorbidity among shift workers. In the final model, variables that were statistically significant in the univariate analysis were included. We additionally conducted a subgroup analysis based on age, using 50 years as the threshold. Lastly, network analysis was used for exploring the patterns of multimorbidity clusters among shift workers. The network model was graphically represented by nodes (circles representing each morbidity) and edges (lines connecting the nodes). The structural importance of multimorbidity patterns was analyzed using node centrality measures, including closeness, betweenness, strength, and expected influence. Clustering and naming of subgroups within multimorbidity groups were determined through discussion among the researchers based on prior studies [7,34].

Results

Participant Characteristics
The final selected shift worker sample size was 1704 and the weighted population size was 2,697,228. The weighted percentage of male workers (56.7%) was slightly more than female workers, and the mean age of the participants was 41.93 (SE 0.42; SD 13.74) years, with an age range of 19 to 80 years (Table 1). The majority of the participants had household incomes in the third (n=538, 32.9%) and fourth (n=561, 33.4%) quartiles. As for educational attainment levels, 87.2% (n=1419) of participants had a high school education or higher. Additionally, there were more married (n=1209, 64.6%) than single participants. Participants were divided by shift work type: 920 (53.4%) in evening work, 191 (12.2%) in night work, 454 (27.3%) in regular shift work, and 139 (7.1%) in irregular shift work. The proportion of nonregular workers was high at 63.7% (n=821), and the average number of working hours per week was 38.01 (SD 17.67).
Table 1. General characteristics of shift workers (N=1704; weighted N=2,697,228).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
</tr>
<tr>
<td>Sex, n (weighted %)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>849 (56.7)</td>
</tr>
<tr>
<td>Female</td>
<td>855 (43.3)</td>
</tr>
<tr>
<td>Age (years), mean (SE; SD)</td>
<td>41.93 (0.42; 13.74)</td>
</tr>
<tr>
<td>House income (quartile), n (weighted %)</td>
<td></td>
</tr>
<tr>
<td>Low (first)</td>
<td>170 (9.3)</td>
</tr>
<tr>
<td>Lower middle (second)</td>
<td>432 (25.4)</td>
</tr>
<tr>
<td>Higher middle (third)</td>
<td>538 (32.9)</td>
</tr>
<tr>
<td>High (fourth)</td>
<td>561 (33.4)</td>
</tr>
<tr>
<td>Education, n (weighted %)</td>
<td></td>
</tr>
<tr>
<td>≤Elementary school</td>
<td>153 (5.7)</td>
</tr>
<tr>
<td>Middle school</td>
<td>132 (6.1)</td>
</tr>
<tr>
<td>High school</td>
<td>743 (46.4)</td>
</tr>
<tr>
<td>≥University</td>
<td>676 (41.8)</td>
</tr>
<tr>
<td>Marital status, n (weighted %)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1209 (64.6)</td>
</tr>
<tr>
<td>Single</td>
<td>495 (35.4)</td>
</tr>
<tr>
<td><strong>Occupational status</strong></td>
<td></td>
</tr>
<tr>
<td>Shift work type, n (weighted %)</td>
<td></td>
</tr>
<tr>
<td>Evening work</td>
<td>920 (53.4)</td>
</tr>
<tr>
<td>Night work</td>
<td>191 (12.2)</td>
</tr>
<tr>
<td>Regular shift work</td>
<td>454 (27.3)</td>
</tr>
<tr>
<td>Irregular shift work</td>
<td>139 (7.1)</td>
</tr>
<tr>
<td>Regular work, n (weighted %)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>417 (36.3)</td>
</tr>
<tr>
<td>No</td>
<td>821 (63.7)</td>
</tr>
<tr>
<td>Working hours per week, mean (SE; SD)</td>
<td>38.01 (0.59; 17.67)</td>
</tr>
<tr>
<td><strong>Health</strong></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td></td>
</tr>
<tr>
<td>Overall (kg/m²), mean (SE; SD)</td>
<td>24.20 (0.11; 3.47)</td>
</tr>
<tr>
<td>Normal, n (weighted %)</td>
<td>618 (35.7)</td>
</tr>
<tr>
<td>Underweight, n (weighted %)</td>
<td>70 (4)</td>
</tr>
<tr>
<td>Overweight, n (weighted %)</td>
<td>410 (23.6)</td>
</tr>
<tr>
<td>Obese, n (weighted %)</td>
<td>600 (36.7)</td>
</tr>
<tr>
<td>Smoking, n (weighted %)</td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>954 (51.9)</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>331 (19.9)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>419 (28.2)</td>
</tr>
<tr>
<td>Alcohol use, n (weighted %)</td>
<td></td>
</tr>
<tr>
<td>Low-risk drinking</td>
<td>969 (67.2)</td>
</tr>
<tr>
<td>At-risk drinking</td>
<td>133 (10.5)</td>
</tr>
</tbody>
</table>
The mean BMI was 24.20 (SE 0.11; SD 3.47) kg/m\(^2\), and overweight and obese participants (those with a BMI of 23.0 kg/m\(^2\) or higher) accounted for 60.3% of the total. Nonsmokers were the most common at 51.9% (n=954), and low-risk drinking was engaged in by the majority of the workers at 67.2% (n=969). The proportion of workers engaged in aerobic physical activity was 50.05% (n=808), which was similar to the proportion of participants not engaged in aerobic physical activity. The average sleep time was 7.34 (SD 1.33) hours, and workers who showed a poor sleep pattern accounted for the majority at 55.9% (n=940).

**Multimorbidity Characteristics of Shift Workers**

Table 2 shows the multimorbidity characteristics of the shift workers. Hypertension (n=287, 14.2%) was the most common of the 13 chronic conditions, followed by dyslipidemia (n=235, 11%), diabetes (n=107, 5.3%), and arthritis (n=126, 5.1%). The multimorbidity prevalence in shift workers was found to be 13.7% (n=302) and the number of chronic conditions ranged from 0 to 6.
Table 2. Multimorbidity characteristics of shift workers (N=1704; weighted N=2,697,228).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Workers, n (weighted %)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chronic conditions</strong></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>287 (14.2)</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>235 (11)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>107 (5.3)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>126 (5.1)</td>
</tr>
<tr>
<td>Cancer</td>
<td>90 (3.9)</td>
</tr>
<tr>
<td>Asthma</td>
<td>50 (3.3)</td>
</tr>
<tr>
<td>Depression</td>
<td>60 (3.1)</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>68 (2.7)</td>
</tr>
<tr>
<td>Thyroid disease</td>
<td>52 (2.7)</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>33 (1.3)</td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>14 (0.9)</td>
</tr>
<tr>
<td>Stroke</td>
<td>12 (0.5)</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>10 (0.5)</td>
</tr>
<tr>
<td><strong>Number of chronic conditions</strong></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>1046 (67.1)</td>
</tr>
<tr>
<td>One</td>
<td>356 (19.2)</td>
</tr>
<tr>
<td>Two</td>
<td>175 (8.3)</td>
</tr>
<tr>
<td>Three</td>
<td>84 (3.8)</td>
</tr>
<tr>
<td>Four</td>
<td>32 (1.2)</td>
</tr>
<tr>
<td>Five</td>
<td>8 (0.3)</td>
</tr>
<tr>
<td>Six</td>
<td>3 (0.1)</td>
</tr>
</tbody>
</table>

*The order of chronic conditions is from the highest weighted percentage to the lowest.*

*The average number of chronic conditions was 0.54 (SE 0.02; SD 0.87; range 0-6 for the total of 13 chronic conditions).*

**Differences in Characteristics by Presence of Multimorbidity**

The differences between the multimorbidity group and the nonmultimorbidity group are presented in Table 3. The average age of the multimorbidity group was 57.21 (SE 0.69; SD 10.47) years, which was statistically significantly higher than that of the nonmultimorbidity group (mean 39.51, SE 0.42, SD 12.69 years). The multimorbidity group had lower household income and educational level, a lower number of unmarried people, higher irregular work level, higher BMI, fewer current smokers, less physical activity, and a higher proportion of participants with poor sleep compared to the nonmultimorbidity group, with all differences being statistically significant.
Table 3. Differences in characteristics of the multimorbidity and nonmultimorbidity groups of shift workers (N=1704; weighted N=2,697,228).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Nonmultimorbidity (n=1402, 86.3%)</th>
<th>Multimorbidity (n=302, 13.7%)</th>
<th>F test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>710 (49.7)</td>
<td>139 (7.1)</td>
<td>2.71 (862)</td>
<td>.10</td>
</tr>
<tr>
<td>Female</td>
<td>139 (36.6)</td>
<td>163 (6.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age, mean (SE)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>956 (64.4)</td>
<td>45 (2.7)</td>
<td>497.79</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≥50</td>
<td>446 (21.9)</td>
<td>257 (11)</td>
<td>240.28</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>House income (quartile), n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (first)</td>
<td>113 (6.9)</td>
<td>58 (2.4)</td>
<td>10.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lower middle (second)</td>
<td>330 (20.3)</td>
<td>102 (4.1)</td>
<td>55.52</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Higher middle (third)</td>
<td>458 (28.9)</td>
<td>80 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (fourth)</td>
<td>499 (30.2)</td>
<td>62 (3.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤Elementary school</td>
<td>67 (2.7)</td>
<td>86 (3.0)</td>
<td>92.36</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Middle school</td>
<td>82 (4.1)</td>
<td>50 (2.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>632 (40.8)</td>
<td>111 (5.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥University</td>
<td>621 (38.7)</td>
<td>55 (3.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Marital status, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>922 (51.7)</td>
<td>287 (12.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>480 (34.6)</td>
<td>15 (0.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shift work type, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evening work</td>
<td>769 (46.4)</td>
<td>151 (7.0)</td>
<td>1.59</td>
<td>.19</td>
</tr>
<tr>
<td>Night work</td>
<td>162 (10.9)</td>
<td>29 (1.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular shift work</td>
<td>369 (23.1)</td>
<td>85 (4.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irregular shift work</td>
<td>102 (5.8)</td>
<td>37 (1.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Regular work, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>369 (33.0)</td>
<td>48 (3.3)</td>
<td>8.10</td>
<td>.005</td>
</tr>
<tr>
<td>No</td>
<td>652 (53.7)</td>
<td>169 (10.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Working hours per week, mean (SE)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.90 (0.62)</td>
<td>38.72 (1.42)</td>
<td></td>
<td>0.30</td>
<td>.58</td>
</tr>
<tr>
<td><strong>BMI, mean (SE)</strong></td>
<td>23.98 (0.12)</td>
<td>25.57 (0.24)</td>
<td>34.45</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>BMI category, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>536 (32.5)</td>
<td>82 (3.3)</td>
<td>11.87</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Underweight</td>
<td>67 (3.9)</td>
<td>3 (0.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>330 (19.9)</td>
<td>80 (3.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obese</td>
<td>463 (30.0)</td>
<td>137 (6.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Smoking, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>780 (44.7)</td>
<td>174 (7.2)</td>
<td>5.90</td>
<td>.003</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>257 (16.1)</td>
<td>74 (3.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>365 (25.5)</td>
<td>54 (2.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Alcohol use, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk drinking</td>
<td>830 (59.5)</td>
<td>139 (7.7)</td>
<td>1.19</td>
<td>.31</td>
</tr>
</tbody>
</table>
Table 4 presents the factors affecting multimorbidity in shift workers. Examining the relationship between each influencing factor and the 2 variables (crude odds ratio [OR]) of multimorbidity showed that age, household income, education, marital status, regularity of work, BMI, smoking, physical activity, and poor sleep quality were all factors influencing multimorbidity. In the final adjusted logistic model, factors affecting multimorbidity were age (OR 1.106, 95% CI 1.069-1.144; P<.001) third-quartile (OR 0.302, 95% CI 0.119-0.768; P=.01) vs first-quartile household income, fourth-quartile (OR 0.366, 95% CI 0.142-0.942; P=.04) vs first-quartile household income, nonregular work (OR 1.804, 95% CI 1.008-3.228; P=.047), and obese BMI (OR 2.152, 95% CI 1.155-4.010; P=.02) vs normal BMI. That is, when the age increased by 1 year, the risk of multimorbidity increased by 1.11 times, and the risk of multimorbidity in the third-quartile and fourth-quartile household income groups decreased by 69.8% and 63.4%, respectively, compared to the first-quartile household income group. The risk of multimorbidity in the nonregular work group increased by 1.8 times compared with the regular work group, and the risk of multimorbidity in the obese group increased by 2.15 times compared to the normal BMI group. Subgroup analysis showed that age, regular work, and obese BMI were associated with multimorbidity in the subgroup younger than 50 years, while household income, education, marital status, and obese BMI were associated in the subgroup aged 50 years and older (Multimedia Appendix 2, Table S1).
Table 4. Factors affecting multimorbidity in shift workers (N=1704; weighted N=2,697,228).

<table>
<thead>
<tr>
<th>Factors</th>
<th>Crude OR(^a) (95% CI)</th>
<th>(P) value</th>
<th>Adjusted OR (95% CI)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.106 (1.091-1.120)</td>
<td>&lt;.001</td>
<td>1.106 (1.069-1.144)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>House income (quartiles)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (first)</td>
<td>Reference</td>
<td><em>b</em></td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>Lower middle (second)</td>
<td>0.597 (0.378-0.943)</td>
<td>.03</td>
<td>0.467 (0.193-1.129)</td>
<td>.09</td>
</tr>
<tr>
<td>Higher middle (third)</td>
<td>0.406 (0.261-0.630)</td>
<td>&lt;.001</td>
<td>0.302 (0.119-0.768)</td>
<td>.01</td>
</tr>
<tr>
<td>High (fourth)</td>
<td>0.311 (0.195-0.496)</td>
<td>&lt;.001</td>
<td>0.366 (0.142-0.942)</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤Elementary school</td>
<td>Reference</td>
<td>—</td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>Middle school</td>
<td>0.439 (0.257-0.750)</td>
<td>.003</td>
<td>0.503 (0.202-1.255)</td>
<td>.14</td>
</tr>
<tr>
<td>High school</td>
<td>0.126 (0.082-0.196)</td>
<td>&lt;.001</td>
<td>0.611 (0.293-1.276)</td>
<td>.19</td>
</tr>
<tr>
<td>≥University</td>
<td>0.074 (0.046-0.119)</td>
<td>&lt;.001</td>
<td>0.528 (0.213-1.304)</td>
<td>.17</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>Reference</td>
<td>—</td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>Single</td>
<td>0.089 (0.049-0.164)</td>
<td>&lt;.001</td>
<td>0.749 (0.245-2.292)</td>
<td>.61</td>
</tr>
<tr>
<td><strong>Regular work</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Reference</td>
<td>—</td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>No</td>
<td>1.847 (1.199-2.847)</td>
<td>.006</td>
<td>1.804 (1.008-3.228)</td>
<td>.047</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>Reference</td>
<td>—</td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>Underweight</td>
<td>0.251 (0.074-0.859)</td>
<td>.03</td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>Overweight</td>
<td>1.815 (1.261-2.612)</td>
<td>.001</td>
<td>0.918 (0.466-1.809)</td>
<td>.81</td>
</tr>
<tr>
<td>Obese</td>
<td>2.216 (1.581-3.106)</td>
<td>&lt;.001</td>
<td>2.152 (1.155-4.010)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>Reference</td>
<td>—</td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>Ex-smoker</td>
<td>1.472 (1.025-2.115)</td>
<td>.04</td>
<td>0.950 (0.474-1.903)</td>
<td>.89</td>
</tr>
<tr>
<td>Current smoker</td>
<td>0.668 (0.446-0.999)</td>
<td>.05</td>
<td>0.987 (0.490-1.990)</td>
<td>.97</td>
</tr>
<tr>
<td><strong>Physical activity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>—</td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>Yes</td>
<td>0.701 (0.516-0.951)</td>
<td>.02</td>
<td>1.435 (0.819-2.514)</td>
<td>.21</td>
</tr>
<tr>
<td><strong>Poor sleep</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No (7-9 h)</td>
<td>Reference</td>
<td>—</td>
<td></td>
<td>_</td>
</tr>
<tr>
<td>Yes (&lt;7 h or &gt;9 h)</td>
<td>1.650 (1.213-2.245)</td>
<td>.001</td>
<td>1.467 (0.880-2.445)</td>
<td>.14</td>
</tr>
</tbody>
</table>

\(^a\)OR: odds ratio.
\(^b\)Not applicable.

Network Analysis

As presented in Figure 1, the network analysis revealed that chronic diseases were clustered into three groups: (1) cardiometabolic multimorbidity (hypertension, dyslipidemia, diabetes, CHD, and stroke), (2) musculoskeletal multimorbidity (arthritis and osteoporosis), and (3) unclassified diseases (depression, CLD, thyroid disease, asthma, cancer, and CKD). A centrality plot of multimorbidity patterns in shift workers shows hypertension, dyslipidemia, arthritis, diabetes, and osteoporosis by strength and expected influence in order (Multimedia Appendix 3, Figure S2).
Figure 1. Multimorbidity network in shift workers. CKD: chronic kidney disease; HNT: hypertension; CHD: coronary heart disease; CLD: chronic liver disease.

Discussion

Overview

This study aimed to identify patterns of multimorbidity and examine the factors associated with multimorbidity among shift workers in Korea. Among the 1704 shift workers included in the study, approximately 14% had multimorbidity. A recent meta-analysis of multimorbidity in the worldwide adult population found that there was a prevalence of approximately 37% [15]. The difference in prevalence could be due to several factors. First, the mean age of our sample was relatively low at 41.9 (SD 13.7) years, compared to 56.9 years in the meta-analysis. It has been recognized that older age is associated with an increased number of chronic diseases [35]. Furthermore, women are more prone to developing chronic diseases than men [36,37], and a higher proportion of our sample was male.

It is recognized that shift work is associated with adverse health outcomes. Previous research reported that shift work increases the risk of cardiovascular incidence and mortality, cancer, and stroke [1,7,38-40]. Approximately 14% of the Korean population is involved in shift work, and although some shift work is inevitable and may provide economic benefits, policies are needed to modify and manage shift schedules. Socioeconomic and behavioral factors are also known to increase the risk for multimorbidity. Consistent with previous research, we found that shift workers in the low-income and obese groups had increased risks of developing multimorbidity. The relationship between low income and the development of chronic diseases and multimorbidity may be explained by less physical activity and lower fruit and vegetable consumption in the low-income group [41,42]. Furthermore, obesity is a major risk factor associated with developing several chronic diseases, including diabetes, heart disease, asthma, arthritis, and depression [43]. We found that a high proportion of Korean workers are involved in nonregular work (63.7%), and having regular work was a significant factor influencing multimorbidity. Many nonregular workers are paid low wages (less than US $10 per hour), are employed on 2-year fixed-term contracts, and may be required to extend their employment or ask to terminate the employment every 2 years. Those who work less than a certain number of hours per week are not covered by social insurance and are not entitled to weekly holidays or paid annual leave, which could be important factors contributing to poor health outcomes among nonregular workers in Korea [44]. Labor laws may require reform so that nonregular workers are guaranteed stability in their jobs, social insurance coverage, and proper wages to enable a minimum standard of living. However, longitudinal and cohort studies are needed to assess and analyze the relationships between shift work, socioeconomic and behavioral factors, and multimorbidity. Further, the subgroup analysis revealed that education and marital status were significant factors influencing multimorbidity in shift workers aged 50 years and older. Consistent with previous research, the risk of multimorbidity increased when workers were less educated and when they were single [45,46]. A deeper understanding of these relationships

https://publichealth.jmir.org/2024/1/e55014

JMIR Public Health Surveill 2024 | vol. 10 | e55014 | p.1271

(page number not for citation purposes)
may require further analysis to identify the subgroups at greatest risk for multimorbidity and develop targeted interventions, as there could be interactions between socioeconomic and demographic factors.

The most common chronic disease in this study was hypertension, followed by dyslipidemia, diabetes, and arthritis. This is similar to the findings of a previous meta-analysis of multimorbidity in the United States, United Kingdom, and Spain [20]. Additionally, we performed a network analysis and found 2 distinct groups: hypertension, dyslipidemia, diabetes, CHD, and stroke clustered as one group; arthritis and osteoporosis clustered as another group. Previous research found that the coexistence of hypertension and at least 1 other comorbidity was most common among patients with multimorbidity [47]. Similar to our findings, Yang et al [7] found that shift workers with hypertension in the United Kingdom were more prone to cardiometabolic multimorbidity, including diabetes, coronary artery disease, and stroke. However, the only known multimorbidity cluster is cardiovascular multimorbidity, and the evidence on multimorbidity in other chronic diseases and their clusters in shift workers remains limited. In the general population, cardiovascular diseases and metabolic diseases tend to cluster together, and osteoarthritis tended to cluster together in an Australian sample, finding that is similar to ours [48]. Some chronic diseases are more likely to cluster together, as seen in our study and the studies in the United Kingdom and Australia. Some chronic diseases are more likely to cluster together as they may share similar pathophysiological pathways [24]. Also, shift work is known to disrupt circadian rhythms and sleep and is associated with overweight and with blood glucose levels, which are important factors for developing cardiometabolic diseases [49,50]. Although it is controversial, some have suggested that patients with arthritis may have greater bone loss [51]. Considering the relationship between arthritis and osteoporosis and that the incidence of both conditions increases with age [52,53], and that the mean age of our participants with multimorbidity was 57 (SD 10.47) years, the co-occurrence of these chronic disease may be understandable. Identifying these disease clusters can be important, as we can target and tailor interventions for specific groups. However, caution is needed when interpreting and comparing clusters of multimorbidity or patterns across studies, as researchers use different statistical methods to analyze clusters or patterns and may include a different list of chronic diseases in their analyses. Further research is needed to better understand and assess multimorbidity clusters and their trajectories and patterns over time among shift workers using unified and valid analytic methods with the most prevalent chronic diseases in shift workers.

Most chronic care models and guidelines focus on the treatment and management of individual chronic diseases [54]. Multimorbidity requires more complex care that prioritizes what is most important for each patient. Several types of interventions for multimorbidity have been implemented in the past, but the evidence to deliver specific interventions remains limited. Furthermore, interventions specific for shift workers with multimorbidity are relatively unknown, despite shift workers being at greater risk for developing multimorbidity than nonshift workers. Nevertheless, occupational or organization-based interventions may benefit shift workers. These potentially include policy development regarding shift schedules, early screening diagnostic interventions, and diet and physical activity programs. However, additional research remains necessary to confirm the multimorbidity clusters in shift workers and to confirm the differences between socioeconomic and behavioral factors to enable the development of specific intervention programs for each cluster group.

To our knowledge, this study is the first to assess multimorbidity and its associated factors in Korean shift workers. We analyzed a 2016-2020 series of national population-based study data (KNHANES) using network analysis to reveal multimorbidity patterns. However, this study has some limitations. First, KNHANES is a cross-sectional study; thus, we cannot assume temporal relationships between variables. We also could not assess the long-term patterns of multimorbidity development and the effect of socioeconomic behavioral factors on multimorbidity. Therefore, we recommend longitudinal studies on multimorbidity among shift workers in the future. Second, the data were based on self-reports from participants. Thus, recall bias cannot be excluded. Some objective data, for example, medical records to identify multimorbidity, may be needed in a future study. Third, the specific patterns of shift work were not specified in the current data set. For example, we do not know if the shift workers had rotating shifts, only night shifts, or how many night or evening shifts were included in their work schedules. Further studies should specify the shift work schedules; a subgroup analysis may be beneficial. Fourth, we also recommend subgroup analysis of different work sectors, since working conditions and shift schedules may differ. For example, the health care and manufacturing industries may use more complex shift schedules and require more specific interventions based on subgroup analysis. Fifth, network analysis provides only a graphical presentation of multimorbidity patterns and does not allow for statistical analyses, such as regression, to identify factors associated with each multimorbidity pattern. Future studies may consider using other statistical analyses, such as latent class analysis, if researchers are interested in exploring factors associated with multimorbidity patterns in shift workers. Finally, the study results may have excluded some confounding variables, such as family support and use of health care, that may have influenced multimorbidity.

Conclusion

Multimorbidity is a crucial factor influencing premature death, poor health and quality of life, and use of health care. Our findings indicate that approximately 14% of Korean shift workers have multimorbidity and that several socioeconomic and behavioral factors are associated with multimorbidity. This suggests that policy development regarding work schedule modification is necessary. Furthermore, screening and tailored intervention programs at the organizational level may benefit efforts to prevent and monitor multimorbidity among shift workers. However, we also suggest a future longitudinal study to assess and confirm multimorbidity patterns among shift workers.
Acknowledgments
The authors would like to thank the Korea Disease Control and Prevention Agency for offering the raw data from the Korea National Health and Nutrition Examination Survey (KNHANES). This paper’s contents are solely the responsibility of the authors and do not necessarily represent the official view of KNHANES. This research was supported by Chung-Ang University research grants in 2023 and a National Research Foundation of Korea grant funded by the Ministry of Science and ICT of the Korean government (2022R1F1A1066262)

Data Availability
The data supporting the findings of this study are derived from the Korea National Health and Nutrition Examination Survey (KNHANES), which is publicly accessible at no cost. Interested researchers can access the data by visiting the KNHANES website [55]. Please note that while the data are freely available, users are encouraged to comply with the terms and conditions specified on the KNHANES portal.

Authors' Contributions
HCH and YMK conceptualized the study idea; YMK collected and analyzed the data; HCH supervised the study; and HCH and YMK led manuscript writing.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Flow diagram of study sample selection.
[DOCX File, 45 KB - publichealth_v10i1e55014_app1.docx]

Multimedia Appendix 2
Age-subgroup analysis of factors affecting multimorbidity in shift workers.
[DOCX File, 35 KB - publichealth_v10i1e55014_app2.docx]

Multimedia Appendix 3
Centrality plot of multimorbidity patterns in shift workers presenting hypertension, dyslipidemia, arthritis, diabetes, and osteoporosis according to strength and expected influence in order.
[DOCX File, 155 KB - publichealth_v10i1e55014_app3.docx]

References


Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test-Concise
CHD: coronary heart disease
CLD: chronic liver disease
CKD: chronic kidney disease
KNHANES: Korea National Health and Nutrition Examination Survey
OR: odds ratio
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WHO: World Health Organization

©Hye Chong Hong, Young Man Kim. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 10.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Pandemic Fatigue and Preferences for COVID-19 Public Health and Social Measures in China: Nationwide Discrete Choice Experiment

Meng Yang, MBBS; Zonglin He, MBBS; Yin Zhang, MPH; Taoran Liu, PhD; Wai-kit Ming, MPH, MD, PhD

1School of Medicine, Jinan University, Guangzhou, China
2Division of Life Science, Hong Kong University of Science and Technology, Hong Kong, China (Hong Kong)
3Department of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
4Department of Infectious Diseases and Public Health, City University of Hong Kong, Hong Kong, China (Hong Kong)

*these authors contributed equally

Corresponding Author:
Wai-kit Ming, MPH, MD, PhD
Department of Infectious Diseases and Public Health
City University of Hong Kong
Room 1A-503, 5/F, Block 1, To Yuen Building, 31 To Yuen Street
Hong Kong
China (Hong Kong)
Phone: 852 34426956
Email: wkming2@cityu.edu.hk

Abstract

Background: Information on the public’s preferences for current public health and social measures (PHSMs) and people’s mental health under PHSMs is insufficient.

Objective: This study aimed to quantify the public’s preferences for varied PHSMs and measure the level of pandemic fatigue in the COVID-19 normalization stage in China.

Methods: A nationwide cross-sectional study with a discrete choice experiment and psychometric scales was conducted to assess public preferences for and attitudes toward PHSMs, using the quota sampling method. The COVID-19 Pandemic Fatigue Scale (CPFS) was used to screen fatigue levels among respondents. The multinomial logit model, latent class model, and Mann-Whitney test were used for statistical analysis. We also conducted subgroup analysis based on sex, age, monthly income, mental health status, and pandemic fatigue status.

Results: A total of 689 respondents across China completed the survey. The discrete choice experiment revealed that respondents attached the greatest importance to the risk of COVID-19 infection within 3 months (45.53%), followed by loss of income within 3 months (30.69%). Vulnerable populations (low-income populations and elderly people) were more sensitive to the risk of infection, while younger respondents were more sensitive to income loss and preferred nonsuspension of social places and transportation. Migrants and those with pandemic fatigue had less acceptance of the mandatory booster vaccination and suspension of transportation. Additionally, a higher pandemic fatigue level was observed in female respondents, younger respondents, migrants, and relatively lower-income respondents (CPFS correlation with age: r=–0.274, P<.001; correlation with monthly income: r=–0.25, P<.001). Mandatory booster COVID-19 vaccination was also not preferred by respondents with a higher level of pandemic fatigue, while universal COVID-19 booster vaccination was preferred by respondents with a lower level of pandemic fatigue.

Conclusions: Pandemic fatigue is widely prevalent in respondents across China, and respondents desired the resumption of normal social life while being confronted with the fear of COVID-19 infection in the normalization stage of COVID-19 in China. During future pandemics, the mental burden and adherence of residents should be considered for the proper implementation of PHSMs.

(JMIR Public Health Surveill 2024;10:e45840) doi:10.2196/45840
KEYWORDS
pandemic fatigue; preference; public health and social measures; discrete choice experiment; COVID-19

Introduction

The transmission of the SARS-CoV-2 Omicron variant led to a sharp rise in infected cases in mainland China, spreading from major cities like Guangzhou and Shanghai to the entire country [1]. To contain the transmission of the virus, various public health and social measures (PHSMs) have been adopted at municipal and provincial levels in China under the dynamic zero-COVID policy [2]. These measures include suspension of public transport, closure of public places, closed-off community management, mandatory nucleic acid testing, home quarantine, and isolation of infected and suspected cases, among others [3-5]. However, the severity of the epidemic combined with high-level public health policies during the Omicron wave had significant impacts on the normal life of citizens from different dimensions and may have resulted in mental health issues [6].

Fatigue issues have been noticeable during the COVID-19 pandemic, especially the adverse psychological impacts of nonpharmacological interventions (NPIs) [7]. In China, the prevalences of anxiety and depression symptoms were reported to be 29% and 37.1%, respectively, during the COVID-19 pandemic in 2020 [8]. A study in Italy found that 38% of the general population had psychological distress during the early stage of the COVID-19 pandemic [9]. Among them, vulnerable populations, including elderly people [10,11], migrant workers [12,13], children [14-16], adolescents [16,17], and individuals with pre-existing mental illness [18,19], may have a greater risk of psychological disturbances due to increased exposure to external adverse circumstances. The prevalence of mental disorders was found to be higher during the Omicron wave than during the wild-type wave. For example, the study by Lu et al [20] found that among nonmedical and medical staff, the prevalence rates of anxiety were 55.0% and 47.3%, respectively, and the rates of depression were 62.4% and 53.4%, respectively.

Long-term COVID-19 public health policy may result in pandemic fatigue [21,22], causing a decline in public compliance [23,24]. Changes in people’s perceptions of risk assessment have also led to behavioral changes [23,25]. Rayani et al [26] reported that higher levels of risk perception might allow people to maintain positive preventive behaviors. A study by Alijanzadeh et al [27] in Iran also showed that the risk perception of individuals can influence preventive COVID-19 behaviors through their fear of COVID-19 and trust in the health care system. Meanwhile, public participation at the policy level in preventive behavior, disease response, and surveillance has become increasingly important [28]. Information on public perceptions and attitudes toward social distancing measures is prominent in the unofficial media. In contrast, formal research evidence on the public’s preferences toward the current PHSMs and people’s mental health problems during the Omicron wave under the strengthened COVID-19 policy has been insufficient. Moreover, no studies to date have captured the desirability of the different PHSMs toward the pandemic in China or have captured the general public’s willingness to trade. Such insufficient information on the general public’s pandemic fatigue and preference may hinder priority settings when no single PHSM can sufficiently combat the transmission of the virus.

In the context of PHSMs, this study aimed to explore the public’s preferences and preference homogeneities and heterogeneities for varied PHSMs. Furthermore, based on an assessment of the current level of prevention and control measures in participants’ regions, this study considered the impact of PHSM fatigue on preferences according to the epidemiological fatigue scale [29,30].

Methods

Overview

In this study, we used various instruments to investigate mental health problems among the general population, especially migrant workers and those who work in nonregistered locations for 3 months or more [31]. The first instrument was a discrete choice experiment (DCE) questionnaire, which had a survey-based experiment design that solicited and quantified respondents’ utilities and preferences toward a set of attributes and levels of PHSMs. Following the DCE questionnaire, a Likert psychometric scale of pandemic fatigue was used to measure the respondents’ perceptions of the current PHSMs and levels of pandemic fatigue. Additionally, we conducted a subgroup analysis [32] to explore the heterogeneities based on demographic information and socioeconomic status, and a comparison was conducted of the preferences of respondents with relatively low pandemic fatigue levels and those with high fatigue levels.

Respondents

The inclusion criteria of this study were age of at least 18 years and absence of cognitive impairments (self-report). Respondents were recruited and selected through an online social media advertising platform (Credamo Inc), which has over 3 million samples and covers all provincial administrative regions in China [33-36]. Credamo randomly distributed the survey in 31 provinces of China (excluding Hong Kong, Macao, and Taiwan). Specifically, we provided a quota size of 350 per sex group and 140 per age group (oversampled). According to this census, the population of China was approximately 1.411 billion, with 51.24% male individuals (723.34 million) and 48.76% female individuals (688.44 million) [37,38], leading to a sex ratio of approximately 105.07 male individuals for every 100 female individuals. We also referred to the National Bureau of Statistics of China [37] for age-specific quota design, with 63.35% in the age group of 15 to 59 years and 21.76% in the age group of 60 years or older. However, due to budget restrictions and practical issues during the pandemic, we considered a 1:1 ratio per sex group and 140 individuals per age group for the data collection platform. No personally identifiable information was collected as the survey was anonymous. Consent was obtained when the respondents actively pressed the button marked “I have been informed with sufficient information of the study and agree to participate in this study” after viewing the introductory section.
of the questionnaire where the background and objectives of the study were presented. Respondents could only access the questionnaire if they consented and reported that they were 18 years or older and did not have cognitive impairments. The translation of the original survey has been provided in Multimedia Appendix 1. Respondents received RMB 20 (US $2.76) as an incentive for participation.

Data Collection
An anonymous self-administered survey created using Lighthouse Studio (version 9.9.1; Sawtooth Software) was distributed from July 01, 2022, to September 30, 2022, and respondents from the entire country were considered for inclusion. The minimum sample size requirement of this study was calculated using the rule of thumb approach proposed previously [39]. Specifically, the equation for sample size calculation was as follows:

\[ N > \frac{500c}{(t \times a)} \]

where \( t \) refers to the choice tasks in the survey, \( a \) refers to the number of alternatives, and \( c \) refers to the number of analysis cells. Specifically, the number of analysis cells \( c \) in this study refers to the largest number of levels for any of the attributes. As such, the minimum sample size in this study should be 125 respondents. Moreover, according to the standard parametric approach [40] of sample size calculation, the minimum sample size is 267 (Multimedia Appendix 2).

All the questions were close-ended, with tick boxes provided for responses and no question skipping allowed. No data were stored if the questionnaire website was closed before the completion of the survey.

Survey and DCE Design
The survey of this study had 4 main sections. Specifically, in the first section, we aimed to solicit respondents’ demographic information, including socioeconomic information (age, sex, education level, religion, marital status, occupation, income level, current residence, and registered permanent residence city).

In the second section of the survey, respondents’ vaccination history and medical history were collected. Respondents were asked how many doses of COVID-19 vaccination they have received, whether they have ever been diagnosed with or are currently experiencing psychological diseases (eg, depression, anxiety, obsessive-compulsive disorder, phobia, bipolar disorder, neurasthenia, schizophrenia, and personality disorder), and whether they have ever been infected with COVID-19. If respondents answered that they had been diagnosed with or are currently experiencing psychological diseases, they were required to answer what specific disease they encountered, the severity of the disease, and whether they have ever received or are currently receiving treatment. In addition, if respondents indicated that they had been infected with COVID-19, they were required to provide information about how they found out that they had been infected, their symptoms and complications, and their date of hospitalization and discharge.

The third section of the survey was the DCE. Respondents were presented with 9 sets of scenarios, and in each scenario, respondents faced 3 hypothetical responses, namely, “option A,” “option B,” and “neither.” Respondents were required to select the measure that they felt most satisfied with. The attributes and levels of different measures of the DCE were determined by a literature review [32,41,42] and consultation with local epidemiologists and experts, and according to design guidelines for DCE [43]. As a result, we determined 8 attributes in our study: (1) Risk of COVID-19 infection within 3 months; (2) Closure of social occasions; (3) Suspension of on-campus educational activities; (4) Suspension of public transportation; (5) Contact tracing, isolation, and quarantine; (6) Nucleic acid screening program; (7) Mandatory booster vaccination; and (8) Loss of income in 3 months. All the attributes and levels selected in the study have been summarized in Table 1.
Table 1. Attributes and levels selected in the discrete choice experiment survey.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk of COVID-19 infection within 3 months</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>Closure of social occasions</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Suspension of on-campus educational activities</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Suspension of public transportation</td>
<td>Full suspension</td>
</tr>
<tr>
<td></td>
<td>Suspension in high-risk areas</td>
</tr>
<tr>
<td></td>
<td>Normal operation</td>
</tr>
<tr>
<td>Contact tracing, isolation, and quarantine</td>
<td>Voluntary</td>
</tr>
<tr>
<td></td>
<td>Compulsory</td>
</tr>
<tr>
<td></td>
<td>None</td>
</tr>
<tr>
<td>Nucleic acid screening program</td>
<td>Only high-risk units, workplaces, and vulnerable public</td>
</tr>
<tr>
<td></td>
<td>Nucleic acid screening for all</td>
</tr>
<tr>
<td></td>
<td>None</td>
</tr>
<tr>
<td>Mandatory booster vaccination</td>
<td>Universal vaccination</td>
</tr>
<tr>
<td></td>
<td>Only high-risk groups are vaccinated (long-term patients, people over 60 years old, etc)</td>
</tr>
<tr>
<td></td>
<td>None</td>
</tr>
<tr>
<td>Loss of income in 3 months</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

A sample of a hypothetical choice task is shown in Figure 1. The levels in task choices of different versions were designed according to the principles of (1) orthogonality and (2) balance. The task choices in the DCE section were 8 random choices and 1 fixed choice. We used the fixed choice for further data quality control. Since the DCE questionnaire is relatively more complicated for respondents to understand and such cognitive burden imposed on respondents may lead to some bias in their selection, we added specific text and forced respondents to stay on the questionnaire page for at least 1 minute and carefully read the text to help them better understand what discrete choice tasks are and how to select the choices subsequently. Details are provided in Multimedia Appendix 1.
Figure 1. A sample of a hypothetical choice task in the discrete choice experiment (DCE) survey. Nine task choices in total were present in the DCE part. Each task choice contained 2 options (option 1 and option 2) and an “opt-out” option (none of both). Options were characterized by 8 attributes and random levels. Respondents were required to select an option from among the 2 options and the “opt-out” option.

Assuming that the government will implement two different sets of public health measures to mitigate the spread of the coronavirus, please select one of the following two hypothetical public health measures that you prefer.

(Q1. 9 questions in total) Please note: some scenarios may not be realistic, but please also choose the one that better suits your preference based on the information given.

<table>
<thead>
<tr>
<th>Risk of COVID-19 infection within 3 months</th>
<th>Option 1</th>
<th>Option 2</th>
<th>None of both</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Closure of social occasions</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Suspension of on-campus educational activities</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Suspension of public transportation</td>
<td>Suspension in high-risk areas</td>
<td>Normal operation</td>
<td>None of both</td>
</tr>
<tr>
<td>Contact tracing, isolation, and quarantine</td>
<td>Voluntarily</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Nucleic acid screening program</td>
<td>Only high-risk units, workplaces and vulnerable public</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Mandatory booster vaccination</td>
<td>None</td>
<td>Universal vaccination</td>
<td></td>
</tr>
<tr>
<td>Loss of income in 3 months</td>
<td>100%</td>
<td>40%</td>
<td></td>
</tr>
</tbody>
</table>

The description was as follows:

In this part, you will face a series of tasks; these are called discrete choice tasks, a method we use to understand preferences and decision-making processes. Each task will offer you two hypothetical options and a “none” option, each with a set of attributes or features. Your task is to choose the option that you prefer or that you would most likely choose in real life. Please read the descriptions of each option carefully. Each option is different, with its unique set of attributes or characteristics. Remember, there’s no right or wrong answer here. We are interested in your genuine preferences. Choose the option that best aligns with what you
would prefer in real life, based on the attributes presented. Some scenarios may not happen in real life since they were hypothetical; however, please also select an alternative based on your own preferences.

According to the full factorial design, there were 11,664 (6×2×2×3×3×3×6) policies and 11,664×11,663 task choices. To significantly reduce the complication of the design in order to ensure that respondents could complete the tasks, we applied a fractional factorial design based on balance (the frequencies of attribute levels are roughly equal across all tasks) and orthogonality (the frequencies of attribute pairs are roughly equal across the tasks) principles.

Psychological Likert scales were included in the fourth section of the survey, and pandemic fatigue was assessed on a 5-point Likert scale. The reliability (α=.885) and validity (Kaiser-Meyer-Olkin measure=0.737) of the Likert scale for pandemic fatigue were tested. The pandemic fatigue model designed by Lilleholt et al [29] was used to ask about demotivation toward COVID-19 PHSMs and the desire to know the development of the epidemic. We adjusted the pandemic fatigue model in our study. The adjusted pandemic fatigue model contained a series of questions on public attitudes or views on the strengthening of relevant measures for epidemic prevention and control at the current stage and a series of scales to measure the public’s pandemic fatigue. The first question asked about local confirmed cases in the respondent’s living area (town, county, and district), and it was followed by a question that asked about the current PHSMs in that area. The third question contained a scale on the respondent’s perceived risk of being infected with COVID-19. The fourth question contained a scale to measure the respondent’s perceptions of the current measures. Subsequently, the fifth question assessed the epidemic prevention fatigue situation under the current situation of strengthened epidemic prevention measures.

Statistical Analysis

Descriptive analysis was performed to describe respondents’ demographic information and socioeconomic information, as well as information regarding migration, COVID-19 vaccination history, mental health disease history, exposure to COVID-19, infection with COVID-19, and experience of closed-off community management (entrance numbers were minimized, checking points were set up in communities, entry permits were limited, face mask wearing was required, health monitoring was enhanced, and only registered personnel and vehicles were allowed to pass through).

We used the multinomial logit (MNL) model to quantify respondents’ relative utilities among all respondents. The MNL model of this study followed the random utility maximization theory [44]. We calculated the odds ratio (OR) and 95% CI based on respondents’ relative utilities among levels and attributes to further measure respondents’ preferences. In addition, we applied the latent class model (LCM) to determine how respondents’ preferences differed according to group membership. We used the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to determine the appropriate number of groups among the respondents. The Mann-Whitney test was applied for the analysis of quantitative variables. The scale data were analyzed using SPSS version 23 (IBM Corp), and the MNL model and LCM were adopted in Lighthouse Studio (version 9.9.1).

Subgroup Analysis Procedure

The LCM was robust in identifying unobserved heterogeneity within the data, and this method allowed for the identification of latent classes of individuals who exhibited similar preferences or characteristics. However, the LCM was not appropriate for investigating the association between pandemic fatigue and preference heterogeneities, as covariates, such as sex, age, and mental health status, were not controlled. Therefore, to further explore the preference heterogeneities among the respondents on controlling sex, age, and mental health status, we also conducted subgroup analyses based on respondents’ demographic information, including sex, age, monthly income, and mental health status. Moreover, we conducted a subgroup analysis based on respondents’ levels (high level or low level) of pandemic fatigue according to the results of the pandemic fatigue scale.

Ethical Considerations

We collected consent from respondents through an online consent form in the survey. This study has been approved by the Institutional Review Board (IRB) at the City University of Hong Kong (reference number: 11-2022-65-E). We adhered to the International Society for Pharmacoeconomics and Outcome Research (ISPOR) reporting guidelines for designing and reporting the research questions, assessing attributes and levels, and performing statistical analysis for the DCE.

Results

Respondent Characteristics

A total of 1183 respondents clicked on the link of our survey, and of these, 855 completed the survey. After the control process for data quality involving the exclusion of respondents who wrongly answered a trap question, 689 respondents were included in the final analysis. Among the 689 respondents, 341 (49.5%) were male and 348 (50.5%) were female. Additionally, 286 (41.5%) respondents were aged 35 years or younger. Most respondents (509/689, 73.9%) had a monthly income equal to or less than RMB 10,000 (US $1378.15), and 30.5% (210/689) of respondents were migrants (Table 2).
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value (N=689), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>341 (49.5)</td>
</tr>
<tr>
<td>Female</td>
<td>348 (50.5)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>128 (18.6)</td>
</tr>
<tr>
<td>26-35</td>
<td>159 (22.9)</td>
</tr>
<tr>
<td>36-45</td>
<td>135 (19.6)</td>
</tr>
<tr>
<td>46-55</td>
<td>129 (18.7)</td>
</tr>
<tr>
<td>≥56</td>
<td>139 (20.2)</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
</tr>
<tr>
<td>Below bachelor’s degree</td>
<td>181 (26.3)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>401 (58.2)</td>
</tr>
<tr>
<td>Above bachelor’s degree</td>
<td>107 (15.6)</td>
</tr>
<tr>
<td><strong>Current residence</strong></td>
<td></td>
</tr>
<tr>
<td>Northeast China</td>
<td>28 (4.1)</td>
</tr>
<tr>
<td>North China</td>
<td>122 (17.7)</td>
</tr>
<tr>
<td>East China</td>
<td>244 (35.4)</td>
</tr>
<tr>
<td>Central China</td>
<td>105 (15.2)</td>
</tr>
<tr>
<td>South China</td>
<td>103 (14.9)</td>
</tr>
<tr>
<td>Southwest China</td>
<td>66 (9.6)</td>
</tr>
<tr>
<td>Northwest China</td>
<td>21 (3.1)</td>
</tr>
<tr>
<td><strong>Original residence</strong></td>
<td></td>
</tr>
<tr>
<td>Northeast China</td>
<td>29 (4.2)</td>
</tr>
<tr>
<td>North China</td>
<td>108 (15.7)</td>
</tr>
<tr>
<td>East China</td>
<td>233 (33.8)</td>
</tr>
<tr>
<td>Central China</td>
<td>143 (20.8)</td>
</tr>
<tr>
<td>South China</td>
<td>86 (12.5)</td>
</tr>
<tr>
<td>Southwest China</td>
<td>70 (10.2)</td>
</tr>
<tr>
<td>Northwest China</td>
<td>20 (2.9)</td>
</tr>
<tr>
<td><strong>Religion</strong></td>
<td></td>
</tr>
<tr>
<td>Christianity</td>
<td>16 (2.3)</td>
</tr>
<tr>
<td>Mohammedanism</td>
<td>3 (0.4)</td>
</tr>
<tr>
<td>Buddhism</td>
<td>63 (9.1)</td>
</tr>
<tr>
<td>Others</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td>None</td>
<td>605 (87.8)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Unmarried and single</td>
<td>188 (27.3)</td>
</tr>
<tr>
<td>Unmarried and cohabiting</td>
<td>21 (3.0)</td>
</tr>
<tr>
<td>Married</td>
<td>469 (68.1)</td>
</tr>
<tr>
<td>Divorced</td>
<td>9 (1.3)</td>
</tr>
<tr>
<td>Widow</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td><strong>Migrant</strong></td>
<td></td>
</tr>
</tbody>
</table>
Occupation and working status

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value (N=689), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>210 (30.5)</td>
</tr>
<tr>
<td>No</td>
<td>479 (69.5)</td>
</tr>
</tbody>
</table>

Monthly income (RMB$^a$)

<table>
<thead>
<tr>
<th>Value</th>
<th>(N=689), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000 or below</td>
<td>509 (73.8)</td>
</tr>
<tr>
<td>10,001 or above</td>
<td>180 (26.2)</td>
</tr>
</tbody>
</table>

History of mental health disease

<table>
<thead>
<tr>
<th>Value</th>
<th>(N=689), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>42 (6.1)</td>
</tr>
<tr>
<td>No</td>
<td>634 (92.0)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>13 (1.9)</td>
</tr>
</tbody>
</table>

$^a$A currency exchange rate of 1 RMB=0.138 USD is applicable.

Preferences for Public Health Measures

The attribute that had the most weighted importance in respondents’ decision-making was the risk of being infected with COVID-19 in 3 months (45.53%), followed by the loss of income due to COVID-19 measure (30.69%). Suspension of on-campus educational activities (1.29%) had the weakest weighted preference (Figure 2). Weaker preferences were observed when increasing the risk of infection with COVID-19. Compared with the full suspension of public transportation, respondents believed that suspension in only high-risk areas (areas with 10 or more local confirmed cases were designated as high-risk areas) would be associated with larger utility (OR 1.168, 95% CI 1.106-1.234; *P*<.002). Moreover, compulsory contact tracing was favored by respondents compared with voluntary contact tracing (OR 1.294, 95% CI 1.225-1.366; *P*<.001). In addition, respondents were willing to accept booster doses of COVID-19 vaccines, and their utility decreased along with loss of income within 3 months due to PHSMs (Table 3).
Figure 2. Weighted attribute importance among public and different latent classes of respondents. (A) Weighted attribute importance among all respondents. (B-E) Weighted attribute importance among 4 latent classes of respondents (groups 1-4). A larger proportion represents a higher attribute importance.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient&lt;sup&gt;a&lt;/sup&gt;</th>
<th>SE</th>
<th>P value</th>
<th>OR&lt;sup&gt;b&lt;/sup&gt; (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risk of COVID-19 infection within 3 months</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0% (reference)</td>
<td>0.832</td>
<td>0.044</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0.516</td>
<td>0.044</td>
<td>&lt;.001</td>
<td>0.729 (0.669-0.795)</td>
</tr>
<tr>
<td>40%</td>
<td>0.299</td>
<td>0.044</td>
<td>&lt;.001</td>
<td>0.587 (0.538-0.639)</td>
</tr>
<tr>
<td>60%</td>
<td>-0.171</td>
<td>0.046</td>
<td>&lt;.001</td>
<td>0.367 (0.335-0.401)</td>
</tr>
<tr>
<td>80%</td>
<td>-0.535</td>
<td>0.049</td>
<td>&lt;.001</td>
<td>0.255 (0.231-0.281)</td>
</tr>
<tr>
<td>100%</td>
<td>-0.941</td>
<td>0.055</td>
<td>&lt;.001</td>
<td>0.170 (0.152-0.189)</td>
</tr>
<tr>
<td><strong>Closure of social occasions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (reference)</td>
<td>-0.025</td>
<td>0.018</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.025</td>
<td>0.018</td>
<td>.16</td>
<td>1.052 (1.016-1.090)</td>
</tr>
<tr>
<td><strong>Suspension of on-campus educational activities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (reference)</td>
<td>-0.025</td>
<td>0.018</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.025</td>
<td>0.018</td>
<td>.16</td>
<td>1.051 (1.015-1.089)</td>
</tr>
<tr>
<td><strong>Suspension of public transportation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full suspension (reference)</td>
<td>-0.067</td>
<td>0.028</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Suspension in high-risk areas</td>
<td>0.088</td>
<td>0.028</td>
<td>.002</td>
<td>1.168 (1.106-1.234)</td>
</tr>
<tr>
<td>Normal operation</td>
<td>-0.021</td>
<td>0.028</td>
<td>.46</td>
<td>1.048 (0.992-1.106)</td>
</tr>
<tr>
<td><strong>Contact tracing, isolation, and quarantine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voluntary (reference)</td>
<td>-0.098</td>
<td>0.028</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Compulsory</td>
<td>0.159</td>
<td>0.028</td>
<td>&lt;.001</td>
<td>1.294 (1.225-1.366)</td>
</tr>
<tr>
<td>None</td>
<td>-0.061</td>
<td>0.028</td>
<td>.03</td>
<td>1.038 (0.983-1.097)</td>
</tr>
<tr>
<td><strong>Nucleic acid screening program</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only high-risk units, workplaces, and vulnerable public (reference)</td>
<td>0.083</td>
<td>0.028</td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>Nucleic acid screening for all staff</td>
<td>0.096</td>
<td>0.028</td>
<td>&lt;.001</td>
<td>1.013 (0.959-1.070)</td>
</tr>
<tr>
<td>None</td>
<td>-0.179</td>
<td>0.029</td>
<td>&lt;.001</td>
<td>0.770 (0.728-0.814)</td>
</tr>
<tr>
<td><strong>Mandatory booster vaccination</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universal vaccination (reference)</td>
<td>0.082</td>
<td>0.028</td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>Only high-risk groups are vaccinated (long-term patients, people over 60 years old, etc)</td>
<td>-0.055</td>
<td>0.028</td>
<td>.047</td>
<td>0.872 (0.826-0.921)</td>
</tr>
<tr>
<td>None</td>
<td>-0.026</td>
<td>0.028</td>
<td>.35</td>
<td>0.898 (0.850-0.949)</td>
</tr>
<tr>
<td><strong>Loss of income in 3 months</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0% (reference)</td>
<td>0.541</td>
<td>0.044</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0.426</td>
<td>0.044</td>
<td>&lt;.001</td>
<td>0.891 (0.817-0.972)</td>
</tr>
<tr>
<td>40%</td>
<td>0.207</td>
<td>0.045</td>
<td>&lt;.001</td>
<td>0.716 (0.656-0.782)</td>
</tr>
<tr>
<td>60%</td>
<td>-0.148</td>
<td>0.047</td>
<td>.002</td>
<td>0.564 (0.514-0.617)</td>
</tr>
<tr>
<td>80%</td>
<td>-0.371</td>
<td>0.048</td>
<td>&lt;.001</td>
<td>0.561 (0.510-0.616)</td>
</tr>
<tr>
<td>100%</td>
<td>-0.655</td>
<td>0.051</td>
<td>&lt;.001</td>
<td>0.302 (0.274-0.334)</td>
</tr>
</tbody>
</table>

<sup>a</sup>The results were calculated using the multinomial logit model. A positive sign represents a positive utility for respondents choosing the specific level, and a negative sign represents a negative utility for respondents choosing the specific level.

<sup>b</sup>OR: odds ratio.

<sup>c</sup>Not applicable.
Subgroup Analysis of Preferences for PHSMs

To better trace the heterogeneities of the preferences, subgroup analyses were conducted in terms of age and different monthly income levels (Figure 3). Compared with respondents having high monthly income, those having low monthly income were less sensitive to the risk of infection with COVID-19 within 3 months but more sensitive to the loss of income due to the measure within 3 months. Moreover, low-income respondents cared more about nucleic acid test screening for all and preferred the suspension of public transportation in only high-risk areas. Additionally, low-income respondents preferred not to be close to social and living places, which was in contrast with the findings for high-income respondents.

Figure 3. Subgroup analysis based on sex, age, mental health disease history, and residence status. (A) Subgroup analysis of preferences of respondents with a monthly income of ≤10,000 or >10,000 RMB. A currency exchange rate of 1 RMB=0.138 USD is applicable. (B) Subgroup analysis of preferences of respondents aged ≤35 years or >35 years. (C) Subgroup analysis of preferences of respondents with or without mental health diseases diagnosed previously. (D) Subgroup analysis of preferences of nonmigrant or migrant respondents.

Similarly, respondents older than 35 years were more sensitive to the risk of COVID-19 infection and less sensitive to the loss of income within 3 months. Moreover, compared with older respondents, younger respondents preferred nucleic acid test screening for only high-risk units, workplaces, and vulnerable public, while older respondents preferred screening for all. Furthermore, younger respondents preferred not suspending on-campus educational activities and not closing social and living places.

Respondents diagnosed with mental health diseases did not favor contact tracing, isolation, and quarantine, as well as closure of social and living places compared with those without mental health diseases. The subgroup analysis for migrants and nonmigrants indicated that migrants had less acceptance of the mandatory booster vaccination and accepted the suspension of transportation in high-risk areas or normal operations.
Subgroup Analysis of Pandemic Fatigue and Preference Heterogeneities

A higher pandemic fatigue level was observed in female respondents, younger respondents, migrants, and relatively lower-income respondents (COVID-19 Pandemic Fatigue Scale [CPFS] correlation with age: \( r = -0.274, P < .001 \); correlation with monthly income: \( r = -0.25, P < .001 \) (Table 4). Based on the results of the CPFS, some preference heterogeneities were also found among respondents with a lower or higher level of pandemic fatigue (Figure 4). Respondents with a higher level of fatigue tended to be less sensitive to the risk of COVID-19 infection within 3 months and more sensitive to income loss within 3 months. Additionally, compared with respondents with a lower level of pandemic fatigue, those with a higher level of fatigue preferred the nonsuspension of social places and nonsuspension of on-campus educational activities. Mandatory booster COVID-19 vaccination was also not preferred by respondents with a higher level of pandemic fatigue, while universal COVID-19 booster vaccination was preferred by respondents with a lower level of pandemic fatigue.
Table 4. Results of the COVID-19 Pandemic Fatigue Scale.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Respondents, n</th>
<th>Score, mean (SD)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Correlation&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>All respondents</td>
<td>689</td>
<td>15.24 (5.262)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>341</td>
<td>14.71 (5.350)</td>
<td>.01</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Female</td>
<td>348</td>
<td>15.75 (5.120)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>128</td>
<td>18.28 (4.836)</td>
<td>&lt;.001</td>
<td>-.274</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>26-35</td>
<td>158</td>
<td>15.01 (4.939)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-45</td>
<td>135</td>
<td>14.50 (4.982)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46-55</td>
<td>129</td>
<td>15.08 (5.758)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥56</td>
<td>139</td>
<td>13.55 (4.671)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle school education or below</td>
<td>26</td>
<td>16.54 (5.798)</td>
<td>.18</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>High school education</td>
<td>63</td>
<td>14.87 (3.744)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocational school education</td>
<td>92</td>
<td>14.63 (5.353)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>401</td>
<td>15.07 (5.214)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree</td>
<td>97</td>
<td>16.12 (5.938)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PhD degree</td>
<td>10</td>
<td>17.50 (5.255)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Religion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Christianity</td>
<td>16</td>
<td>13.19 (4.490)</td>
<td>.29</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mohammedanism</td>
<td>3</td>
<td>18.33 (7.506)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buddhism</td>
<td>63</td>
<td>14.89 (5.873)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>2</td>
<td>17.00 (6.272)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>605</td>
<td>15.30 (5.197)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmarried and single</td>
<td>188</td>
<td>17.65 (4.989)</td>
<td>&lt;.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Unmarried and cohabiting</td>
<td>21</td>
<td>16.81 (4.633)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>469</td>
<td>14.08 (4.979)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>9</td>
<td>20.44 (5.175)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Widow</td>
<td>2</td>
<td>18.50 (10.607)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Migrant</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>210</td>
<td>16.12 (5.288)</td>
<td>.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>479</td>
<td>14.85 (5.209)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occupation and working area</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>130</td>
<td>18.58 (4.767)</td>
<td>&lt;.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Manager</td>
<td>95</td>
<td>14.18 (5.357)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technician and associate professional</td>
<td>124</td>
<td>14.42 (5.516)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clerical support worker</td>
<td>105</td>
<td>13.65 (4.218)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service and sales worker</td>
<td>106</td>
<td>15.26 (5.231)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled agricultural, forestry, and fishery worker</td>
<td>28</td>
<td>16.54 (56.215)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plant and machine operator and assembler</td>
<td>50</td>
<td>13.10 (3.754)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>51</td>
<td>15.25 (4.560)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Respondents, n</td>
<td>Score, mean (SD)</td>
<td>P value&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Correlation&lt;sup&gt;b&lt;/sup&gt;</td>
<td>P value&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>----------------------</td>
<td>-------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Monthly income (RMB)&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5000</td>
<td>220</td>
<td>16.76 (4.985)</td>
<td>&lt;.001</td>
<td>–0.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>5000-10,000</td>
<td>289</td>
<td>14.78 (4.768)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10,001-15,000</td>
<td>103</td>
<td>14.45 (5.656)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15,001-20,000</td>
<td>47</td>
<td>13.45 (6.064)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥20,000</td>
<td>30</td>
<td>14.00 (6.623)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of mental health disease</td>
<td></td>
<td></td>
<td>.02</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>42</td>
<td>16.95 (5.635)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>634</td>
<td>15.04 (5.194)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure to novel coronavirus pneumonia</td>
<td></td>
<td></td>
<td>.14</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>198</td>
<td>15.84 (5.613)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>491</td>
<td>14.99 (5.099)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closed-off community management&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>.42</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>90</td>
<td>15.93 (6.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>599</td>
<td>15.13 (5.138)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Mann-Whitney test.

<sup>b</sup>Spearman correlation coefficients for noncontinuous variables.

<sup>c</sup>Not applicable.

<sup>d</sup>Continuous variable.

<sup>e</sup>A currency exchange rate of 1 RMB=0.138 USD is applicable.

<sup>f</sup>Entrance numbers were minimized, checking points were set up in communities, entry permits were limited, face mask wearing was required, health monitoring was enhanced, and only registered personnel and vehicles were allowed to pass through.
Figure 4. Weighted importance of attributes and levels among respondents based on pandemic fatigue levels. (A) Weighted attribute importance among respondents with a lower level of COVID-19 pandemic fatigue. (B) Weighted attribute importance among respondents with a higher level of COVID-19 pandemic fatigue. (C) Relative utility of levels among the 2 groups of respondents.

Latent Class Analysis

According to the AIC and BIC of the LCM, 4 latent groups of respondents were determined, with the lowest BIC value of 9680.29 and AIC value of 9104.82. All the other model fitting values have been presented in Multimedia Appendix 3. The segmented sizes were 15.1%, 16.5%, 34.6%, and 33.8% for groups 1, 2, 3, and 4, respectively. As shown in Figure 2, respondents in groups 1 and 4 attached the most importance to the risk of COVID-19 infection within 3 months, while respondents in groups 2 and 3 attached importance to the loss of income within 3 months. Additionally, following the risk of COVID-19 infection and loss of income within 3 months, groups 2 and 4 considered contact tracing and nucleic acid test screening to be the third and fourth most important attributes,
respectively. Group 3 believed that nucleic acid test screening and suspension of public transportation were very essential.

Discussion

The COVID-19 pandemic posed tremendous challenges for delivering mental and physiological health services throughout China. This study sought to comprehensively investigate public mental health and preferences for PHSMs. This is the first study to estimate public preferences for PHSMs using a DCE for a nationally representative population in China. The risk of COVID-19 infection within 3 months; contact tracing, isolation, and quarantine; nucleic acid screening program; and loss of income within 3 months significantly influenced the preferences for PHSMs.

In our study, we found that the respondents placed the greatest importance on the risk of COVID-19 infection in the last 3 months when considering public health measures for COVID-19 mitigation. With its rapid spread and serious complications, COVID-19 caused fear in the vast majority of people irrespective of whether they were in the risk group. In a population-based survey conducted in America [45], the population was fearful, worried, and uncertain about COVID-19, especially in more densely populated communities, communities with higher presumptive and reported COVID-19 case concentrations, and urban locations. Additionally, an online survey in Italy that asked about health behaviors and the psychological and overall impact of COVID-19 found that only the fear of infection significantly dissuaded people from violating epidemic prevention rules [46]. Hence, the risks of infection and adverse outcomes secondary to infection should be clearly outlined by the media or the government to the public to enhance mutual understanding, reduce their psychological burden, and improve the compliance of people’s epidemic prevention behavior.

Furthermore, the respondents in our survey attached more importance to income loss in their preferences. According to an analysis based on economic forecasts in the European Union, the COVID-19 crisis had an indispensable impact on household disposable income, similar to the one experienced during the 2008-2009 financial crisis, with lower-income households being more severely hit [47]. The high preference may be due to the large negative economic, living, and psychological effects of lower income [48]. This was consistent with the fact that migrant workers, accounting for about one-fifth of the whole Chinese population, were faced with large housing stress and psychological burden from the sudden loss of income and further quarantine enforcement during the COVID-19 pandemic [31]. Therefore, the government should consider subsidies related to epidemic prevention, particularly for the low-income population; take fiscal policy measures as appropriate to reduce the risk and scale of income reduction; and cushion the impact of the epidemic crisis on inequality and poverty through policy interventions.

Respondents in our study showed preference heterogeneity for epidemic prevention measures. Understanding the heterogeneity of information and differences in personal values toward epidemic prevention measures can help policy makers understand individuals’ preferences so that more rational and customized PHSMs can be formulated to reduce the negative emotions caused by epidemic prevention. For example, younger participants preferred not to undergo nucleic acid screening, but older people were more afraid of having novel coronavirus pneumonia. The probable cause is that the case fatality rate of novel coronavirus pneumonia is low in young people and increases in a log-linear model by age among individuals older than 30 years [49]. Therefore, relevant departments should be responsible for community humanistic care, appeasing the mood of the masses, eliminating panic, guiding the community to carry out scientific and orderly epidemic prevention work, implementing vaccine booster shots in the population, and publicizing the scientific knowledge of COVID-19.

Our study showed that migrant workers had a high level of pandemic fatigue related to the suspension of transportation and closure of social places, which aligned with existing literature [31,50]. These findings indicate that vulnerable groups, including migrants and the older population [51], are more prone to experience psychological pressure due to unemployment, suspension of the public transportation network, and loss of income [12,52]. These findings emphasize the importance of psychological placation for susceptible populations during the outbreak to help provide support and managed care for individuals at risk of psychological impact. On the other hand, we found higher pandemic fatigue scores in young participants than in participants from other age groups, which is consistent with the finding in a previous study [53] reporting that university students had significantly reduced mood and reduced social interactions during lockdown periods. For migrant workers, elderly people, and other susceptible populations, governments should develop effective mental health interventions and strategies, carefully assess and manage the mental health needs of vulnerable groups, and provide mental health services through community management or digital platforms during the epidemic.

Although COVID-19 PHSMs are dynamic, our findings contributed to the existing literature by providing a better understanding of the psychological impact of the pandemic, and this may be useful for formulating and planning effective prevention strategies and psychological counseling for the public and susceptible populations. Moreover, the findings of this study may provide insights for PHSM design when managing epidemic outbreaks in the future. Through the analysis of heterogeneous populations that have been affected by the pandemic mentally and emotionally, our research provides key insights that can inform formulation and priority settings and the planning of more effective prevention strategies and psychological support mechanisms. This is particularly relevant for public health authorities and policy makers who are challenged and tasked with conditions involving the physical and mental well-being of the public and vulnerable groups during such crises. Furthermore, the implications of our findings extend far beyond the current pandemic context. As we investigated the psychological effects of COVID-19 and the preferences of various PHSMs for mitigating these impacts, we were able to provide insights that can be pivotal in the face of future infectious disease outbreaks. Moreover, our research
highlights the necessity of incorporating psychological considerations into the priority settings of PHSMs. This approach ensures that interventions are holistic, addressing both the epidemiological and emotional aspects of disease control.

There are limitations in this study, especially in the sampling methods. As we applied quota sampling without providing the quotas of regions in this study owing to budget issues, the results may have a potential bias for inferring the general population, and selection bias may also exist. Moreover, in our study, we collected preference data from 689 respondents living in 31 provinces in China. However, considering the 1.4 billion population in China, the presence of only around 22 respondents in each province may reduce the representativeness of the sampling. Owing to the limited budget for data collection and the restricted offline sampling procedure related to the COVID-19 pandemic lockdown, the flexibility of the sampling was largely limited. Further studies with larger and more representative samples for investigating the mental health of the general population under the conditions of the pandemic may be required to more accurately quantify the perspectives for PHSMs. In addition, the study acknowledges the limitations imposed by the use of quota sampling, particularly the equal representation of sexes and the simplified categorization of age groups, which may not accurately reflect the complex demographics of the adult population in China. Owing to the challenges posed by the pandemic and the budgetary constraints for data collection, the study could not completely adhere to the exact adult age structure of the Chinese population in the sampling methods. This limitation may affect the generalizability of the findings to the entire adult population of China. This is a limitation that future research might overcome with alternative strategies or under different circumstances. Finally, we acknowledge that the DCE questionnaire may impose some cognitive burden on respondents, and this may lead to some biases when selecting among the alternative options. Therefore, a face-to-face approach is considered to be better than an online approach. However, due to the pandemic lockdown, a face-to-face approach was not feasible. In future research performed to understand people’s pandemic fatigue and preferences, a face-to-face approach should be applied if there is no lockdown.

Variability in the preference for COVID-19 policies was found between different groups. Pandemic fatigue and fear of COVID-19 infection contributed to the public’s mental health problems. Hence, at the late stage of the pandemic, policy makers should consider reducing people’s mental burden by introducing approaches to relieve people’s fear of infection when PHSMs are being relaxed. The findings provide insights on PHSM implementation for outbreaks in the future as our research highlights the necessity of incorporating heterogeneous psychological considerations into the priority settings of PHSMs. This may ensure that interventions are holistic, addressing both the epidemiological and emotional aspects of disease control.

Data Availability
The data related to this study can be obtained upon reasonable request to the corresponding author.

Authors' Contributions
TL, ZH, and WM contributed to conceptualization. MY and TL contributed to data curation. TL and MY contributed to the formal analysis. MY and WM contributed to funding acquisition. MY, TL, ZH, and YZ contributed to the investigation. MY, YZ, TL, and ZH contributed to the methodology. WM and TL contributed to project administration. WM contributed to resources. WM and TL contributed to supervision. TL, MY, ZH, and WM contributed to validation. TL and MY contributed to writing-original draft. MY, TL, and ZH contributed to writing-review and editing.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Study survey.
[PDF File (Adobe PDF File), 169 KB - publichealth_v10i1e45840_app1.pdf ]

Multimedia Appendix 2
Sample size calculation (standard parametric approach).
[PDF File (Adobe PDF File), 96 KB - publichealth_v10i1e45840_app2.pdf ]

Multimedia Appendix 3
Akaike Information Criterion, Bayesian Information Criterion, and other fitting values.
[PDF File (Adobe PDF File), 131 KB - publichealth_v10i1e45840_app3.pdf ]

References


Abbreviations

AIC: Akaike Information Criterion
BIC: Bayesian Information Criterion
CPFS: COVID-19 Pandemic Fatigue Scale
DCE: discrete choice experiment
LCM: latent class model
MNL: multinomial logit
OR: odds ratio
PHSM: public health and social measure
Quantifying Disparities in COVID-19 Vaccination Rates by Rural and Urban Areas: Cross-Sectional Observational Study

Wenyong Dong1, Dr Med; Yudong Miao2, PhD; Zhanlei Shen2, DrPH; Wanliang Zhang2, MD; Junwen Bai2, MD; Dongfang Zhu2, MD; Ruizhe Ren2, MD; Jingbao Zhang3, MD; Jian Wu3, PhD; Clifford Silver Tarimo3, DrPH; Theodora Ojangba2, DrPH; Yi Li2, MD

1Department of Hypertension, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou, China
2Department of Health Management, College of Public Health, Zhengzhou University, Zhengzhou, China
3Department of Science and Laboratory Technology, Dar Es Salaam Institute of Technology, Dar es Salaam, United Republic of Tanzania

Corresponding Author:
Yi Li, MD
Department of Health Management
College of Public Health
Zhengzhou University
No. 100 Science Avenue, High-Tech Industrial Development Zone, Zhengzhou City, Henan Province
Zhengzhou, 450001
China
Phone: 86 13298157145
Email: ly07263026@163.com

Abstract

Background: Vaccination plays an important role in preventing COVID-19 infection and reducing the severity of the disease. There are usually differences in vaccination rates between urban and rural areas. Measuring these differences can aid in developing more coordinated and sustainable solutions. This information also serves as a reference for the prevention and control of emerging infectious diseases in the future.

Objective: This study aims to assess the current coverage rate and influencing factors of COVID-19 (second booster) vaccination among Chinese residents, as well as the disparities between urban and rural areas in China.

Methods: This cross-sectional study used a stratified random sampling approach to select representative samples from 11 communities and 10 villages in eastern (Changzhou), central (Zhengzhou), western (Xining), and northeast (Mudanjiang) Mainland China from February 1 to February 18, 2023. The questionnaires were developed by experienced epidemiologists and contained the following: sociodemographic information, health conditions, vaccine-related information, information related to the Protective Motivation Theory (PMT), and the level of trust in the health care system. Vaccination rates among the participants were evaluated based on self-reported information provided. Binary logistic regression models were performed to explore influencing factors of vaccination among urban and rural participants. Urban-rural disparities in the vaccination rate were assessed using propensity score matching (PSM).

Results: A total of 5780 participants were included, with 53.04% (3066/5780) being female. The vaccination rate was 12.18% (704/5780; 95% CI 11.34-13.02) in the total sample, 13.76% (341/2478; 95% CI 12.40-15.12) among the rural participants, and 10.99% (363/3302; 95% CI 9.93-12.06) among the urban participants. For rural participants, self-reported health condition, self-efficacy, educational level, vaccine knowledge, susceptibility, benefits, and trust in the health care system were independent factors associated with vaccination (all \( P < .05 \)). For urban participants, chronic conditions, COVID-19 infection, subjective community level, vaccine knowledge, self-efficacy, and trust in the health care system were independent factors associated with vaccination (all \( P < .05 \)). PSM analysis uncovered a 3.42% difference in vaccination rates between urban and rural participants.

Conclusions: The fourth COVID-19 vaccination coverage rate (second booster) among the Chinese population was extremely low, significantly lower than the previous vaccine coverage rate. Given that COVID-19 infection is still prevalent at low levels, efforts should focus on enhancing self-efficacy to expand the vaccine coverage rate among the Chinese population. For rural residents, building awareness of the vaccine’s benefits and improving their overall health status should be prioritized. In urban areas, a larger proportion of people with COVID-19 and patients with chronic illness should be vaccinated.

https://publichealth.jmir.org/2024/1/e50595
Introduction

Vaccination plays an important role in preventing COVID-19 infection, reducing the severity of the disease and the case fatality rate [1-4]. Currently, the prevalence of COVID-19 infection rate is low [5]. However, COVID-19 still has a significant impact on people's health, causing issues such as long COVID-19, preterm birth, stillbirth, myocarditis, and pericarditis [6-8]. A study from Australia has shown that vaccination is effective in protecting the population [9]. Similar studies carried out in the United States indicated that COVID-19 vaccines have reduced the severity of the disease and prevented serious complications such as respiratory failure and death [10]. Other studies have shown that individuals who received booster doses of COVID-19 vaccine have stronger immune protection than those who received only a single dose [11,12]. Therefore, for countries and regions aiming to reduce the spread of COVID-19 by expanding the protection of immune barriers, increasing vaccination rates remains an important task for health care systems at this stage [13].

Establishing herd immunity through COVID-19 vaccination requires a vaccination rate ranging from 50% to 85% of the population [14,15]. An intensive vaccination campaign during the initial phase of the pandemic wave led to a lower optimal level of doses administered per 100 inhabitants (roughly 47 doses of vaccines administered) to reduce the number of infected individuals; however, as the pandemic wave progressed, the optimal level of vaccines increased to about 90 doses per 100 inhabitants to produce the same effect [16]. The maximum vaccination rate that could be achieved without a mandate was 70%, with the remaining 30% associated with people’s hesitancy toward vaccinations [17]. The factors influencing COVID-19 vaccination are complex and only partially explored. Previous studies have shown that COVID-19 vaccination was related to the severity of the outbreak, perceived susceptibility to the disease, and concerns related to the vaccine’s safety and effectiveness [18-21]. Our previous research reported correlations between COVID-19 vaccination and trust in the health care system, vaccine accessibility, lifestyle, and psychological experience [22,23]. Additionally, our previous nationwide investigation conducted during primary COVID-19 vaccination hesitancy confirmed a 2.38% gap between rural residents in China and their urban compatriots [24]. Similar findings were reported in studies related to influenza and HPV vaccines [25,26].

Disparities in COVID-19 vaccination rates between rural and urban regions are meaningful [27]. In China and other limited-income countries, frequent movement of the population between urban and rural areas is common [28]. Given the extremely high transmission capacity of COVID-2019, increasing vaccination rates among urban or rural residents unilaterally is not cost-effective in building herd immunity [29]. Currently, India, Brazil, and other limited-income countries are experiencing significant internal migration patterns similar to China’s [30,31]. Effective public governance improves prevention and preparedness to face pandemic threats [32-34]. Quantifying disparities in COVID-19 vaccination rates in rural and urban areas based on Chinese national-level evidence is crucial for global efforts to end the COVID-19 pandemic and can provide a reference for the prevention and control of emerging infectious diseases.

Therefore, we conducted a nationwide survey during the fourth COVID-19 (second booster) vaccination period. This study aims to assess the current coverage rate of COVID-19 vaccination among Chinese residents, identify influencing factors, and examine disparities between urban and rural areas in China. This study’s findings will reveal the real-time status of COVID-19 vaccination among people in China. A quantitative assessment of the disparity between urban and rural areas will deepen our understanding of the complexity of COVID-19 infection prevention, helping produce a more coordinated and sustainable solution to end the pandemic.

Methods

Sample and Data

From February 1 to February 18, 2023, we selected 11 communities and 10 villages in eastern, central, western, and northeastern Mainland Chinese as representative samples using stratified random sampling. Within each region, namely Changzhou, Zhengzhou, Xining, and Mudanjiang, a random sampling approach was employed to select representative urban and rural areas. Specifically, 2 or more communities and villages were randomly chosen from each city. Furthermore, to ensure comprehensive coverage, households were also randomly selected within each region. Finally, all members of the selected family (aged ≥18 years) participated in the survey and completed a web-based or paper-based questionnaire with the investigator’s assistance. To ensure data consistency and reliability, questionnaires that contained contradictory responses to key questions were excluded. Additionally, questionnaires completed in less than 5 minutes were excluded to maintain an adequate level of response detail. Furthermore, to prevent duplicate entries and potential bias in the analysis, questionnaires that were filled out repeatedly by the same individuals were excluded. Finally, questionnaires that lacked clear identification of the survey object’s source were excluded to ensure transparency and accuracy in the data collection process. Following the exclusion criteria, the questionnaires underwent a thorough review process conducted by trained staff. From this review, a total of 5891 questionnaires were collected. Subsequently, after further screening and data validation, a final sample of 5780 participants was included in our study. Figure 1 shows the study flowchart.
Measures of Variables

Experienced epidemiologists developed the questionnaires, which contained the following components: sociodemographic information (sex, age, religion, marital status, educational level, career, social status in China, and social status in community), health conditions (chronic disease, history of allergy, self-reported health condition, and COVID-19 infection), vaccine-related information (vaccine knowledge and vaccine accessibility), information related to the Protective Motivation Theory (PMT; severity, susceptibility, benefits, barriers and self-efficacy), and the level of trust in the health care system (trust in doctors and vaccine developers).

We assigned a unique code to each participant, which was used to distinguish whether the participant came from an urban or a rural area.

COVID-19 vaccination status was measured based on participants' responses to 3 items. The first item asked, “Have you been vaccinated with the COVID-19 basic vaccine?” We set 3 answers: 1 (“Yes, I have completed COVID-19 basic vaccination”) 2 (“Yes, but I haven't completed COVID-19 basic vaccination”), and 3 (“No”). The second question was, “Have you been vaccinated with the third COVID-19 (first booster) vaccine?” We set 2 answers: 1 (“Yes”) and 2 (“No”). The third question was, “Have you been vaccinated with the fourth COVID-19 (second booster) vaccine?” We set 2 answers: 1 (“Yes”) and 2 (“No”). We defined participants who completed all 3 items as those who had been vaccinated with the fourth COVID-19 (second booster) vaccine.

Data Analysis Procedure

An independent sample t test or chi-square test was carried out to test the differences in COVID-19 vaccination across groups between the urban and rural participants. Binary logistic regression models were performed to explore the factors influencing vaccination among the urban and rural participants. Both univariate and multivariate analyses for urban and rural participants were conducted separately. PSM was used to minimize potential confounding biases. A probit regression model was used to estimate the propensity scores for urban and rural participants. Finally, 1956 pairs of homogeneous participants were matched in a 1:1 ratio using propensity scores from a total of 5780 participants. Differences were deemed statistically significant if \( P < .05 \). We performed all statistical analyses using Stata software (version 16.1; StataCorp).

Ethical Considerations

This study was reviewed and approved by the Life Science Ethics Review Committee of Zhengzhou University (2021-01-12-05). Written informed consent outlining the study’s objectives was obtained from each participant before the survey commenced. All data were used solely for research purposes. The study data are anonymous, and the answers are protected by privacy laws.
**Results**

**Prevalence Of COVID-19 Vaccination and Characteristics Among Urban and Rural Participants**

A total of 5780 participants, among them 57.13% (3302/5780) urban residents, completed the survey. A summary of COVID-19 vaccination rates, along with characteristics like sociodemographic information, health conditions, vaccine-related information, information related to PMT, and trust in the health care system between the urban and rural participants is shown in Table S1 in Multimedia Appendix 1. In the total sample, 12.18% (704/5780) of participants (95% CI 11.34-13.02) expressed they had been vaccinated with the fourth COVID-19 (second booster) vaccine. Additionally, 10.99% (363/3302) of urban participants (95% CI 9.93-12.06) had a relatively lower COVID-19 vaccination rate than rural participants at 13.76% (341/2478; 95% CI 12.40-15.12). Individuals aged 30 to 39 years, without chronic disease or previous COVID-19 infection, who had a higher subjective community level, higher level of vaccine knowledge, higher perceived benefits, higher self-efficacy, and trust doctors and vaccine developers (at levels 1 and 4) were more likely to receive a COVID-19 vaccination in urban areas (all P<.05). Conversely, individuals aged 18 to 29 years, without chronic disease, with vaccine accessibility within 15 minutes, higher self-reported health condition, lower level of vaccine knowledge, lower perceived benefits, higher perceived barriers, lower self-efficacy, and lower trust in doctors and vaccine developers were more likely to receive a COVID-19 vaccination in rural areas (all P<.05).

Additionally, we discovered that rural participants tended to be older and less educated, with fewer reporting a history of COVID-19 infection compared to urban participants. In both rural and urban participants, the proportion of female (n=1488, 45.06%) and male (n=1226, 49.48%) participants was relatively similar. Moreover, 45.34% (n=1497) of urban participants and 38.58% (n=956) of rural participants took less than 15 minutes to reach the vaccination site (Table S2 in Multimedia Appendix 1).

**Factors Influencing COVID-19 Vaccination Among Urban and Rural Participants**

After adjusting for potential confounding variables, we discovered that urban participants without chronic disease (adjusted odds ratio [AOR] 1.629, 95% CI 1.108-2.396), without COVID-19 infection (AOR 1.977, 95% CI 1.977-2.479), who had a higher subjective community level (AOR 1.073, 95% CI 1.013-1.136), level 2 self-efficacy (AOR 2.162, 95% CI 1.550-3.014), and level 3 self-efficacy (AOR 1.842, 95% CI 1.120-3.029) had a higher COVID-19 vaccination rate. Participants from urban areas who had level 4 vaccine knowledge (AOR 0.667, 95% CI 0.476-0.935), level 2 trust in doctors (AOR 0.530: 95% CI 0.377-0.743), and level 3 trust in vaccine developers (AOR 0.638, 95% CI 0.425-0.959) had a lower COVID-19 vaccination rate. The results showed that rural participants who had a higher self-reported health condition (AOR 1.009, 95% CI 1.001-1.017), level 2 self-efficacy (AOR 2.524, 95% CI 1.753-3.633), and level 3 self-efficacy (AOR 4.162, 95% CI 2.369-7.315) had a higher COVID-19 vaccination rate. Participants from rural areas who had a university degree or above (AOR 0.619, 95% CI 0.427-0.898), level 2 vaccine knowledge (AOR 0.680, 95% CI 0.501-0.923), level 4 vaccine knowledge (AOR 0.522, 95% CI 0.351-0.776), level 2 susceptibility (AOR 0.591, 95% CI 0.416-0.840), level 2 benefits (AOR 0.537, 95% CI 0.376-0.765), level 3 benefits (AOR 0.539, 95% CI 0.312-0.933), level 2 trust in doctors (AOR 0.665, 95% CI 0.456-0.969), level 3 trust in doctors (AOR 0.545, 95% CI 0.340-0.874), level 4 trust in doctors (AOR 0.364, 95% CI 0.200-0.661), and level 4 trust in vaccine developers (AOR 0.489, 95% CI 0.253-0.945) had a lower COVID-19 vaccination rate (Table S3 in Multimedia Appendix 1).

**Propensity Score Matching Analysis**

In total, 3912 samples were captured using propensity score matching (PSM) from the 5780 participants. After PSM for sex, age, religion, career, self-reported health condition, educational level, COVID-19 infection, vaccine accessibility, vaccine knowledge, susceptibility, barriers, trust in doctors, and trust in vaccine developers, no statistically significant discrepancies were discerned between the urban and rural participants in all covariates (all P>.05). The balance test and common support domain of PSM for urban and rural samples are shown in Table S4 and Figure S1 in Multimedia Appendix 1. Based on the balanced samples, differences in COVID-19 vaccination rates among the urban and rural participants were assessed. As shown in Figure 2, the prevalence of COVID-19 vaccination among rural participants (14.57%; 95% CI 13.01-16.14) was still higher than that among urban participants (11.15%; 95% CI 9.75-12.54) by 3.42% (P<.05) after PSM. In addition, the vaccination rate of COVID-19 in uninfected participants was higher than that of participants who were infected in most age groups (Figure 3).
**Discussion**

**Principal Findings**

This study showed that 12.18% (704/5780) participants completed the fourth COVID-19 (second booster) vaccination, with 10.99% (363/3302) urban participants and 13.76% (341/2478) rural participants completing the fourth COVID-19 (second booster) vaccination, respectively. Vaccine knowledge, self-efficacy, and trust in the health care system had an impact on vaccination rates among both urban and rural residents. Educational level, self-reported health condition, susceptibility, and benefits had an impact on the vaccination rate among rural residents. According to the PSM analysis, a 42% disparity in vaccine coverage rates was confirmed between urban and rural participants. In addition, the vaccination rate in uninfected participants was higher than that of participants who were infected in most age groups.

Our study unveiled differences in vaccination coverage rates between urban and rural residents, with rural residents having a higher COVID-19 vaccination rate than urban residents. This finding is consistent with those of previous studies, not only for COVID-19 but also for other vaccines [35-37]. Prior research has shown that the vaccination rate in rural areas is higher than that in urban areas [38,39]. In a survey of men who have sex
with men in England, the vaccination rate for human papillomavirus (HPV) was higher in rural areas than in urban areas [38]. A COVID-19 vaccination coverage survey conducted among US veterans showed that the vaccination rate of rural veterans was higher than that of urban veterans [39]. On the other hand, some research results showed higher vaccination rates in urban areas compared to rural areas [40,41]. In a survey of COVID-19 vaccination in the United States, the findings showed that the vaccination rate in urban areas was higher compared to rural areas [40]. In a survey of influenza vaccination among adults in the United States, results showed that the vaccination rate was higher in cities than in rural areas [36]. A survey conducted among adolescents in the United States assessing the use of vaccination services for vaccines, such as MenACWY (meningococcal serogroups A, C, W, and Y), Tdap (tetanus, diphtheria, pertussis), HPV, and influenza, revealed that urban adolescents had a higher rate of utilization compared to their rural counterparts [41].

Differences in vaccination rates between urban and rural areas can be influenced by many factors [42]. We believe that there are several reasons that the rural residents had a higher COVID-19 vaccination rate than their urban counterparts. First, due to the unbalanced allocation of medical resources, rural residents have a higher mortality rate related to COVID-19, and they are more willing to be vaccinated for their own safety [43,44]. Second, with the development of urbanization in China, rural population mobility and the increase in migrant workers have led to a smaller rural population base, resulting in a higher vaccination rate [45]. Third, the Chinese government has begun paying more attention to health issues faced by the rural older adult population and has taken relevant measures to ensure that every rural older adult is vaccinated [46]. Finally, the decision to vaccinate is also easily influenced by information found on the internet. Because rural residents have lower education levels, they have higher trust in web-based information and a stronger intention to vaccinate than urban residents [47-49].

Our results showed that several factors, including vaccine knowledge level, self-efficacy, trust in medical staff, and trust in vaccine developers, had an impact on the decision to vaccinate among urban and rural residents. Studies have shown that an individual's comprehension of vaccine knowledge influences their decision to get vaccinated [50,51]. Individuals who lack sufficient vaccine knowledge may be hesitant about vaccination [52]. The same applies to other types of vaccines, such as influenza vaccines, where individuals with higher levels of vaccine knowledge are more likely to receive the influenza vaccine [53]. This is similar to the results of other studies showing that an individual's self-efficacy influences their decision to receive vaccinations. Higher self-efficacy encourages individuals to get vaccinated [54,55]. Similarly, trust in doctors and vaccine developers also affects an individual's decision to get vaccinated [56,57]. A survey in Germany showed that trust in the medical establishment impacts the intention to receive the COVID-19 vaccine [58]. Confidence in public service delivery also leads to favorable responses to mass immunization efforts [59].

In summary, when formulating a vaccination plan, it is important to focus on early promotion and publicity, highlighting the necessity, safety, effectiveness, and other relevant information about the vaccine. In addition, emphasis should be placed on improving residents' self-efficacy and on doctor-patient communication to increase vaccination rates [60].

Furthermore, our analysis revealed that in both urban and rural areas, the rate of fourth COVID-19 (second booster) vaccination was generally higher among residents without a history of infection compared to those who had been previously infected with COVID-19. SARS-CoV-2 constantly changes through mutations that generate variants [61]. A study showed that the antibody response produced after vaccination better neutralizes certain prevalent variants [62]. A statistically significant decreased risk for reinfection was found among individuals who were previously infected and then vaccinated versus those who were previously infected but remained unvaccinated [63]. In addition, there was a decreased risk for symptomatic disease among previously infected and vaccinated persons compared with those who were not vaccinated after infection [64]. Therefore, even if residents have a history of infection, they should be vaccinated again to safeguard against future infections. We suggest that when promoting booster vaccinations, it should be emphasized that individuals should receive the booster even if they have been infected with SARS-CoV-2 and have acquired antibodies.

**Strengths and Limitations**

This study was the first nationwide survey on the vaccination status of residents in China during the fourth COVID-19 (second booster) vaccination campaign. Given the complexity of factors influencing COVID-19 vaccination, we used PSM for the first time to mitigate confounding variables and quantify disparities in COVID-19 vaccination rates in rural and urban areas.

Additionally, this study used 1:1 matching with PSM, highlighting 2 advantages. First, the PSM method ensured that all indicators were collected from homogeneous respondents from both urban and rural areas. Second, 1:1 matching maximized the utilization of group information to ensure that individuals within each group were closely matched, enhancing the strength of the principal findings.

However, this study has numerous limitations that must be acknowledged. First, the assessment of COVID-19 vaccination relied on self-reported questionnaires, which inevitably introduced subjectivity, self-bias, and response bias in the data collection. Second, while the study employed the PSM framework to comprehensively collect COVID-19 vaccination covariates, it is possible that some unknown related factors were not collected.

**Conclusion**

The fourth COVID-19 (second booster) vaccination coverage rate among the Chinese population was exceptionally low, significantly lower than previous vaccine coverage rates. With COVID-19 infections still prevalent at a low level, efforts should be concentrated on enhancing self-efficacy and building trust in the health care system to expand vaccine coverage among the Chinese population. For rural residents, building confidence in vaccination benefits and improving overall health status must be prioritized. In urban areas, focusing on vaccinating a larger
proportion of people with COVID-19 and patients with chronic disease is crucial. In conclusion, the vaccination rate is higher in rural areas than in urban areas. Therefore, attention should be given to urban-rural differences and targeted vaccination plans should be formulated when implementing vaccination programs. The results of this study offer valuable insights for future responses to emerging major infectious diseases.

Acknowledgments
This study was funded by the National Social Science Fund of China (21BGL222), the Collaborative Innovation Key Project of Zhengzhou (20XTZX05015), and the Henan Provincial University Philosophy and Social Science Innovation Talent Support Plan.

Data Availability
All data can be acquired by contacting the corresponding author.

Authors' Contributions
YM and WD conceived, designed, and supervised the study. YL, WZ, ZS, JB, DZ, RR, and JZ participated in the data collection, YM and YL analyzed the data and prepared the tables and figures. YM and YL drafted the original manuscript. WD and JW acquired the funding. WD, YM, and JW administrated the project. YM, ZS, WD, TO, and CST reviewed and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Additional Tables and Figures.

References


Abbreviations

AOR: adjusted odds ratio

HPV: human papillomavirus

Men ACWY: meningococcal serogroups A, C, W, and Y

PMT: Protective Motivation Theory

PSM: propensity score matching

Tdap: tetanus, diphtheria, pertussis

©Wenyong Dong, Yudong Miao, Zhanlei Shen, Wanliang Zhang, Junwen Bai, Dongfang Zhu, Ruizhe Ren, Jingbao Zhang, Jian Wu, Clifford Silver Tarimo, Theodora Ojangba, Yi Li. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 19.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Smoking Behavior Change and the Risk of Heart Failure in Patients With Type 2 Diabetes: Nationwide Retrospective Cohort Study

Jung Eun Yoo¹, MD, PhD; Su-Min Jeong², MD; Kyu Na Lee³, BS; Heesun Lee⁴, MD, PhD; Ji Won Yoon⁵, MD, PhD; Kyungdo Han³*, PhD; Dong Wook Shin⁶,7,8*, MD, MBA, DrPH

¹Department of Family Medicine, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Republic of Korea
²Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
³Department of Statistics and Actuarial Science, Soongsil University, Seoul, Republic of Korea
⁴Division of Cardiology, Department of Internal Medicine, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Republic of Korea
⁵Division of Endocrinology and Metabolism, Department of Internal Medicine, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Republic of Korea
⁶Department of Family Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
⁷Supportive Care Center, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
⁸Department of Clinical Research Design & Evaluation, Samsung Advanced Institute for Health Science & Technology, Sungkyunkwan University, Seoul, Republic of Korea
* these authors contributed equally

Corresponding Author:
Dong Wook Shin, MD, MBA, DrPH
Department of Family Medicine
Samsung Medical Center
Sungkyunkwan University School of Medicine
81 Irwon-ro, Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 5252
Fax: 82 2 3410 0388
Email: dwshin.md@gmail.com

Abstract

Background: Heart failure (HF) is one of the most common initial manifestations of cardiovascular disease in patients with type 2 diabetes. Although smoking is an independent risk factor for HF, there is a lack of data for the incidence of HF according to changes in smoking behaviors in patients with type 2 diabetes.

Objective: We aimed to examine the association between interval changes in smoking behavior and the risk of HF among patients with type 2 diabetes.

Methods: We conducted a retrospective cohort study using the National Health Insurance Service database. We identified 365,352 current smokers with type 2 diabetes who had 2 consecutive health screenings (2009-2012) and followed them until December 31, 2018, for the incident HF. Based on smoking behavior changes between 2 consecutive health screenings, participants were categorized into quitter, reducer I (≥50% reduction) and II (<50% reduction), sustainer (reference group), and increaser groups.

Results: During a median follow-up of 5.1 (IQR 4.0-6.1) years, there were 13,879 HF cases (7.8 per 1000 person-years). Compared to sustainers, smoking cessation was associated with lower risks of HF (adjusted hazard ratio [aHR] 0.90, 95% CI 0.86-0.95), whereas increasers showed higher risks of HF than sustainers; heavy smokers who increased their level of smoking had a higher risk of HF (aHR 1.13, 95% CI 1.04-1.24). In the case of reducers, the risk of HF was not reduced but rather increased slightly (reducer I: aHR 1.14, 95% CI 1.08-1.21; reducer II: aHR 1.03, 95% CI 0.98-1.09). Consistent results were noted for subgroup analyses including type 2 diabetes severity, age, and sex.
Conclusions: Smoking cessation was associated with a lower risk of HF among patients with type 2 diabetes, while increasing smoking amount was associated with a higher risk for HF than in those sustaining their smoking amount. There was no benefit from reduction in smoking amount.

(JMIR Public Health Surveill 2024;10:e46450) doi:10.2196/46450

KEYWORDS

smoking; change in smoking behavior; cessation; heart failure; type 2 diabetes; diabetes; cardiovascular disease; smoking cessation; smoker; risk factor

Introduction

Background

We are currently in a global type 2 diabetes public health crisis; an estimated 426 million individuals have type 2 diabetes [1]. Patients with type 2 diabetes are at increased risk of cardiovascular disease, and heart failure (HF) is one of the most common initial manifestations of cardiovascular disease in patients with type 2 diabetes [2]. Furthermore, type 2 diabetes contributes greatly to the increased morbidity and mortality associated with HF [3]. Thus, identifying modifiable risk factors for HF is critical in this population at high risk of type 2 diabetes.

Smoking is the leading predictor of death in patients with type 2 diabetes [4] and also an independent risk factor for HF and contributes to coronary heart disease, a major cause of HF [5-7]. In particular, smoking has been linked with several factors associated with left ventricular hypertrophy and cardiac dysfunction [8,9]. In addition, recent studies have found that individuals who smoke have at least twice the risk of developing HF than those who never smoke [8,10,11]. The risk is almost 4-fold among smokers who smoke ≥25 cigarettes per day [10].

Limited evidence is currently available, demonstrating that smoking cessation is associated with a reduced risk of incident HF and adverse events related to HF [5]. Some studies have reported that the risks of HF decreased over time after smoking cessation [11,12]. Those who ceased smoking for over 30 years experienced a similar HF risk to never smokers [13]. Former smokers also had similar cardiac structure and function as measured by echocardiography when compared with never smokers [9].

Given the increased risk for HF posed by type 2 diabetes, defining the effect of smoking cessation for HF in a type 2 diabetes cohort is important. Indeed, smoking and type 2 diabetes interrelate in the generation of cardiovascular events [14]. However, most studies have not focused on this high-risk population, and the HF incidence associations between current smoking and smoking cessation with type 2 diabetes are still unknown. Indeed, although smoking behavior can change over time, previous studies measured the association at only 1 timepoint [8-12,15,16], and there is a lack of data for the incidence of HF according to changes in smoking frequency among current smokers. One additional limitation of the previous studies is that the study population was comprised of only older individuals (aged ≥65 years [11]). Another limitation was the small sample sizes and the small number of outcomes of interest. The largest previous study had 3874 HF events among 188,167 individuals [10].

Study Design and Aim

To fill this knowledge gap, we conducted a nationwide cohort study to investigate whether changes in smoking behavior among patients with type 2 diabetes resulted in subsequent altered risk of HF. By measuring smoking behavior repetitively, we were able to investigate the effects of reduction, increase, and cessation of smoking.

Methods

Study Setting

We used a database provided by the National Health Insurance Service (NHIS), a single insurer in Korea [17]. The NHIS provides a mandatory universal insurance system that covers ~97% (about 51 million) of the Korean population and medical aid to ~3% (about 1.5 million) of the population in the lowest income bracket. NHIS provides a free biennial cardiovascular health screening for all insured individuals. Hence, the NHIS retains an extensive data set of the entire Korean population that includes demographics; medical treatment and procedures; disease diagnosis according to the International Classification of Diseases, 10th revision (ICD-10) codes; and the results of health screenings consisting of a self-questionnaire on health behavior, anthropometric measurements, and laboratory test results.

Study Population

Among individuals who underwent a health screening in 2009-2012, we identified 2,745,689 individuals with type 2 diabetes: (1) those with a previous history of type 2 diabetes which was defined as ICD-10 code (E11-14) diagnosis with at least 1 claim for prescription of antidiabetic agents before the health screening or (2) those with fasting plasma glucose 126 mg/dL at the health screening. This definition was based on the consensus of relevant findings widely used in previous studies [18,19]. We then selected current smokers (n=758,049) according to the World Health Organization definition of current smoker [20]. Among these, 485,547 underwent a follow-up health examination within 2 years. We excluded those who had previously been diagnosed with (1) any cancer (n=54,192); (2) myocardial infarction, atrial fibrillation, or valvular heart diseases (n=18,205); or (3) HF (n=9757) before the second health screening. Those with at least 1 missing variable used in the study were also excluded (n=33,891). To reduce the effect of reverse causality, we applied a 1-year lag time by excluding participants who were diagnosed with HF (n=2463) and who

https://publichealth.jmir.org/2024/1/e46450
A total of 365,352 participants were included in the primary analysis (Figure 1).

**Flowchart for participant recruitment.**

- **Participants (≥20 years) with type 2 diabetes who underwent health screening in 2009-2012 (n=2,745,689)**
- **Participants who currently smoke (n=758,049)**
- **Participants who had a follow-up screening within 2 years (n=485,547)**
  - Previous history
    - Cancer (n=54,192)
    - Myocardial infarction, atrial fibrillation or valvular heart disease (n=18,205)
    - Heart failure (n=9757)
  - Incomplete information (n=33,891)
  - Follow-up <1 year (n=4150)
- **Participants eligible for inclusion (n=365,352)**
- **Followed up from 1 year after the second screening date to the date of incident heart failure, death or outmigration, or the end of the study period (December 31, 2018), whichever came first.**

**Ethical Considerations**

This study was approved by the institutional review board of Samsung Medical Center (IRB SMC 2022-07-072). The review board waived the requirement for written informed consent from the patients because the data are public and anonymized under confidentiality guidelines.

**Definition of Change in Amount of Smoking Cigarettes**

Information on smoking status was obtained from a self-administered questionnaire completed as a part of the NHIS health screening. The participants who identified themselves as current smokers were asked to provide their daily smoking amount (average number of cigarettes per day) and smoking duration in years. According to the number of cigarettes per day at the time of the first screening, participants were categorized into three groups: (1) light smokers (<10 cigarettes per day), (2) moderate smokers (10-19 cigarettes per day), and (3) heavy smokers (≥20 cigarettes per day). Each of the 3 groups was then divided into 5 subgroups by comparing the number of cigarettes per day between the first screening and the follow-up screening. The subgroups were (1) quitters (those who completely ceased smoking), (2) reducers I (those who reduced the number of cigarettes by 50% or more), (3) reducers II (those who reduced the number of cigarettes by 20% or more and less than 50%), (4) sustainers (those who reduced the number of cigarettes by less than 20% or increased by less than 20%), and (5) increasers (those who increased the number of cigarettes by 20% or more).

**Outcome: Ascertainment of HF and Follow-Up**

The end point of the study was a newly diagnosed case of HF. HF was defined as the first hospitalization under a primary diagnosis of ICD-10 code I50 as used in previous studies [21-23]. The study participants were followed up from 1 year after the second screening date to the date of incident HF, death or outmigration, or the end of the study period (December 31, 2018), whichever came first.

**Covariates**

We considered socioeconomic position as a potential covariate including income level and place of residence. Household income was categorized into quartiles based on insurance premium levels with those covered by Medical Aid being merged into the lowest income quartile. Alcohol consumption was classified as none, mild (<15 g of alcohol per day), moderate (15-30 g of alcohol per day), and heavy (≥30 g of alcohol per day). Regular exercise was defined as >30 minutes of moderate physical activity at least 5 times per week or >20 minutes of strenuous physical activity at least 3 times per week [24]. BMI was calculated as subject weight (kg) divided by square of height (m$^2$).
Hypertension was defined as a history of a claim (I10-13 or 15 codes) and antihypertensive medication or systolic blood pressure ≥140 mm Hg or diastolic blood pressure ≥90 mm Hg. Dyslipidemia was defined as a history of a claim with E78 codes and lipid-lowering medications, or total cholesterol level ≥240 mg/dL. Chronic kidney disease was defined as a glomerular filtration rate <60 mL/minute per 1.73 m² as estimated by the Modification of Diet in Renal Disease equation. Chronic obstructive pulmonary disease was defined as a claim with J41-44 codes, and stroke was defined as a claim with codes I63-64.

**Statistical Analysis**

Cox proportional hazards regression analysis was performed to evaluate the association between smoking behavior change and incident HF. Model 1 was unadjusted. Model 2 was adjusted for age, sex, income, area of residence, alcohol consumption, duration of smoking, physical activity, BMI, and comorbidities (hypertension, dyslipidemia, chronic kidney disease, chronic obstructive pulmonary disease, and stroke). Model 3 was further adjusted for fasting glucose level, duration of type 2 diabetes, and insulin use. Stratification analyses by smoking levels at the first examination (light, moderate, and heavy smoker), type 2 diabetes severity (duration of type 2 diabetes: new-onset, <5 years, and 5 years, the number of oral antidiabetic agents taken: 0, 1-2, and 3, and use of insulin), age (20-64 years and 65 years), and sex were performed to assess the association of change in smoking behavior with incidence of HF. Statistical analyses were performed using SAS (version 9.4; SAS Institute Inc), and a P value <.05 was considered statistically significant.

**Results**

**Baseline Characteristics of the Study Population**

Table 1 shows the baseline characteristics according to the change in smoking behavior. During the 2 years before initiation of the study, 44.5% (n=162,645) of current smokers sustained their smoking amount, 18.6% (n=67,843) became quitters, 20.7% (n=75,615) reduced their smoking amount (n=27,987, 7.7% and n=47,628, 13% in reducer I and reducer II groups, respectively), and the remaining 16.2% (n=59,249) increased their smoking amount. Compared to sustainers, quitters tended to be older, women, nondrinkers, and light smokers, engage in more regular exercise, have more comorbid conditions and longer duration of type 2 diabetes, and use multiple oral antidiabetic agents and insulin.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Total (N=365,352)</th>
<th>Smoking behavior change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quitter (n=67,643)</td>
<td>Reducer I (n=27,987)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>51.2 (11.1)</td>
<td>51.5 (11.0)</td>
</tr>
<tr>
<td>Sex (male), n (%)</td>
<td>347,685 (95.2)</td>
<td>26,216 (93.7)</td>
</tr>
<tr>
<td>Income, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 (lowest)</td>
<td>67,489 (18.5)</td>
<td>5954 (21.3)</td>
</tr>
<tr>
<td>Q2</td>
<td>70,644 (19.3)</td>
<td>5776 (20.6)</td>
</tr>
<tr>
<td>Q3</td>
<td>107,573 (29.4)</td>
<td>7890 (28.2)</td>
</tr>
<tr>
<td>Q4 (highest)</td>
<td>119,646 (32.8)</td>
<td>8367 (29.9)</td>
</tr>
<tr>
<td>Place of residence (urban), n (%)</td>
<td>162,216 (44.4)</td>
<td>12,324 (44.7)</td>
</tr>
<tr>
<td>Alcohol consumption, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>104,751 (28.7)</td>
<td>8453 (30.2)</td>
</tr>
<tr>
<td>Mild</td>
<td>115,893 (31.7)</td>
<td>10,357 (37)</td>
</tr>
<tr>
<td>Moderate</td>
<td>79,341 (21.7)</td>
<td>5538 (19.8)</td>
</tr>
<tr>
<td>Heavy</td>
<td>65,367 (17.9)</td>
<td>3639 (13)</td>
</tr>
<tr>
<td>Smoking status (CPDb), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light (&lt;10)</td>
<td>34,791 (9.5)</td>
<td>1122 (4)</td>
</tr>
<tr>
<td>Moderate (10-19)</td>
<td>132,370 (36.2)</td>
<td>5605 (20)</td>
</tr>
<tr>
<td>Heavy (≥20)</td>
<td>198,191 (54.3)</td>
<td>21,260 (76)</td>
</tr>
<tr>
<td>Duration of smoking (years)a, n (%)</td>
<td>10,473 (2.9)</td>
<td>690 (2.5)</td>
</tr>
<tr>
<td>Pack-years of smokinga, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No, n (%)</td>
<td>289,791 (79.3)</td>
<td>21,728 (77.6)</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>55,630 (15.2)</td>
<td>4555 (16.2)</td>
</tr>
<tr>
<td>Regular, n (%)</td>
<td>19,931 (5.5)</td>
<td>1724 (6.16)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>24.8 (3.3)</td>
<td>24.7 (3.3)</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg), mean (SD)</td>
<td>126.8 (14.6)</td>
<td>127.0 (14.9)</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg), mean (SD)</td>
<td>78.9 (9.9)</td>
<td>78.8 (9.8)</td>
</tr>
<tr>
<td>Fasting glucose (mg/dL), mean (SD)</td>
<td>137.8 (51.8)</td>
<td>138.4 (50.8)</td>
</tr>
</tbody>
</table>
Change in Smoking Behavior and HF

During a median follow-up of 5.1 (IQR 4.0-6.1) years, there were 13,879 HF cases (7.8 per 1000 person-years). Table 2 shows the associations between change in smoking behavior and the risk of HF, with sustainers as a reference group.

There was no overall difference in HF risk between increasers and sustainers (adjusted hazard ratio [aHR] 1.02, 95% CI 0.97-1.07), but heavy smokers who increased their level of smoking had a 13% higher HF risk (aHR 1.13, 95% CI 1.04-1.24). In the case of reducers, the risk of HF was not reduced but rather increased slightly (aHR 1.14, 95% CI 1.08-1.21, for reducer I; aHR 1.03, 95% CI 0.98-1.09, for reducer II). Conversely, quitters had a lower risk of incident HF after adjusting for confounding factors including type 2 diabetes severity (aHR 0.90, 95% CI 0.86-0.95). When stratified by smoking level at the first screening, moderate-to-heavy smokers who became quitters had a lower risk of HF than sustainers (aHR 0.90, 95% CI 0.83-0.98 for moderate smoker; aHR 0.89, 95% CI 0.83-0.94 for heavy smoker).
Table 2. Associations of smoking behavior change (2009–2011) with heart failure during 5.1 years of follow-up (N=365,352).

<table>
<thead>
<tr>
<th>Smoking behavior change 2009–2011</th>
<th>Participants, n (%)</th>
<th>Events, n (%)</th>
<th>Duration (person-years)</th>
<th>IR</th>
<th>Model 1e HRd</th>
<th>Model 2e HR (95% CI)</th>
<th>Model 3f HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All current smoker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quitter</td>
<td>67,843 (18.6)</td>
<td>2850 (20.5)</td>
<td>333,223.1</td>
<td>8.6</td>
<td>1.17 (1.12-1.23)</td>
<td>0.92 (0.88-0.96)</td>
<td>0.90 (0.86-0.95)</td>
</tr>
<tr>
<td>Reducer I</td>
<td>27,987 (7.7)</td>
<td>1361 (9.8)</td>
<td>136,019.6</td>
<td>10.0</td>
<td>1.38 (1.30-1.46)</td>
<td>1.15 (1.09-1.22)</td>
<td>1.14 (1.08-1.21)</td>
</tr>
<tr>
<td>Reducer II</td>
<td>47,628 (13.0)</td>
<td>1726 (12.4)</td>
<td>234,148.6</td>
<td>7.4</td>
<td>1.01 (0.96-1.07)</td>
<td>1.03 (0.98-1.09)</td>
<td>1.03 (0.98-1.09)</td>
</tr>
<tr>
<td>Sustainer</td>
<td>162,645 (44.5)</td>
<td>5779 (41.6)</td>
<td>796,568.8</td>
<td>7.3</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Increaser</td>
<td>59,249 (16.2)</td>
<td>2163 (15.6)</td>
<td>288,377.6</td>
<td>7.5</td>
<td>1.04 (0.99-1.09)</td>
<td>1.03 (0.98-1.08)</td>
<td>1.02 (0.97-1.07)</td>
</tr>
<tr>
<td>P value</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Smoking status in 2009b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light smoker (n=34,791)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quitter</td>
<td>12,376 (3.4)</td>
<td>585 (4.2)</td>
<td>59,050.7</td>
<td>9.9</td>
<td>1.23 (1.05-1.43)</td>
<td>1.13 (0.97-1.32)</td>
<td>1.12 (0.96-1.30)</td>
</tr>
<tr>
<td>Reducer I</td>
<td>1122 (0.3)</td>
<td>41 (0.3)</td>
<td>5350.9</td>
<td>7.7</td>
<td>0.95 (0.68-1.33)</td>
<td>0.89 (0.64-1.24)</td>
<td>0.89 (0.64-1.24)</td>
</tr>
<tr>
<td>Reducer II</td>
<td>2536 (0.7)</td>
<td>155 (1.1)</td>
<td>11,967.5</td>
<td>13.0</td>
<td>1.61 (1.31-1.97)</td>
<td>1.43 (1.16-1.75)</td>
<td>1.45 (1.18-1.78)</td>
</tr>
<tr>
<td>Sustainer</td>
<td>5905 (1.6)</td>
<td>227 (1.6)</td>
<td>28,215.5</td>
<td>8.0</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Increaser</td>
<td>12,852 (3.5)</td>
<td>534 (3.8)</td>
<td>61,458.2</td>
<td>8.7</td>
<td>1.08 (0.92-1.26)</td>
<td>1.16 (0.99-1.35)</td>
<td>1.16 (0.99-1.35)</td>
</tr>
<tr>
<td>Moderate smoker (n=132,370)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quitter</td>
<td>26,278 (7.2)</td>
<td>1020 (7.3)</td>
<td>128,926.6</td>
<td>7.9</td>
<td>1.18 (1.10-1.28)</td>
<td>0.93 (0.86-1.01)</td>
<td>0.90 (0.83-0.98)</td>
</tr>
<tr>
<td>Reducer I</td>
<td>5605 (1.5)</td>
<td>233 (1.7)</td>
<td>27,180.4</td>
<td>8.6</td>
<td>1.29 (1.13-1.48)</td>
<td>0.97 (0.84-1.11)</td>
<td>0.96 (0.84-1.10)</td>
</tr>
<tr>
<td>Reducer II</td>
<td>15,609 (4.3)</td>
<td>536 (3.9)</td>
<td>76,381.6</td>
<td>7.0</td>
<td>1.06 (0.96-1.16)</td>
<td>0.99 (0.90-1.09)</td>
<td>0.98 (0.89-1.08)</td>
</tr>
<tr>
<td>Sustainer</td>
<td>51,907 (14.2)</td>
<td>1666 (12.0)</td>
<td>252,254.2</td>
<td>6.6</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Increaser</td>
<td>32,971 (9.0)</td>
<td>1082 (7.8)</td>
<td>160,352</td>
<td>6.7</td>
<td>1.02 (0.95-1.10)</td>
<td>1.05 (0.97-1.13)</td>
<td>1.03 (0.95-1.11)</td>
</tr>
<tr>
<td>Heavy smoker (n=198,191)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quitter</td>
<td>29,189 (8.0)</td>
<td>1245 (9.0)</td>
<td>145,245.8</td>
<td>8.6</td>
<td>1.13 (1.06-1.20)</td>
<td>0.90 (0.85-0.96)</td>
<td>0.89 (0.83-0.94)</td>
</tr>
<tr>
<td>Reducer I</td>
<td>21,260 (5.8)</td>
<td>1087 (7.8)</td>
<td>103,488.2</td>
<td>10.5</td>
<td>1.40 (1.31-1.50)</td>
<td>1.19 (1.12-1.28)</td>
<td>1.19 (1.11-1.28)</td>
</tr>
<tr>
<td>Reducer II</td>
<td>29,483 (8.1)</td>
<td>1035 (7.5)</td>
<td>145,799.5</td>
<td>7.1</td>
<td>0.94 (0.88-1.01)</td>
<td>1.03 (0.96-1.10)</td>
<td>1.03 (0.96-1.10)</td>
</tr>
<tr>
<td>Sustainer</td>
<td>104,833 (28.7)</td>
<td>3886 (28.0)</td>
<td>516,099.2</td>
<td>7.5</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Increaser</td>
<td>13,426 (3.7)</td>
<td>547 (3.9)</td>
<td>66,567.3</td>
<td>8.2</td>
<td>1.09 (0.99-1.19)</td>
<td>1.17 (1.07-1.28)</td>
<td>1.13 (1.04-1.24)</td>
</tr>
<tr>
<td>P for interaction</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aQuitters are those who ceased smoking; reducer I are those who reduced the number of cigarettes by 50% or more; reducer II are those who reduced the number of cigarettes by 20% or more; sustainers are those who reduced the number of cigarettes by less than 20% or increased by less than 20%; increasers are those who increased the number of cigarettes by 20% or more.
bIR: incidence rate per 1000 person-years.
cModel 1: unadjusted.
dHR: hazard ratio.
eModel 2: adjusted for age, sex, socioeconomic position (income level and place of residence), alcohol consumption, duration of smoking, physical activity, BMI, and comorbidities (hypertension, dyslipidemia, chronic kidney disease, chronic obstructive pulmonary disease, and stroke).
fModel 3: Model 2 + adjusted for fasting glucose level, duration of type 2 diabetes, and use of insulin.
gN/A: not applicable.
hLight smoker: <10 cigarettes per day; moderate smokers: 10-19 cigarettes per day; heavy smokers: ≥20 cigarettes per day.

Stratified Analysis

Figure 2 presents the associations between smoking behavior change and risk of HF incidence according to type 2 diabetes severity. Smoking cessation remained predictive of a lower incidence of HF, regardless of the duration of type 2 diabetes, number of oral antidiabetic agents, and use of insulin. In stratified analyses according to age and sex, the results were also consistent with the main findings (Table 3).
Figure 2. Association of smoking behavior change (2009-2012) with the incidence of heart failure stratified by (A) duration of type 2 diabetes, (B) number of oral antidiabetic agents, and (C) use of insulin. HRs are adjusted for age, sex, socioeconomic position (income level and place of residence), alcohol consumption, duration of smoking, physical activity, BMI, comorbidities (hypertension, dyslipidemia, chronic kidney disease, chronic obstructive pulmonary disease, and stroke), fasting glucose level, duration of type 2 diabetes, and use of insulin. HR: hazard ratio; quitter: those who ceased smoking; reducer I: those who reduced the number of cigarettes by 50% or more; reducer II: those who reduced the number of cigarettes by 20% or more and by less than 50%; sustainer: those who reduced the number of cigarettes by less than 20% or increased by less than 20%; increaser: those who increased the number of cigarettes by 20% or more.
Table 3. Stratified analysis on the associations of smoking behavior change (2009-2011)* and the risk of heart failure.

<table>
<thead>
<tr>
<th>Smoking behavior in 2009-2011</th>
<th>Participants, n (%)</th>
<th>Events, n (%)</th>
<th>Duration (Person-years)</th>
<th>IRb</th>
<th>Model 1: HRc (95% CI)</th>
<th>Model 2: HRd (95% CI)</th>
<th>Model 3: HRd (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>20-64 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quitter</td>
<td>55,007 (15.1)</td>
<td>1594 (11.5)</td>
<td>272,533.4</td>
<td>5.8</td>
<td>1.04 (0.98-1.10)</td>
<td>0.92 (0.87-0.98)</td>
<td>0.91 (0.85-0.96)</td>
</tr>
<tr>
<td>Reducer I</td>
<td>23,080 (6.3)</td>
<td>783 (5.6)</td>
<td>113,242.4</td>
<td>6.9</td>
<td>1.24 (1.15-1.34)</td>
<td>1.16 (1.08-1.25)</td>
<td>1.15 (1.06-1.24)</td>
</tr>
<tr>
<td>Reducer II</td>
<td>42,174 (11.5)</td>
<td>1156 (8.3)</td>
<td>208,554.5</td>
<td>5.5</td>
<td>0.99 (0.93-1.06)</td>
<td>1.02 (0.95-1.09)</td>
<td>1.02 (0.96-1.09)</td>
</tr>
<tr>
<td>Sustainer</td>
<td>144,252 (39.5)</td>
<td>3957 (28.5)</td>
<td>709,688.7</td>
<td>5.6</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Increaser</td>
<td>52,126 (14.3)</td>
<td>1443 (10.4)</td>
<td>255,321.3</td>
<td>5.7</td>
<td>1.02 (0.96-1.08)</td>
<td>1.04 (0.98-1.10)</td>
<td>1.03 (0.97-1.09)</td>
</tr>
<tr>
<td>≥65 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quitter</td>
<td>12,836 (3.5)</td>
<td>1256 (9.0)</td>
<td>60,669.8</td>
<td>20.7</td>
<td>0.99 (0.92-1.06)</td>
<td>0.92 (0.85-0.99)</td>
<td>0.90 (0.84-0.97)</td>
</tr>
<tr>
<td>Reducer I</td>
<td>4907 (1.3)</td>
<td>578 (4.2)</td>
<td>22,777.2</td>
<td>25.4</td>
<td>1.22 (1.11-1.34)</td>
<td>1.14 (1.04-1.25)</td>
<td>1.14 (1.03-1.25)</td>
</tr>
<tr>
<td>Reducer II</td>
<td>5454 (1.5)</td>
<td>570 (4.1)</td>
<td>25,594.1</td>
<td>22.3</td>
<td>1.06 (0.97-1.17)</td>
<td>1.06 (0.96-1.16)</td>
<td>1.06 (0.96-1.16)</td>
</tr>
<tr>
<td>Sustainer</td>
<td>18,393 (5.0)</td>
<td>1822 (13.1)</td>
<td>86,880.1</td>
<td>21.0</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Increaser</td>
<td>7123 (1.9)</td>
<td>720 (5.2)</td>
<td>33,056.3</td>
<td>21.8</td>
<td>1.05 (0.96-1.14)</td>
<td>1.01 (0.93-1.10)</td>
<td>1.01 (0.92-1.10)</td>
</tr>
<tr>
<td><strong>P value for interaction</strong></td>
<td>N/A²</td>
<td>N/A</td>
<td>N/A</td>
<td>.33</td>
<td>.90</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quitter</td>
<td>61,788 (16.9)</td>
<td>2479 (17.9)</td>
<td>304,628.3</td>
<td>8.1</td>
<td>1.14 (1.08-1.19)</td>
<td>0.91 (0.87-0.95)</td>
<td>0.89 (0.85-0.94)</td>
</tr>
<tr>
<td>Reducer I</td>
<td>26,216 (7.2)</td>
<td>1245 (9.0)</td>
<td>127,646.9</td>
<td>9.8</td>
<td>1.37 (1.29-1.46)</td>
<td>1.16 (1.09-1.24)</td>
<td>1.16 (1.09-1.23)</td>
</tr>
<tr>
<td>Reducer II</td>
<td>45,969 (12.6)</td>
<td>1632 (11.8)</td>
<td>226,394.6</td>
<td>7.2</td>
<td>1.01 (0.96-1.07)</td>
<td>1.03 (0.98-1.09)</td>
<td>1.04 (0.98-1.10)</td>
</tr>
<tr>
<td>Sustainer</td>
<td>157,461 (43.1)</td>
<td>5487 (39.5)</td>
<td>772,035.7</td>
<td>7.1</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Increaser</td>
<td>56,251 (15.4)</td>
<td>1997 (14.4)</td>
<td>274,163.9</td>
<td>7.3</td>
<td>1.03 (0.98-1.08)</td>
<td>1.04 (0.99-1.09)</td>
<td>1.03 (0.98-1.08)</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quitter</td>
<td>6055 (1.7)</td>
<td>371 (2.7)</td>
<td>28,594.9</td>
<td>13.0</td>
<td>1.09 (0.93-1.27)</td>
<td>0.98 (0.84-1.14)</td>
<td>0.97 (0.83-1.13)</td>
</tr>
<tr>
<td>Reducer I</td>
<td>1771 (0.5)</td>
<td>116 (0.8)</td>
<td>8372.7</td>
<td>13.9</td>
<td>1.16 (0.94-1.44)</td>
<td>1.03 (0.83-1.27)</td>
<td>1.04 (0.84-1.29)</td>
</tr>
<tr>
<td>Reducer II</td>
<td>1659 (0.5)</td>
<td>94 (0.7)</td>
<td>7754.0</td>
<td>12.1</td>
<td>1.02 (0.81-1.29)</td>
<td>0.97 (0.77-1.22)</td>
<td>0.97 (0.77-1.22)</td>
</tr>
<tr>
<td>Sustainer</td>
<td>5184 (1.4)</td>
<td>292 (2.1)</td>
<td>24,533.2</td>
<td>11.9</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Increaser</td>
<td>2998 (0.8)</td>
<td>166 (1.2)</td>
<td>14,213.7</td>
<td>11.7</td>
<td>0.98 (0.81-1.19)</td>
<td>0.94 (0.78-1.14)</td>
<td>0.95 (0.79-1.15)</td>
</tr>
<tr>
<td><strong>P value for interaction</strong></td>
<td>N/A²</td>
<td>N/A</td>
<td>N/A</td>
<td>.67</td>
<td>.26</td>
<td>.30</td>
<td></td>
</tr>
</tbody>
</table>

*Quitters are those who quit smoking; reducer I are those who reduced the number of cigarettes by 50% or more; reducer II are those who the number of cigarettes by 20% or more and less than 50%; sustainer are those who reduced the number of cigarettes by less than 20% or increased by less than 20%; increaser are those who increased the number of cigarettes by 20% or more.

IR: incidence rate per 1000 person-years.

Model 1: unadjusted.

HR: hazard ratio.

Model 2: Adjusted for age, sex, socioeconomic position (income level and place of residence), alcohol consumption, duration of smoking, physical activity, BMI, and comorbidities (hypertension, dyslipidemia, chronic kidney disease, chronic obstructive pulmonary disease, and stroke).

Model 3: Model 2 + adjusted for fasting glucose level, duration of type 2 diabetes, and use of insulin.

Sex: not applicable.

Discussion

Principal Findings

In this large cohort study with repetitive measurements of smoking behavior, we found that increasers, particularly those who were initially heavy smokers, were associated with higher risks of HF than sustainers. In the case of reducers, the incidence of HF was not decreased but slightly increased. Smoking cessation was associated with a decreased risk of incident HF among moderate-to-heavy smokers with type 2 diabetes. Our stratified analyses also demonstrated that smoking cessation...
was consistently associated with a lower HF risk among type 2 diabetes with various characteristics.

The association between smoking and atherosclerotic coronary artery disease, a major cause of HF [7], is well established. The acute prothrombotic, proadrenergic, and proinflammatory properties of smoking are believed to underlie smoking-associated atherosclerosis and cardiovascular disease [11]. In addition, smoking directly leads to left ventricular hypertrophy. In vitro, nicotine stimulates the proliferation of endothelial cells and vascular smooth muscle cells [25], and chronic inhalation of carbon monoxide induces cardiac hypertrophy [25]. Consistently, we showed that heavy smokers with type 2 diabetes who increased their amount of smoking during 2 consecutive screenings had a higher risk of HF than those who sustained their smoking amount. In type 2 diabetes, hyperglycemia, insulin resistance, and hyperinsulinemia trigger a cascade of deleterious effects that contribute to the development of HF via ischemic cardiomyopathy and diabetic cardiomyopathy [26,27]. There have been many studies showing that the onset and progression of macrovascular and microvascular complications of type 2 diabetes are highly associated with smoking [28]. Therefore, it is possible that smoking and type 2 diabetes interact and further increase the risk of HF.

Harm reduction strategy is aimed at reducing the adverse health effects of smoking by reducing the number of cigarettes smoked each day [29]. However, no evidence that cardiovascular risks are reduced by smoking reduction have been provided [29-31]. Consistently, we found that smoking reduction among patients with type 2 diabetes had no benefit for reducing HF risk. Type 2 diabetes itself is a high-risk condition of HF and also interrelates with smoking in relation to cardiovascular events [14]. Thus, the threshold at which smoking affects the development of HF appears to be low, and smoking reduction may not be sufficient to reduce the risk of HF. Moreover, substantial smoking reduction of ≥50% worsened rather than leaving unchanged HF risk, particularly in heavy smokers. The disease burden attributable to smoking is undoubtedly highest in heavy smokers. For heavy smokers, the effect of smoking may persist after reducing smoking, leading to further deterioration of cardiovascular health. Another possible explanation for this phenomenon is that smoking is primarily a nicotine-seeking behavior, and smokers who reduce tend to compensate by smoking more of each cigarette and by taking more and deeper puffs [29]. This results in a much smaller proportional reduction in nicotine intake (and in the intake of associated tar and other toxins) than the reduction in number of cigarettes suggests [29]. The results of this study support health warnings that there is no safe level of smoking for preventing cardiovascular disease including HF. Smokers with type 2 diabetes need to have a goal of complete smoking cessation, even from heavy smoking, substantially reduced the risk of HF.

Smoking cessation consistently decreased the risk of HF even after various stratifications. Interestingly, while the risk of HF among patients with type 2 diabetes usually increases with age and severity or duration of type 2 diabetes, the preventive effect of smoking cessation on HF was also observed among young participants, those with new-onset type 2 diabetes, and those taking fewer oral antidiabetic agents without insulin use. These findings suggest that increased HF risk associated with smoking among patients with type 2 diabetes can be attenuated by smoking cessation.

Current guidelines, such as those of the American Diabetes Association, emphasize that smoking cessation is a major target for the prevention of type 2 diabetes–associated cardiovascular diseases [38]. However, many patients continue to smoke even after a diagnosis of type 2 diabetes [39]. The success rate of smoking cessation among type 2 diabetes is not high, with some studies reporting cessation to be under 20% [40]. In addition to cardiovascular disease, we showed that smoking cessation had a consistent effect on reducing the risk of HF, regardless of type 2 diabetes severity. This study which demonstrates the additional hazards of HF associated with continued smoking and the benefits of smoking cessation among patients with type 2 diabetes provide direct evidence for the need for clinical practitioners to intervene to change smoking behavior.

**Limitations**

There are several limitations of our study. First, we defined type 2 diabetes from health claim data not from clinical records and, therefore, could be subject to the risk of under- or overascertainment. However, we used the diagnosis code and medication records together, which have been shown to have high accuracy [41]. Second, since smoking behaviors were based on self-reported questionnaire without using biochemical verification, misclassification from recall or social desirability bias could exist. However, self-reported smoking behavior has been shown to reverse smoking-associated endothelial dysfunction, which reduces the cardiovascular disease risk of smoking soon after cessation [6]. Smoking-associated alterations of the myocardium, such as left ventricular remodeling and dysfunction, were also reversible after smoking cessation [9]. We confirmed that smoking cessation, even from heavy smoking, substantially reduced the risk of HF.

It is well known that smoking and type 2 diabetes interrelate. Type 2 diabetes is one of the smoking-related diseases, and smoking cessation reduces the risk of type 2 diabetes [34-36]. In type 2 diabetes, glycemic control tends to worsen after quitting, which can last up to 3 years [37]. Consistently, in this study, analyzing 2-year interval changes, quitters were prescribed more glucose-lowering therapies but did not show a greater decrease in fasting glucose than other groups, which possibly reflects worsened short-term glycermic control in quitters (Table S1 and S2 in Multimedia Appendix 1). Therefore, the reduced risk of HF from smoking cessation is mainly from the effect of smoking cessation per se, and the benefits of smoking cessation may outweigh the lighter, temporarily increased glycermic control from smoking cessation with respect to HF among patients with type 2 diabetes. If we could track quitters over a longer period, long-term quitters might achieve improved glycermic control.
been reported to be relatively accurate with 87.5% sensitivity and 89.2% specificity [42]. Third, because we used administrative data, we did not have sufficient clinical information on HF including phenotypes of incident HF, etiology of HF, or plasma brain natriuretic peptide levels.

**Conclusions**

In this large population-based cohort study, we demonstrated that increasing smoking amount was associated with a higher risk for HF compared to maintaining smoking amount, while smoking cessation was associated with a lower risk of HF among patients with type 2 diabetes. There was no benefit from the reduction in smoking amount. These findings suggest that smoking cessation should be reinforced to prevent HF in populations at high risk and with type 2 diabetes.

**Acknowledgments**

This study was performed using the Korean National Health Insurance Service database, and the results do not necessarily represent the opinion of the National Health Insurance Corporation.

**Data Availability**

The data sets generated during and analyzed during this study are not publicly available because restrictions apply to the availability of the data, which were used with permission for the present study. However, they may be available from the corresponding author on reasonable request and with permission from the Korean National Health Insurance Service.

**Conflicts of Interest**

None declared.

**Authors' Contributions**

Authors DWS (dwshin.md@gmail.com) and KH (hkd917@naver.com) are co-corresponding authors for this article.

**Multimedia Appendix 1**

Changes in lifestyle and type 2 diabetes severity according to smoking behavior change (2009 through 2011). Changes in weight and laboratory findings according to smoking behavior change (2009 through 2011).

[DOCX File, 29 KB - publichealth_v10i1e46450_app1.docx ]

**References**


Abbreviations

aHR: adjusted hazard ratio
HF: heart failure
ICD-10: International Classification of Diseases, 10th revision
NHIS: National Health Insurance Service
The Impact of Prostate-Specific Antigen Screening on Prostate Cancer Incidence and Mortality in China: 13-Year Prospective Population-Based Cohort Study

Xiaohao Ruan1*, MD; Ning Zhang1*, MD; Dawei Wang1*, MD; Jingyi Huang1, MD; Jinlun Huang1, MM; Da Huang1, MD; Tsun Tsun Stacia Chun2, BSc; Brian Sze Ho Ho3, MBBS; Ada Tsui-Lin Ng3, MBBS; James Hok-Leung Tsu3, MBBS; Yongle Zhan2, PhD; Rong Na2, MPH, MD, PhD

1Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
2Department of Surgery, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
3Division of Urology, Department of Surgery, Queen Mary Hospital, Hong Kong, China (Hong Kong)

*these authors contributed equally

Corresponding Author:
Yongle Zhan, PhD
Department of Surgery
LKS Faculty of Medicine
The University of Hong Kong
102 Pokfulam Road
Hong Kong
China (Hong Kong)
Phone: 852 22554852
Email: ylzhan@connect.hku.hk

Abstract

Background: The status of prostate-specific antigen (PSA) screening is unclear in China. Evidence regarding the optimal frequency and interval of serial screening for prostate cancer (PCa) is disputable.

Objective: This study aimed to depict the status of PSA screening and to explore the optimal screening frequency for PCa in China.

Methods: A 13-year prospective cohort study was conducted using the Chinese Electronic Health Records Research in Yinzhou study’s data set. A total of 420,941 male participants aged ≥45 years were included between January 2009 and June 2022. Diagnosis of PCa, cancer-specific death, and all-cause death were obtained from the electronic health records and vital statistic system. Hazard ratios (HRs) with 95% CIs were estimated using Cox regression analysis.

Results: The cumulative rate of ever PSA testing was 17.9% with an average annual percent change (AAPC) of 8.7% (95% CI 3.6%-14.0%) in the past decade in China. People with an older age, a higher BMI, higher waist circumference, tobacco smoking and alcohol drinking behaviors, higher level of physical activity, medication use, and comorbidities were more likely to receive PSA screening, whereas those with a lower education level and a widowed status were less likely to receive the test. People receiving serial screening ≥3 times were at a 67% higher risk of PCa detection (HR 1.67; 95% CI 1.48-1.88) but a 64% lower risk of PCa-specific mortality (HR 0.36; 95% CI 0.18-0.70) and a 28% lower risk of overall mortality (HR 0.72; 95% CI 0.67-0.77). People following a serial screening strategy at least once every 4 years were at a 25% higher risk of PCa detection (HR 1.25; 95% CI 1.13-1.36) but 70% (HR 0.30; 95% CI 0.16-0.57) and 23% (HR 0.77; 95% CI 0.73-0.82) lower risks of PCa-specific and all-cause mortality, respectively.

Conclusions: This study reveals a low coverage of PSA screening in China and provides the first evidence of its benefits in the general Chinese population. The findings of this study indicate that receiving serial screening at least once every 4 years is beneficial for overall and PCa-specific survival. Further studies based on a nationwide population and with long-term follow-up are warranted to identify the optimal screening interval in China.

(JMIR Public Health Surveill 2024;10:e47161) doi:10.2196/47161
KEYWORDS
prostate-specific antigen; PSA; prostate cancer; prostate screening; screening interval; incidence; mortality; cohort study; electronic health record; China

Introduction

Globally, prostate cancer (PCa) is the second-most commonly diagnosed male cancer, with an estimated 1.41 million new cases and 0.37 million deaths reported in 2020 [1]. In China, among all cancers, PCa ranks sixth in incidence and seventh in mortality [2], which have been increasing rapidly over the past decades [3]. The aging population and the widespread use of prostate-specific antigen (PSA) testing could partially explain the significant increase in disease burden.

The uptake of PSA testing has been well characterized in Western countries. In Switzerland, for example, the coverage rate of PSA screening has increased to 70% [4]. In the United Kingdom, 55.3% of men aged 40-75 years have undergone at least 1 PSA screening [5]. In the United States, 32.1% of men aged 50 years or older have had PSA screening for routine reasons [6]. On the contrary, anecdotal evidence suggests a low PSA screening rate in most Asian countries [7]. A wide range of factors have been associated with screening behavior in the general population. For instance, people with an older age, a family history of PCa, and a high BMI were more likely to receive PSA testing, whereas socioeconomic deprivation, heavy smoking, and comorbid diabetes were associated with a lower likelihood of PSA testing [5].

PSA screening policy and guidelines to date have been well established in the West, including the recommendation statement of screening for prostate cancer by the US Preventive Services Task Force [8], the National Comprehensive Cancer Network’s guidelines for Prostate Cancer Early Detection [9], European Association of Urology–European Association of Nuclear Medicine–European Society for Radiotherapy & Oncology–European Society of Urological Radiology–International Society of Geriatric Oncology Guidelines on Prostate Cancer [10]. However, PSA screening also leads to overdiagnosis and overtreatment of insignificant PCa, which has been increasingly criticized [11]. Several large randomized controlled trials (RCTs) have suggested that such screening contributes to little benefit for overall and cancer-specific survival, especially when taking its efficiency or cost-effectiveness into account [12]. In China, the effect of PSA screening is poorly understood, and few epidemiological studies have depicted the rate and trend of PSA testing from a population-based perspective.

On the other hand, a consensus on screening strategy has yet to be determined. Some studies suggest that one-time PSA screening is not beneficial [13], while others suggest that annual screening is not cost-effective [14]. The optimal frequency and interval for serial screening remains to be concurred. Kobayashi et al [15] reported an appropriate interval of ≥3 years for PSA rescreening, while Shao et al [16] concluded that more frequent PSA testing could aggravate the risk of overdiagnosis. In well-known clinical trials, no consensus on screening intensity has been established. For example, one-time screening was applied in the Cluster Randomised Trial of PSA Testing for Prostate Cancer; annual screening was undertaken in the Prostate, Lung, Colorectal and Ovarian Cancer Screening trial; and screening once every 2-4 years was performed in the European Randomised Study of Screening for Prostate Cancer [12].

Given a paucity of evidence regarding the depiction of PSA testing, and the benefits of different screening strategies in China, we performed a population-based prospective cohort study using electronic health records (EHRs) data. The primary aims of this study are to evaluate the effects of PSA screening and to explore the optimal screening frequency and interval in the Chinese population.

Methods

Study Setting and Population

Data for this study were obtained from the Chinese Electronic Health Records Research in Yinzhou (CHERRY) study, the first population-based cohort study linking big data of integrated individual-level EHRs. Detailed information regarding the study design, inclusion criteria, data collection, and procedure of the study is published previously [17]. Briefly, the CHERRY study to date has comprised a total of more than 1 million permanent residents living in Yinzhou, an economically advanced region of southeastern China. Participants were included in the CHERRY study if they (1) registered in the system, (2) were older than 18 years on January 1, 2009, (3) were living in Yinzhou for more than 6 months, and (4) provided consent to participate in the study. The data set of the CHERRY study was integrated from the population census, primary care, electronic medical records, health check, surveillance, and vital statistic, where all health-related activities (eg, inpatient and outpatient visits) within the region were recorded.

As 45 years is the initiation age for PCa screening recommended by the National Cancer Center of China [18], we accordingly included male participants aged ≥45 years in this study. We further excluded those (1) aged <45 years at recruitment (n=189,333), (2) who consented to participate but died before enrollment in the cohort (n=574), and (3) with illogical recorded information (eg, recorded death occurred before the uptake of PSA testing; n=19); 420,941 eligible participants were retained in the final analysis. A 3×3 matrix data quality assessment framework proposed by Weiskopf et al [19] was further applied to comprehensively evaluate the data quality of the data set. A detailed data quality assessment report is presented in Multimedia Appendix 1. The flowchart of this study is shown in Figure 1.
Serum PSA was tested using an immunoenzymatic assay with the World Health Organization’s international standard or the Hybritech Standard based on different protocols among hospitals [20]. The value and date of each PSA test were extracted from the CHERRY study’s data set. We then developed a series of variables related to PSA screening. Ever screening was defined as ever having any PSA test or not (ever vs never). The subsequent variables were related to screening strategy and were thus restricted to those ever-having a PSA test. Number of PSA screenings was defined as a total count of screenings during the follow-up years; this variable was further categorized into 3 groups (1, 2, and ≥3) for comparison. Four variables were proxies of screening frequency: receiving 1 PSA test at least annually (yes or no), receiving 1 PSA test at least biennially (yes or no), receiving 1 PSA test at least triennially (yes or no), and receiving 1 PSA test at least quadrennially (yes or no). Screening interval was defined as the average interval of having serial tests (irregular, 1 year, 2 years, 3 years, and 4 years). Detailed definitions of the abovementioned variables are provided in Table S1 in Multimedia Appendix 1.

Outcomes of Interest
Participants were followed up for morbidity and mortality using records linked with the regional system of disease surveillance, chronic disease management, and EHRs based on diagnostic codes from the ICD-10 (International Statistical Classification of Diseases, Tenth Revision). The primary outcomes of this study were prostate cancer (C61) and PCa-specific death. The secondary outcome was all-cause death.

Covariates
Sociodemographics (age, education level, and marital status), anthropometric factors (BMI and waist circumference [WC]), lifestyle (smoking, drinking, and physical activity [PA]), use of medications (5-α reductase inhibitors), and comorbidity were included and used for adjustment in the analyses. Detailed definitions of the covariates are provided in Table S1 in Multimedia Appendix 1.

Statistical Analysis
Descriptive statistics were summarized using means (SD) or median (IQR) values for continuous variables (normally or nonnormally distributed) and number (percentage) values for categorical variables. Follow-up person-years (PYs) were calculated from the date of cohort inception for the nonscreening group and the date of the initial PSA test for the screening group to either the date of PCa diagnosis, death, loss to follow-up, or June 15, 2022, whichever occurred first.

Temporal trends in the annual rate of first PSA screening were estimated by joinpoint regression models [21]. The direction and magnitude of the trends were assessed by the annual percentage change (APC) and average annual percent change (AAPC) with corresponding 95% CIs. Missing values were presumed to be missing at random and were filled by multiple imputations based on chained equations. Cox proportional hazards models were used to estimate the hazard ratios (HRs) and 95% CIs for PSA screening on PCa incidence, cause-specific mortality, and overall mortality. The proportional hazards assumption was tested on the basis of Schoenfeld residuals. Cox models were stratified by age-at-risk of PCa or death accordingly (5-year intervals; model 1), and HRs were adjusted for sociodemographics, anthropometric factors, lifestyle, medications, and comorbidities (model 2), and were additionally adjusted for baseline PSA values and age at the first PSA test (model 3) according to the proposed directed acyclic graph (Figure S1 in Multimedia Appendix 1). In addition, the E-value was calculated to assess the robustness of the main results against unmeasured confounders [22].
All $P$ values were 2-sided and were considered significant when less than .05. The abovementioned analyses were performed using Stata (version 17.0; StataCorp), Joinpoint software (version 4.8.0.1; National Cancer Institute), and R statistical software (version 4.1.2; Foundation for Statistical Computing).

**Ethical Considerations**

The study was approved by the institutional review board of the University of Hong Kong (UW 22-766). All procedures were performed in accordance with the tenets of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent has been obtained from all participants before the study.

**Results**

**Baseline Characteristics**

Among the 420,941 participants included in the study, the overall mean age was 50.4 (SD 11.9) years. After a total of 3,177,289 person-years of follow-up, 2160 men were diagnosed with PCa (incidence rate 0.68 per 1000 PYs), 92 men died from PCa (cause-specific mortality rate 0.03 per 1000 PYs), and 20,781 men died from all causes (overall mortality rate 6.78 per 1000 PYs). The mean BMI was 22.9 (SD 3.1) kg/m$^2$, and 1.8% (n=3900) of them had a high WC. In terms of socioeconomic factors, 3.5% (n=14,297) of men had a bachelor’s degree or above, and 92.0% (n=221,845) of them were married. Long-term medication use (n=5319, 1.9%) and comorbidities (n=33,220, 12.1%) were observed in a small proportion of men. Smoking and drinking alcohol was prevalent among 21.3% (n=48,731) and 23.1% (n=52,916), respectively. Detailed characteristics of the study participants are shown in Table 1.
Table 1. Characteristics of the study participants (N=420,941).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>50.4 (11.9)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Bachelors and above</td>
<td>14,297 (3.5)</td>
</tr>
<tr>
<td>Below bachelors</td>
<td>394,644 (96.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>12,000 (N/A&lt;sup&gt;a&lt;/sup&gt;)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>3118 (1.3)</td>
</tr>
<tr>
<td>Married</td>
<td>221,845 (92.0)</td>
</tr>
<tr>
<td>Widowed</td>
<td>14,440 (6.0)</td>
</tr>
<tr>
<td>Divorced</td>
<td>1756 (0.7)</td>
</tr>
<tr>
<td>Missing</td>
<td>179,782 (N/A)</td>
</tr>
<tr>
<td><strong>Medication use, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>270,225 (98.1)</td>
</tr>
<tr>
<td>Yes</td>
<td>5319 (1.9)</td>
</tr>
<tr>
<td>Missing</td>
<td>145,397 (N/A)</td>
</tr>
<tr>
<td><strong>Comorbidity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>242,324 (87.9)</td>
</tr>
<tr>
<td>Yes</td>
<td>33,220 (12.1)</td>
</tr>
<tr>
<td>Missing</td>
<td>145,397 (N/A)</td>
</tr>
<tr>
<td><strong>BMI (kg/m²), mean (SD)</strong></td>
<td>22.9 (3.1)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

**High waist circumference, n (%)**

| No                               | 216,335 (98.2)          |
| Yes                              | 3900 (1.8)              |
| Missing                          | 200,706 (N/A)           |

**Physical activity, n (%)**

| Never                           | 9056 (7.0)              |
| Occasional (<1 per week)        | 39,388 (30.3)           |
| Frequent (≥1 per week)          | 81,561 (62.7)           |
| Missing                         | 290,936 (N/A)           |

**Tobacco smoking, n (%)**

| Never                           | 180,585 (78.7)          |
| Ever                            | 48,731 (21.3)           |
| Missing                         | 191,625 (N/A)           |

**Alcohol drinking, n (%)**

| Never                           | 176,047 (76.9)          |
| Ever                            | 52,916 (23.1)           |
| Missing                         | 191,978 (N/A)           |

<sup>a</sup>N/A: not applicable.

<sup>b</sup>BMI values missing for 210,431 individuals.
PSA Screening

Overall, 17.9% (n=75,362) of participants ever attended PSA testing. Among them, 51.1% (n=38,527), 22.1% (n=16,686), and 26.7% (n=20,149) underwent testing 1, 2, and ≥3 times; two-thirds of them (63.2%, n=47,654) attended PSA screening without regularity; about one-sixth of them (16.8%, n=12,686) had a baseline PSA value of 3 ng/mL or above and had their first test at the age of 75 years or older (17.9%, n=13,514). A detailed description of PSA screening by age stratification is provided in Table 2.

During 2010-2021, the annual rates of first PSA testing ranged from 0.85% to 3.45%. Joinpoint regression analyses suggested a significant increasing tendency of uptake of PSA testing during the past 12 years (AAPC 8.70%; 95% CI 3.61-14.04; P=.001). The increase in trend in the later 6 years (APC 15.65%; 95% CI 7.97-23.88; P=.002) was 5-fold faster than that in the prior 6 years (APC 3.23%; 95% CI –5.58 to 12.86; P=.43; Figure S2 in Multimedia Appendix 1).

Figure 2 shows several common factors that could influence PSA screening behavior. Age (HR 1.02; 95% CI 1.02-1.02; P<.001), BMI (HR 1.02; 95% CI 1.01-1.02; P<.001), high WC (HR 1.31; 95% CI 1.23-1.40; P<.001), tobacco smoking (HR 1.91; 95% CI 1.88-1.95; P<.001), alcohol drinking (HR 1.78; 95% CI 1.74-1.81; P<.001), physical activity (HR 1.10, 95% CI 1.06-1.15; P<.001 for occasional PA and HR 1.05; 95% CI 1.01-1.09; P=0.01 for frequent PA), medication use (HR 3.16, 95% CI 3.02-3.30; P<.001), and comorbidity (HR 1.09; 95% CI 1.07-1.10; P<.001) were associated with higher probability of receiving a PSA test, while those with lower education levels (HR 0.90; 95% CI 0.85-0.95; P<.001) and with a widowed status (HR 0.68; 95% CI 0.61-0.76; P<.001) may be less likely to receive a PSA test.

Table 2. Age-specific characteristics of participants (N=420,941) having undergone prostate-specific antigen (PSA) screening.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total, n (%)</th>
<th>Enrollment age (years), n (%)</th>
<th>45-49</th>
<th>50-54</th>
<th>55-59</th>
<th>60-64</th>
<th>65-69</th>
<th>70-74</th>
<th>≥75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>420,941 (100)</td>
<td>104,178 (24.7)</td>
<td>62,689 (17.3)</td>
<td>56,275 (13.4)</td>
<td>36,476 (8.7)</td>
<td>24,898 (5.9)</td>
<td>38,068 (9.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSA screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>345,579 (82.1)</td>
<td>90,576 (86.9)</td>
<td>74,520 (84.3)</td>
<td>59,402 (81.7)</td>
<td>44,296 (78.7)</td>
<td>27,952 (76.6)</td>
<td>18,681 (75.0)</td>
<td>30,152 (79.2)</td>
<td></td>
</tr>
<tr>
<td>Ever</td>
<td>75,362 (17.9)</td>
<td>13,602 (13.1)</td>
<td>13,837 (15.7)</td>
<td>13,287 (18.3)</td>
<td>11,979 (21.3)</td>
<td>8524 (23.4)</td>
<td>6217 (25.0)</td>
<td>7916 (20.8)</td>
<td></td>
</tr>
<tr>
<td>Number of PSA screenings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>38,527 (51.2)</td>
<td>7465 (54.9)</td>
<td>7541 (54.5)</td>
<td>6973 (52.5)</td>
<td>5947 (49.6)</td>
<td>4093 (48.0)</td>
<td>2861 (46.0)</td>
<td>3647 (46.1)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16,686 (22.1)</td>
<td>2950 (21.7)</td>
<td>3054 (22.1)</td>
<td>2892 (21.8)</td>
<td>2652 (22.1)</td>
<td>1917 (22.5)</td>
<td>1397 (22.5)</td>
<td>1824 (23.0)</td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>20,149 (26.7)</td>
<td>3187 (23.4)</td>
<td>3242 (23.4)</td>
<td>3422 (25.8)</td>
<td>3380 (28.2)</td>
<td>2514 (29.5)</td>
<td>1959 (31.5)</td>
<td>2445 (30.9)</td>
<td></td>
</tr>
<tr>
<td>Screening interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irregularb</td>
<td>47,654 (63.2)</td>
<td>8658 (63.7)</td>
<td>8996 (64.3)</td>
<td>8460 (63.7)</td>
<td>7474 (62.4)</td>
<td>5346 (62.7)</td>
<td>3846 (61.9)</td>
<td>4974 (62.8)</td>
<td></td>
</tr>
<tr>
<td>Quadrennial</td>
<td>3650 (4.8)</td>
<td>555 (4.1)</td>
<td>565 (4.1)</td>
<td>656 (4.9)</td>
<td>605 (5.1)</td>
<td>433 (5.1)</td>
<td>355 (5.7)</td>
<td>481 (6.1)</td>
<td></td>
</tr>
<tr>
<td>Triennial</td>
<td>6967 (9.2)</td>
<td>1093 (8.0)</td>
<td>1222 (8.8)</td>
<td>1155 (8.7)</td>
<td>1201 (10.0)</td>
<td>788 (9.2)</td>
<td>646 (10.4)</td>
<td>862 (10.9)</td>
<td></td>
</tr>
<tr>
<td>Biennial</td>
<td>12,409 (16.5)</td>
<td>2630 (19.3)</td>
<td>2295 (16.6)</td>
<td>2096 (15.8)</td>
<td>1876 (15.7)</td>
<td>1351 (15.8)</td>
<td>962 (15.5)</td>
<td>1199 (15.1)</td>
<td></td>
</tr>
<tr>
<td>Annual</td>
<td>4682 (6.2)</td>
<td>666 (4.9)</td>
<td>859 (6.2)</td>
<td>920 (6.9)</td>
<td>823 (6.9)</td>
<td>606(7.1)</td>
<td>408 (6.6)</td>
<td>400 (5.1)</td>
<td></td>
</tr>
<tr>
<td>Baseline PSA valuea (ng/mL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;3</td>
<td>62,676 (83.2)</td>
<td>12,792 (94.0)</td>
<td>12,562 (90.8)</td>
<td>11,430 (86.0)</td>
<td>9853 (82.3)</td>
<td>6459 (75.8)</td>
<td>4491 (72.2)</td>
<td>5089 (64.3)</td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>12,686 (16.8)</td>
<td>810 (6.0)</td>
<td>1275 (9.2)</td>
<td>1857 (14.0)</td>
<td>2126 (17.7)</td>
<td>2065 (24.2)</td>
<td>1726 (27.8)</td>
<td>2827 (35.7)</td>
<td></td>
</tr>
<tr>
<td>Age at first PSA testa, year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;75</td>
<td>61,848 (82.1)</td>
<td>13,602 (100.0)</td>
<td>13,837 (100.0)</td>
<td>13,287 (100.0)</td>
<td>11,175 (93.3)</td>
<td>6281 (73.7)</td>
<td>2976 (47.9)</td>
<td>690 (8.7)</td>
<td></td>
</tr>
<tr>
<td>≥75</td>
<td>13,514 (17.9)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>804 (6.7)</td>
<td>2243 (26.3)</td>
<td>3241 (52.1)</td>
<td>7226 (91.3)</td>
</tr>
</tbody>
</table>

Statistics were restricted to those who ever had a PSA screening.
Irregular screening referred to attending screening without regularity over time.
Effects of PSA Screening in the General Population

Compared to nonattendees, screening attendees were at 70-, 16-, and 6-fold higher risk of PCa incidence, PCa death, and overall death, respectively, in the crude model (Tables S2-S4 in Multimedia Appendix 1). After adjustment for age, education, marital status, BMI, high WC, smoking, drinking, PA, medication use, and comorbidity, we found that attendees were still at 35-, 11-, and 1.2-fold higher risk of PCa incidence (HR 35.87; 95% CI 31.82-40.85; \( P < .001 \)), PCa death (HR 11.94; 95% CI 7.69-18.54; \( P < .001 \)), and overall death (HR 1.27; 95% CI 1.20-1.35; \( P < .001 \)), respectively.

PSA Screening Strategy and PCa Incidence

All statistics on screening strategy were restricted to individuals ever having had a PSA test (n=75,362). In the full adjustment model, people receiving a PSA test more than 3 times were at a 67% higher risk of PCa detection than those having had a PSA test once (HR 1.67; 95% CI 1.48-1.88; \( P < .001 \)). Frequent attendees were at a higher risk of PCa detection (HR 3.60; 95% CI 3.15-4.10; \( P < .001 \)) for having a PSA test at least annually; HR 1.75; 95% CI 1.57-1.93; \( P < .001 \) for having a PSA test at least biennially; HR 1.40; 95% CI 1.28-1.55; \( P < .001 \) for having a PSA test at least triennially; and HR 1.25; 95% CI 1.13-1.36; \( P < .001 \) for having a PSA test at least quadrennially. People receiving serial screening with a 1-year interval were at a higher risk of PCa detection (HR 3.53; 95% CI 3.03-4.06; \( P < .001 \)), but those with a 2-year interval were at no significant risk of PCa detection (HR 1.07; 95% CI 0.94-1.22; \( P = .29 \)), while those with 3- and 4-year intervals were at a lower risk of PCa detection (HR 0.80; 95% CI 0.69-0.94; \( P = .006 \) for a 3-year interval; HR 0.61; 95% CI 0.49-0.76; \( P < .001 \) for a 4-year interval; Figure 3; Table S5 in Multimedia Appendix 1).

**Figure 3.** Adjusted Cox regression model for prostate-specific antigen screening strategy on prostate cancer (PCa) incidence, PCa-specific mortality, and overall mortality. HR: hazard ratio; PSA: prostate-specific antigen.
PSA Screening Strategy and PCa-Specific Mortality

People receiving PSA screening more than 3 times were at a lower risk of PCa-specific mortality than those receiving PSA testing once (HR 0.36; 95% CI 0.18-0.70; P=0.002). We further found a 26% decreased risk with every additional PSA screenings (HR 0.74; 95% CI 0.61-0.89; P=0.001). Frequent attendees were at a 24%-70% lower risk of PCa death (HR 0.76; 95% CI 0.18-3.25; P=0.72 for at least an annual PSA screening; HR 0.68; 95% CI 0.33-1.38; P=0.29 for at least a biennial screening; HR 0.41; 95% CI 0.22-0.79; P=0.007 for at least a triennial screening; HR 0.30; 95% CI 0.16-0.57; P<0.001 for at least a quadrennial screening). People receiving serial screening with a 3-year interval were at a significant lower risk of PCa death than with other intervals (HR 0.15; 95% CI 0.04-0.62; P=0.009). Such significance was not identified in other interval groups due to low statistical power (Figure 3 and Table S6 in Multimedia Appendix 1).

PSA Screening Strategy and Overall Mortality

People receiving PSA screening more than 3 times were at a lower risk of overall death than those receiving PSA screening once (HR 0.72; 95% CI 0.67-0.77; P<0.001). Furthermore, we found a 3% lower risk with every additional PSA screenings (HR 0.97; 95% CI 0.95-0.99; P=0.001). People receiving at least annual (HR 2.68; 95% CI 2.41-2.97; P<0.001) and biennial PSA screening (HR 1.39; 95% CI 1.30-1.49; P<0.001) were at a higher risk of overall death, while those receiving at least triennial (HR 0.97; 95% CI 0.90-1.03; P=0.28) and quadrennial PSA screening (HR 0.77; 95% CI 0.73-0.82; P<0.001) were at lower risks of overall death. Compared to people with irregular screening intervals, those with a 1-year interval were at a 136% higher risk of overall death (HR 2.36; 95% CI 2.11-2.64; P<0.001), while those with extended intervals were at a 9%-74% lower risk of overall death (HR 0.91; 95% CI 0.84-0.99; P=0.03 for those with a 2-year interval; HR 0.43; 95% CI 0.38-0.48; P<0.001 for those with a 3-year interval; HR 0.26; 95% CI 0.21-0.32; P<0.001 for those with a 4-year interval; Figure 3 and Table S7 in Multimedia Appendix 1).

Discussion

This prospective cohort study including 420,941 male participants suggests that the cumulative rate of ever PSA testing was 17.9% in the past decade, with an upward secular trend in China. Evidence provided by this study supports the benefit of regular screening (in particular, at least once every 4 years) for overall and cancer-specific survival among individuals attending repeated PSA screening.

This study revealed a low coverage of PSA screening in the Chinese population, which was consistent with the results of several prior studies. So et al [23] reported a 9.5% uptake rate of PSA testing among men in Hong Kong, while Lin et al [24] reported a proportion of 29.4% of men in Taiwan who had self-reported having received a PSA test. Lacking knowledge of or misconceptions about PCa and PSA screening, as well as insufficient health promotion, may account for the issue of low PSA uptake coverage in China [23]. As for other Asian countries, except for Japan [25], the uptake rate of PSA screening was at a similarly low level (eg, 16.9% in Iran and 0.4% in the Philippines) [26,27]. Some studies indicate that PSA testing behavior is highly determined by a variety of demographics and lifestyle factors [23,28]. In this study, we identified that Chinese men with older age, a higher BMI, high WC, smoking behavior, drinking behavior, occasional PA, medication use, and comorbidity were more likely to undergo PSA testing, whereas those with lower education background and under widowed were less likely to have undergone PSA testing. These findings help to depict a clearer picture of the status and correlates of PSA testing behavior, which provides important credence for PSA screening practice in China.

We did not find a survival benefit of PSA screening uptake in the general population as well as among annual screeners in this cohort study. One explanation was the self-selection bias [29]. People who attended PSA screening, particularly annual attendees, were more likely to be older, have unhealthy lifestyles, and be in poor physical condition; in turn, these factors were associated with increased PCa mortality. Furthermore, some nonattenders may be potential later screening attendees should they live longer or have long-term follow-up.

The existence of immortal time bias influenced the calculation of true effects on PSA screening [30]. Evidence regarding the optimal screening frequency or interval remains unclear worldwide. In general, annual screening was recommended by guidelines in the United States, while a 2-year interval in Australia, a 3-year interval in Japan, and a 4-year interval in Canada [31]. In Europe, even an 8-year interval was proposed for those who were not at risk [10]. In China, the Anti-Cancer Association Genitourinary Cancer Committee recommends a 2-year interval for serial PSA screening [32]. Several studies have supported the benefit of a moderate expansion for the screening interval. Gulati et al [33] reported that a biennial screening could reduce 2.4% of overdiagnoses and half of the false-positive rate. Leeuwen [34] stated that a 2-year screening interval significantly decreased the advanced PCa incidence. Heijnsdijk [14] indicated that screening with a 2- to 3-year interval was the most cost-effective strategy. As for the survival benefit between annual screening and interval screening, a prior meta-analysis showed that interval screening every 2-4 years was associated with a significantly lower risk of cancer-specific death (incidence rate ratio 0.79; 95% CI 0.69-0.91) than annual screening (incidence rate ratio 1.05; 95% CI 0.87-1.24) [12]. This study suggests a survival benefit of an extended-interval screening strategy, which is consistent with our findings.

In this study, we could not observe significantly lower risks of either PCa-specific mortality or all-cause mortality for people receiving serial screening at least once every 1-3 years. The main reason was cases with limited outcomes in these groups, resulting in relatively lower statistical power. To increase statistical power, participants were dichotomized on the basis of ever having undergone PSA screening once every 4 years. Expectedly, we observed a 70% decreased risk of PCa-specific death and a 23% decreased risk of all-cause death in this group. The result was robust against any unmeasured confounders (E-values are 6.12 for PCa death and 1.92 for overall death; Table S8 in Multimedia Appendix 1). This finding supports evidence that suggests a 2-year screening interval is cost-effective.
serial screening with at least 1 screening every 4 years to improve PCa survival among Chinese men.

Some limitations should be acknowledged in this study. First, selection, information, and immortal time biases were unavoidable in this cohort study, but a large sample size, rigorous analytic strategy, and protracted time frame of follow-up can increase the reliability of our findings. Second, information regarding covariates was poorly documented in the database, but we have used a multiple imputation approach to deal with the missing value. Third, unmeasured confounders may influence the accuracy of the estimates, but E-values were applied to assess the robustness of the results against unmeasured bias, making our findings more reliable and interpretable. Fourth, the limited number of cases of PCa onset and PCa death in the current data set led to insufficient statistical power to identify the optimal PSA screening interval. With longer-term follow-up available and more cases reaching the end point, a significant optimal screening interval would be identified in the future. Finally, the study population was derived from a single municipal district. The implementation of PSA screening and the incidence of PCa may vary greatly in different regions in China. Further studies based on a nationwide population with higher representativeness are warranted.

In conclusion, this large-scale population-based cohort study reveals a cumulative uptake rate of 17.9% for PSA testing, depicts an upward secular trend of screening attendance, and suggests an optimal screening frequency of at least 1 screening every 4 years for serial PSA testing in China. Findings of this study clarify the status of the PSA screening practice, strengthen the evidence base for an extended interval strategy for serial screening, and provide insights into the improvement of PCa survival for patients, health professionals, and policy makers.

Acknowledgments
The authors thank the Yinzhou District Health Bureau (Health and Family Planning Bureau of Yinzhou District) for providing access to the administrative databases used in the study. The authors also thank the Ningbo Wanda Data Application Service Co., Ltd, for data access and providing the service data analytic platform. This work was supported by grants from the National Natural Science Foundation of China (81972645), Shenzhen-Hong Kong-Macau Science and Technology Program (Category C: SGDX2022053011405024), innovative research team of high-level local universities in Shanghai, the Shanghai Youth Talent Support Program, an intramural grant of The University of Hong Kong to Dr Rong Na, and the Shanghai Sailing Program (22YF1440500) to Dr Da Huang. All the funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability
Data of the Chinese Electronic Health Records Research in Yinzhou (CHERRY) study can be applied from the Health and Family Planning Bureau of Yinzhou District for health research purposes. More details are published on the study website [35].

Authors’ Contributions
XR had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. NZ and DW were responsible for data acquisition and administration. Jingyi H, Jinlun H, DH, TTSC, BSHH, ATLN, and JHLT contributed to technical and material support. YZ contributed to data analysis and interpretation and manuscript writing and revision. RN supervised the study and contributed to manuscript review and language editing. RN (narong.hs@gmail.com) and YZ (ylzhan@connect.hku.hk) contributed equally as authors of this paper. RN was the senior author of this paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary materials.

References


Abbreviations

AAPC: average annual percent change

APC: annual percentage change

CHERRY: Chinese Electronic Health Records Research in Yinzhou

EHR: electronic health record

HR: hazard ratio

ICD-10: International Statistical Classification of Diseases, Tenth Revision

PA: physical activity

PCA: prostate cancer

PSA: prostate-specific antigen

PY: person-year

RCT: randomized controlled trial

WC: waist circumference
Assessing Global, Regional, and National Time Trends and Associated Risk Factors of the Mortality in Ischemic Heart Disease Through Global Burden of Disease 2019 Study: Population-Based Study

Tingting Shu¹,², MD; Ming Tang¹,², MD; Bo He¹,², MD; Xiaozhu Liu³, MD; Yu Han¹,², MD; Chang Liu⁴, MD; Pedro A Jose⁵, MD, PhD, Prof Dr; Hongyong Wang¹, MD, PhD; Qing-Wei Zhang⁶*, MD; Chunyu Zeng¹,²,⁷,⁸,⁹*, MD, PhD, Prof Dr

¹Department of Cardiology, Daping Hospital, The Third Military Medical University (Army Medical University), Chongqing, China
²Chongqing Key Laboratory for Hypertension Research, Chongqing Cardiovascular Clinical Research Center, Chongqing Institute of Cardiology, Chongqing, China
³Department of Critical Care Medicine, Beijing Shijitan Hospital, Capital Medical University, Beijing, China
⁴Department of Cardiac Surgery, Daping Hospital, The Third Military Medical University (Army Medical University), Chongqing, China
⁵Division of Renal Diseases and Hypertension, The George Washington University School of Medicine and Health Sciences, Washington, DC, United States
⁶Division of Gastroenterology and Hepatology, Key Laboratory of Gastroenterology and Hepatology, Ministry of Health, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai Institute of Digestive Disease, Shanghai, China
⁷Key Laboratory of Geriatric Cardiovascular and Cerebrovascular Disease Research, Ministry of Education of China, Daping Hospital, The Third Military Medical University (Army Medical University), Chongqing, China
⁸State Key Laboratory of Trauma, Burns and Combined Injury, Daping Hospital, The Third Military Medical University (Army Medical University), Chongqing, China
⁹Cardiovascular Research Center of Chongqing College, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Chongqing, China

*these authors contributed equally

Corresponding Author:
Chunyu Zeng, MD, PhD, Prof Dr
Department of Cardiology
Daping Hospital
The Third Military Medical University (Army Medical University)
10 Daping Branch Road, Yuzhong District
Chongqing, 400016
China
Phone: 86 023 68729501
Email: chunyuzeng01@163.com

Abstract

Background: Ischemic heart disease (IHD) is the leading cause of death among noncommunicable diseases worldwide, but data on current epidemiological patterns and associated risk factors are lacking.

Objective: This study assessed the global, regional, and national trends in IHD mortality and attributable risks since 1990.

Methods: Mortality data were obtained from the Global Burden of Disease 2019 Study. We used an age-period-cohort model to calculate longitudinal age curves (expected longitudinal age-specific rate), net drift (overall annual percentage change), and local drift (annual percentage change in each age group) from 15 to >95 years of age and estimate cohort and period effects between 1990 and 2019. Deaths from IHD attributable to each risk factor were estimated on the basis of risk exposure, relative risks, and theoretical minimum risk exposure level.

Results: IHD is the leading cause of death in noncommunicable disease–related mortality (118.1/598.8, 19.7%). However, the age-standardized mortality rate for IHD decreased by 30.8% (95% CI –34.83% to –27.17%) over the past 30 years, and its net drift ranged from –2.89% (95% CI –3.07% to –2.71%) in high sociodemographic index (SDI) region to –0.24% (95% CI –0.32% to –0.16%) in low-middle–SDI region. The greatest decrease in IHD mortality occurred in the Republic of Korea (high SDI) with...
net drift –6.06% (95% CI –6.23% to –5.88%), followed by 5 high-SDI nations (Denmark, Norway, Estonia, the Netherlands, and Ireland) and 2 high-middle–SDI nations (Israel and Bahrain) with net drift less than –5.00%. Globally, age groups of >60 years continued to have the largest proportion of IHD-related mortality, with slightly higher mortality in male than female group. For period and birth cohort effects, the trend of rate ratios for IHD mortality declined across successive period groups from 2000 to 2004 and birth cohort groups from 1985 to 2000, with noticeable improvements in high-SDI regions. In low-SDI regions, IHD mortality significantly declined in female group but fluctuated in male group across successive periods; sex differences were greater in those born after 1945 in middle- and low-middle–SDI regions and after 1970 in low-SDI regions. Metabolic risks were the leading cause of mortality from IHD worldwide in 2019. Moreover, smoking, particulate matter pollution, and dietary risks were also important risk factors, increasingly occurring at a younger age. Diets low in whole grains and legumes were prominent dietary risks in both male and female groups, and smoking and high-sodium diet mainly affect male group.

Conclusions: IHD, a major concern, needs focused health care attention, especially for older male individuals and those in low-SDI regions. Metabolic risks should be prioritized for prevention, and behavioral and environmental risks should attract more attention to decrease IHD mortality.

(JMIR Public Health Surveill 2024;10:e46821) doi:10.2196/46821

KEYWORDS
age-period-cohort analysis; GBD 2019; Global Burden of Disease 2019 study; ischemic heart disease; mortality; risk factors

Introduction

Ischemic heart disease (IHD) is the leading cause of death among noncommunicable diseases (NCDs) and cardiovascular diseases (CVDs) in the world, accounting for 9.14 million deaths in 2019 [1,2]. The United Nations set, as part of the Sustainable Development Goals, a global target to reduce by one-third the total premature mortality (ages 30-70 years) from NCDs by 2030 [3]. However, the trend and distribution of IHD-related mortality vary greatly in countries from high sociodemographic index (SDI) to low SDI [4]. In Central Asia, the age-standardized mortality rate (ASMR) increased by 16.7% from 1990 to 2017 [5]. By contrast, in Korea, Japan, and China, the ASMR decreased significantly in both male and female groups during this time period [6,7].

Rapid economic development and social advancements have greatly increased the incidence of metabolic risk factors, such as hypertension, hyperlipidemia, diabetes, obesity, and aging, which have largely contributed to the majority of IHD-related deaths shifting from developed to developing countries [2,8]. However, the relative contribution of risk factors to mortality from IHD varies with age, sex, metabolic state, behavior, geographical area, and environmental exposure, and therefore different interventions may be required to effectively address these risk factors [9].

Previous studies have also explored IHD epidemic trends over time in many countries [2,10]. However, typical statistical analysis cannot break down the risks when estimating mortality. Therefore, the age-period-cohort (APC) model was developed to improve the traditional method for mortality description and analysis, which separates age effects, period effects, and cohort effects and quantifies the influence of age, time, and birth cohort factors on mortality rate [11,12]. In this study, we used data from the Global Burden of Disease 2019 Study (GBD) to assess the time trends in IHD mortality at the global, regional, and national levels and associations with age, period, and birth cohort using the APC model over the past 30 years. We also aimed to shed light on attributable risk factors for IHD at global, regional, and national levels to support policy makers making informed decisions about the potential benefits of risk reduction policies.

Methods

Data Sources

GBD, coordinated by the Institute for Health Metrics and Evaluation, which assessed the global burden of 369 diseases and injuries and 87 risk factors in 204 countries and territories from 1990 to 2019, provided a unique opportunity to understand the global burden and landscape of IHD mortality [13,14]. In our study, we obtained the publication estimates of deaths for IHD with “ischemic heart disease” from the “cause blank” of the GBD website across 204 countries and territories [2,13]. The statistical code used for GBD estimation is publicly available on the internet [15]. The GBD study uses deidentified data, and a waiver of informed consent was reviewed and approved by the University of Washington Institutional Review Board. All results and original data sources used to compute estimates are publicly available on the internet [16].

SDI was used to quantify the development level for each location year, which is scaled from 0 to 1, with higher values indicating high SDI levels. GBD assigned 204 countries and territories into 5 groups by SDI quintiles: low SDI, low-middle SDI, middle SDI, high-middle SDI, and high SDI [13,14].

Statistical Analyses

The overall temporal trends in mortality over the study period were assessed by all-age mortality rate (AAMR), ASMR, percentage of mortality, and the relative change in percentage between 1990 and 2019. We calculated ASMR with global age-standard population data from GBD [14].

APC models were developed to determine independent effect estimates of age, period, and birth cohort on IHD mortality [17]. Net drift and local drift are important parameters in the APC models. Net drift represents the overall log-linear trend by period and birth cohort, which indicates the overall annual percentage change of the expected age-adjusted rates over time. Local drift represents the log-linear trend by period and cohort for each
age group, which indicates the annual percentage change. The longitudinal age curve indicates the expected age-specific rate in a reference cohort adjusted for period effects, which was assigned to 19 age groups using successive 5-year age intervals from 15-19 years to over 95 years among individuals with IHD. The age effects represent a differing risk of the outcome associated with different age brackets. The period effects represent variations in the outcome over time that influence all age groups simultaneously. The mortality and population data were arranged into consecutive 5-year periods from 1990 to 2019, with the mid-year of the 2000-2004 survey year as the reference period group [18]. Cohort effects are the ratio of age-specific rates of mortality risks, which were arranged into consecutive 5-year cohorts from the 1895 cohort to the 2000 cohort, as referenced by the mid-year of birth (1945 cohort) [18]. All analyses were conducted with 3 sex groups: the male group, the female group, and both groups (male and female).

A total of 87 risk factors associated with IHD-related death were selected in accordance with the World Cancer Research Fund grades of convincing or probable evidence [19,20]. To estimate the IHD-related deaths attributable to a specific risk, these IHD-related deaths were multiplied by the population attributable fractions (PAF) for the IHD risk-outcome pair for a given sex, age, year, and location. PAF show that the proportion by which the deaths would decrease in a specific year if the exposure to a risk factor in the past was equal to the theoretical minimum risk exposure level, which were adjusted by other risk factors included in this study [21,22].

We obtained the estimated parameters from the APC web tool (freely available R tools) provided by the National Cancer Institute of the United States, the methodological details of which have been described previously [23]. We hypothesized that changes in age, period, and birth cohort would mainly affect IHD mortality. The Joinpoint Trend Analysis Software (version 4.9; National Cancer Institute) was used to determine the average annual percentage change in IHD mortality from 1990 to 2019 [24]. Wald chi-square tests were adapted for the significance of the estimable parameters and functions. Statistical tests were 2-sided, and a P value <.05 was considered significant. We performed all the analysis in R software (version 3.6.3; R Foundation for Statistical Computing).

Ethical Considerations

Ethics approval and consent are not required because the original data in this study were aggregated and nonidentified, which are freely available from the GBD Study.

Results

Overview and Trends for IHD Mortality

In 2019, IHD was the leading cause of NCD-related AAMR in the global and each SDI quintile, accounting for 118.10 deaths per 10,000 (95% UI [uncertainty interval] 125.93-110.51) population worldwide (Table 1 and Figure 1A). IHD-related AAMRs were lower in low-SDI quintiles and about 35% higher in male group than female group. IHD had the largest proportion of NCD-related mortality with 19.7% (118.1/598.8) in all ages and maintained this dominant proportion in the >50-year-old age group (Figure 1B).

Globally, the number of mortalities for IHD increased by 60.43% from 1990 to 2019, and the ASMR for IHD decreased by 30.8%, with marked reduction in high-SDI (–58.68%) and high-middle–SDI (–35.28%) regions (Table 2). However, ASMR for IHD increased in 4 lower-income GBD regions, with a percentage change of 15.75% in East Asia, 11.71% in Central Asia, 8.31% in Oceania, and 4.63% in Southern Sub-Saharan Africa (Table 2).

Over the past 30 years (1990-2019), the IHD mortality rate increased in the older age groups (>60 years), but the increase declined in later periods that occurred mainly in the high-SDI regions (Figure S1 in Multimedia Appendix 1). There was a relatively declining risk of IHD mortality in cohorts born before 1945 in high-SDI regions, but the risk remained high in the low-SDI regions (Figure S1 in Multimedia Appendix 1). Similar to the trend by global and 5 SDI quintiles, IHD ASMRs across nations increased with age and declined between 1990-1994 and 2015-2019 (Figures S1A-S1F in Multimedia Appendix 1).

There was a downward trend of mortality for IHD in high-SDI countries across successive birth cohorts, but it fluctuated and trended to increase with aging in low-SDI countries (Figures S2A-S2F in Multimedia Appendix 1).

Among 204 countries and territories in 2019, ASMRs for IHD in 107 countries were higher than the global average, 32 of which were more than 2-fold higher (Figures 2A and 2B and Table S1 in Multimedia Appendix 2). The highest ASMR for IHD was observed in Uzbekistan (707.51, 95% UI 638.23-780.68 per 100,000 population; middle SDI) with the largest increase of 119.01% (95% UI 96.32%-143.33%) over 30 years, followed by Azerbaijan (middle SDI) with 35.76% (95% UI 19.27%-54.74%) increase, and Tajikistan (low-middle SDI), with 82.1% (95% UI 53.66%-116.8%) increase. However, the net drift increased in Uzbekistan (1.98%, 95% CI 1.8%-2.17%), decreased in Azerbaijan (–0.95%, 95% CI –1.23% to –0.68%), and was essentially unchanged in Tajikistan (0.52%, 95% CI 0.13%-0.9%). In Qatar (high SDI), deaths related to IHD increased by 203.32% (95% CI 126.08%-308.55%), with a net drift of –3.69% (95% CI –4.28% to –3.09%). In Guam (high SDI) and Liberia (low SDI), ASMR in IHD decreased by 14.32% (95% CI –29.31% to 2.29%) and 14.04% (95% CI –33.47% to 10.41%), but the net drifts in mortality were 0.27% (95% CI –0.88% to 1.44%) and –0.36% (95% CI –0.77% to 0.05%), respectively. China and India, the 2 most populous countries, had the largest number of deaths with 1.87 (95% CI 1.61-2.13) million and 1.52 (95% CI 1.31-1.75) million, respectively, with relatively modest change of net drifts, 0.13% in China and –0.48% in India. These results show that IHD mortality trends were uneven across countries and not strictly commensurate with SDI levels. The directions of change in IHD mortality trends were uneven across countries and not strictly consistent with net drift from the APC model, suggesting the necessity to differentiate the relative contributions of period and cohort effects in IHD mortality.
JMIR PUBLIC HEALTH AND SURVEILLANCE

Shu et al

Table 1. Trends of death cases and all-age mortality rate (AAMR) in ischemic heart disease across 5 sociodemographic index (SDI) quintiles and 21
Location

Number of death cases in
a

1990 (95% UI )

Number of death
cases in 2019 (95%
UI)

Percentage change of
cases (95% UI)

AAMR in 1990
per 100,000
population
(95% UI)

AAMR in 2019 Percentage change
per 100,000
of AAMR (95% UI)
population
(95% UI)

Global

5,695,890 (5,405,191 to
5,895,398)

9,137,791
(8,395,682 to
9,743,550)

60.43 (50.23 to 69.14) 106.47 (101.03
to 110.2)

118.1 (108.51
to 125.93)

10.92 (3.88 to 16.94)

High SDI

1,688,791 (1,572,959 to
1,744,989)

1,447,271
(1,270,022 to
1,553,837)

–14.3 (–19.35 to
–9.59)

205.45 (191.35
to 212.28)

142.82 (125.32
to 153.33)

–30.49 (–34.58 to
–26.67)

High-middle 1,870,951 (1,782,685 to
SDI
1,923,524)

2,658,293
(2,411,867 to
2,832,480)

42.08 (32.95 to 50.53) 162.63 (154.96
to 167.2)

185.84 (168.61
to 198.02)

14.27 (6.92 to 21.07)

Middle SDI

1,151,131 (1,087,976 to
1,217,989)

2,824,552
(2,576,474 to
3,047,018)

145.37 (122.94 to
167)

67.05 (63.37 to
70.95)

117.86 (107.51
to 127.14)

75.77 (59.7 to 91.26)

Low-middle
SDI

712,691 (654,082 to
773,458)

1,646,065
(1,488,073 to
1,801,773)

130.96 (104.7 to
156.95)

63.09 (57.9 to
68.47)

93.32 (84.36 to
102.14)

47.91 (31.09 to
64.55)

Low SDI

269,141 (240,679 to
301,919)

556,602 (495,167 to 106.81 (78.43 to
627,058)
134.17)

50.96 (45.57 to
57.17)

49.31 (43.87 to
55.56)

–3.23 (–16.51 to
9.58)

High-income 149,217 (139,101 to
Asia Pacific 155,395)

172,964 (139,728 to 15.91 (0.45 to 25.23)
191,814)

86 (80.17 to
89.56)

92.35 (74.6 to
102.42)

7.39 (–6.94 to 16.01)

High-income 655,487 (606,259 to
North Ameri- 680,515)
ca

606,489 (539,683 to –7.48 (–11.06 to
647,145)
–3.11)

233.33 (215.81
to 242.24)

166.36 (148.04
to 177.51)

–28.7 (–31.46 to
–25.34)

Western Europe

916,253 (860,097 to
945,855)

666,760 (586,750 to –27.23 (–32.04 to
718,140)
–22.79)

238.24 (223.64
to 245.94)

152.82 (134.48
to 164.6)

–35.86 (–40.1 to
–31.94)

Australasia

40,271 (37,696 to 41,643)

33,125 (28,421 to
35,760)

–17.74 (–24.2 to
–13.33)

198.6 (185.9 to
205.36)

113.97 (97.79
to 123.04)

–42.61 (–47.12 to
–39.53)

Andean
19,641 (17,398 to 21,868)
Latin America

34,446 (28,404 to
40,971)

75.37 (41.25 to
114.39)

51.45 (45.57 to
57.28)

54.16 (44.66 to
64.42)

5.28 (–15.21 to 28.7)

Tropical
119,468 (113,614 to
Latin Ameri- 123,476)
ca

175,999 (160,324 to 47.32 (39.13 to 54.49) 78.14 (74.32 to
185,401)
80.77)

78.71 (71.7 to
82.92)

0.73 (–4.87 to 5.63)

Central Latin 92,607 (87,113 to 96,026)
America

219,022 (189,141 to 136.51 (109.03 to
250,607)
167.84)

56.43 (53.08 to
58.51)

87.6 (75.65 to
100.23)

55.25 (37.22 to
75.82)

Southern
66,488 (62,693 to 68,727)
Latin America

61,452 (55,939 to
65,419)

–7.58 (–12.51 to –2.8) 134.21 (126.55
to 138.72)

92.06 (83.8 to
98)

–31.41 (–35.07 to
–27.86)

Caribbean

44,030 (41,268 to 46,039)

63,535 (55,083 to
72,649)

44.3 (27.09 to 62.53)

134.7 (116.78
to 154.03)

7.91 (–4.95 to 21.55)

Central Europe

389,917 (371,996 to
398,754)

354,125 (308,352 to –9.18 (–18.94 to 0.39) 317.09 (302.52
395,305)
to 324.28)

310.03 (269.95
to 346.08)

–2.23 (–12.74 to
8.07)

Eastern Europe

796,168 (762,557 to
815,230)

986,560 (879,471 to 23.91 (13.8 to 34.01)
1,075,171)

469.86 (418.85
to 512.06)

33.67 (22.76 to
44.56)

124.82 (116.99
to 130.52)

351.5 (336.66
to 359.92)

Central Asia 131,655 (124,727 to
135,840)

200,135 (183,535 to 52.01 (40.02 to 65.82) 190.07 (180.07
218,271)
to 196.11)

213.98 (196.23
to 233.37)

12.58 (3.69 to 22.8)

North Africa 444,690 (411,688 to
and Middle 478,245)
East

799,484 (706,349 to 79.78 (58.25 to
909,787)
100.46)

128.89 (119.32
to 138.61)

131.34 (116.04
to 149.46)

1.9 (–10.3 to 13.62)

South Asia

1,857,949
(1,633,946 to
2,091,633)

68.74 (61.52 to
75.62)

102.92 (90.51
to 115.87)

49.73 (25.75 to
72.57)

754,477 (675,262 to
830,037)

https://publichealth.jmir.org/2024/1/e46821

XSL• FO
RenderX

146.26 (106.82 to
183.81)

JMIR Public Health Surveill 2024 | vol. 10 | e46821 | p.1336
(page number not for citation purposes)


<table>
<thead>
<tr>
<th>Location</th>
<th>Number of death cases in 1990 (95% UI)</th>
<th>Number of death cases in 2019 (95% UI)</th>
<th>Percentage change of cases (95% UI)</th>
<th>AAMR in 1990 per 100,000 population (95% UI)</th>
<th>AAMR in 2019 per 100,000 population (95% UI)</th>
<th>Percentage change of AAMR (95% UI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southeast Asia</td>
<td>248,618 (226,364 to 270,439)</td>
<td>588,556 (527,652 to 644,283)</td>
<td>136.73 (107.34 to 167.98)</td>
<td>53.26 (48.49 to 57.94)</td>
<td>87.35 (78.31 to 95.62)</td>
<td>64.01 (43.64 to 85.66)</td>
</tr>
<tr>
<td>East Asia</td>
<td>631,081 (559,209 to 703,062)</td>
<td>1,926,477 (1,662,056 to 2,182,277)</td>
<td>205.27 (158.4 to 261.21)</td>
<td>51.51 (45.64 to 57.39)</td>
<td>130.86 (112.9 to 148.23)</td>
<td>154.04 (115.04 to 200.59)</td>
</tr>
<tr>
<td>Oceania</td>
<td>4,771 (3.879 to 6.050)</td>
<td>12,296 (9.869 to 15,450)</td>
<td>157.72 (112.98 to 215.56)</td>
<td>73.75 (59.95 to 93.51)</td>
<td>92.62 (74.33 to 116.37)</td>
<td>25.59 (3.79 to 53.78)</td>
</tr>
<tr>
<td>Western Sub-Saharan Africa</td>
<td>87,324 (72,430 to 108,234)</td>
<td>166,504 (139,342 to 195,399)</td>
<td>90.67 (45.29 to 128.5)</td>
<td>45.34 (37.61 to 56.2)</td>
<td>36.49 (30.54 to 42.82)</td>
<td>–19.53 (–38.68 to –3.56)</td>
</tr>
<tr>
<td>Eastern Sub-Saharan Africa</td>
<td>59,328 (51,116 to 66,823)</td>
<td>120,310 (96,907 to 143,565)</td>
<td>102.79 (52.77 to 151.07)</td>
<td>31.2 (26.88 to 35.14)</td>
<td>29.22 (23.53 to 34.86)</td>
<td>–6.35 (–29.45 to 15.95)</td>
</tr>
<tr>
<td>Central Sub-Saharan Africa</td>
<td>22,977 (19,148 to 28,107)</td>
<td>47,752 (36,504 to 62,352)</td>
<td>107.83 (66.12 to 162.71)</td>
<td>41.38 (34.49 to 50.63)</td>
<td>36.3 (27.75 to 47.4)</td>
<td>–12.28 (–29.89 to 10.88)</td>
</tr>
<tr>
<td>Southern Sub-Saharan Africa</td>
<td>21,423 (19,366 to 23,428)</td>
<td>43,851 (39,702 to 48,020)</td>
<td>104.69 (86.5 to 125.45)</td>
<td>40.81 (36.89 to 44.63)</td>
<td>55.81 (50.53 to 61.11)</td>
<td>36.75 (24.59 to 50.62)</td>
</tr>
</tbody>
</table>

UI: uncertainty interval.
**Figure 1.** The leading cause of global mortality from noncommunicable diseases (NCDs) in the all-age population. (A) All-age mortality rate in 2019 for NCDs in female and male, individually, and as a group. (B) The proportion of the mortality rate in 2019 for NCDs among age groups. CVD: cardiovascular disease; SDI: sociodemographic index.

**(A)**

<table>
<thead>
<tr>
<th>Cause</th>
<th>Both</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ischemic heart disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>18.6</td>
<td>14.2</td>
<td>22.9</td>
</tr>
<tr>
<td>Other cancer</td>
<td>10.5</td>
<td>14.5</td>
<td>6.9</td>
</tr>
<tr>
<td>Chronic respiratory diseases</td>
<td>38.6</td>
<td>43.5</td>
<td>32.4</td>
</tr>
<tr>
<td>Diabetes and kidney disease</td>
<td>30.6</td>
<td>40.9</td>
<td>26.3</td>
</tr>
<tr>
<td>Digestive diseases</td>
<td>28.9</td>
<td>29.8</td>
<td>28.0</td>
</tr>
<tr>
<td>Neurological disorders</td>
<td>26.7</td>
<td>27.2</td>
<td>26.1</td>
</tr>
<tr>
<td>Tracheal, bronchial, and lung cancer</td>
<td>26.1</td>
<td>59.9</td>
<td>24.4</td>
</tr>
<tr>
<td>Other CVD</td>
<td>22.7</td>
<td>39.5</td>
<td>13.2</td>
</tr>
<tr>
<td>Hypertensive heart disease</td>
<td>14.9</td>
<td>16.7</td>
<td>13.5</td>
</tr>
<tr>
<td>Other noncommunicable diseases</td>
<td>14.7</td>
<td>19.2</td>
<td>10.3</td>
</tr>
<tr>
<td>Colon and rectal cancer</td>
<td>14.0</td>
<td>32.2</td>
<td>6.6</td>
</tr>
<tr>
<td>Stomach cancer</td>
<td>12.7</td>
<td>14.9</td>
<td>10.4</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>9.6</td>
<td>15.3</td>
<td>6.6</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>8.8</td>
<td>17.3</td>
<td>5.2</td>
</tr>
<tr>
<td>Substance use disorders</td>
<td>3.3</td>
<td>11.6</td>
<td>5.3</td>
</tr>
<tr>
<td>Musculoskeletal disorders</td>
<td>1.5</td>
<td>3.9</td>
<td>1.5</td>
</tr>
<tr>
<td>Skin and subcutaneous diseases</td>
<td>1.2</td>
<td>2.3</td>
<td>1.4</td>
</tr>
<tr>
<td>Mental disorders</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**(B)**

<table>
<thead>
<tr>
<th>Cause</th>
<th>Ischemic heart disease</th>
<th>Stroke</th>
<th>Other cancer</th>
<th>Chronic respiratory diseases</th>
<th>Diabetes and kidney disease</th>
<th>Digestive diseases</th>
<th>Neurological disorders</th>
<th>Tracheal, bronchial, and lung cancer</th>
<th>Other CVD</th>
<th>Hypertensive heart disease</th>
<th>Other noncommunicable diseases</th>
<th>Colon and rectal cancer</th>
<th>Stomach cancer</th>
<th>Breast cancer</th>
<th>Pancreatic cancer</th>
<th>Substance use disorders</th>
<th>Musculoskeletal disorders</th>
<th>Skin and subcutaneous diseases</th>
<th>Mental disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke</td>
<td>18.6</td>
<td>14.2</td>
<td>22.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other cancer</td>
<td>10.5</td>
<td>14.5</td>
<td>6.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic respiratory diseases</td>
<td>38.6</td>
<td>43.5</td>
<td>32.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes and kidney disease</td>
<td>30.6</td>
<td>40.9</td>
<td>26.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digestive diseases</td>
<td>28.9</td>
<td>29.8</td>
<td>28.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurological disorders</td>
<td>26.7</td>
<td>27.2</td>
<td>26.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tracheal, bronchial, and lung cancer</td>
<td>26.1</td>
<td>59.9</td>
<td>24.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other CVD</td>
<td>22.7</td>
<td>39.5</td>
<td>13.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertensive heart disease</td>
<td>14.9</td>
<td>16.7</td>
<td>13.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other noncommunicable diseases</td>
<td>14.7</td>
<td>19.2</td>
<td>10.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon and rectal cancer</td>
<td>14.0</td>
<td>32.2</td>
<td>6.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stomach cancer</td>
<td>12.7</td>
<td>14.9</td>
<td>10.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>9.6</td>
<td>15.3</td>
<td>6.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>8.8</td>
<td>17.3</td>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substance use disorders</td>
<td>3.3</td>
<td>11.6</td>
<td>5.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal disorders</td>
<td>1.5</td>
<td>3.9</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skin and subcutaneous diseases</td>
<td>1.2</td>
<td>2.3</td>
<td>1.4</td>
<td>1.08</td>
<td>1.18</td>
<td>1.43</td>
<td>2.71</td>
<td>1.65</td>
<td>2.13</td>
<td>1.88</td>
<td>1.95</td>
<td>1.93</td>
<td>1.83</td>
<td>1.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental disorders</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
JMIR PUBLIC HEALTH AND SURVEILLANCE

Shu et al

Table 2. Trends of age-standardized mortality rate (ASMR) and net drift in ischemic heart disease across 5 socio-demographic index (SDI) quintiles
Location

ASMR in 1990 per
100,000 population (95%
UIa)

ASMR in 2019 per
100,000 population (95%
UI)

Percentage change of
ASMR (95% UI)

AAPCb, % (95% CI)

Net drift, % (95% CI)

Global

170.45 (159.61 to 176.94) 117.95 (107.83 to 125.92) –30.8 (–34.83 to –27.17) –1.26 (–1.38 to –1.13) –1.15 (–1.21 to –1.09)

High SDI

162.39 (150.62 to 168.15) 67.1 (60.07 to 71.54)

High-middle
SDI

209.2 (196.25 to 216.23)

Middle SDI

143.11 (133.18 to 152.13) 134.12 (121.51 to 145.22) –6.28 (–14.41 to 1.84)

–0.22 (–0.35 to –0.1)

Low-middle
SDI

144.21 (132.05 to 156.58) 136.59 (122.96 to 149.5)

–5.28 (–15.25 to 5.09)

–0.14 (–0.19 to –0.08) –0.24 (–0.32 to –0.16)

Low SDI

139.2 (124 to 156.79)

127.99 (113.13 to 143.9)

–8.05 (–20.59 to 3.17)

–0.29 (–0.41 to –0.17) –0.43 (–0.5 to –0.37)

High-income
Asia Pacific

85.48 (78.44 to 89.52)

30.47 (25.73 to 33.34)

–64.36 (–67.28 to
–62.35)

–3.51 (–3.77 to –3.25) –3.46 (–3.73 to –3.19)

High-income
North America

179.79 (166.53 to 186.54) 88.09 (79.57 to 93.4)

Western Europe 156.31 (146.05 to 161.8)

–58.68 (–60.3 to –56.69) –3.08 (–3.19 to –2.97) –2.89 (–3.07 to –2.71)

135.41 (122.68 to 144.44) –35.28 (–39.01 to
–31.69)

–1.56 (–1.77 to –1.35) –2 (–2.17 to –1.83)
–0.31 (–0.37 to –0.25)

–51.01 (–52.49 to –48.8) –2.47 (–2.59 to –2.35) –2.4 (–2.57 to –2.24)

60.69 (54.18 to 64.89)

–61.17 (–62.94 to
–59.05)

–3.27 (–3.39 to –3.14) –3.92 (–4.2 to –3.64)

Australasia

178.69 (165.54 to 185.48) 59.04 (51.31 to 63.44)

–66.96 (–68.94 to
–65.45)

–3.79 (–3.95 to –3.62) –4.32 (–4.66 to –3.97)

Andean Latin
America

109.84 (97.27 to 121.84)

64.48 (52.96 to 76.52)

–41.29 (–52.35 to –28.5) –1.81 (–1.98 to –1.63) –1.98 (–2.07 to –1.88)

Tropical Latin
America

156.35 (145.4 to 162.47)

75.23 (68.22 to 79.42)

–51.88 (–54.21 to
–49.69)

–2.48 (–2.57 to –2.4)

Central Latin
America

131.13 (121.2 to 136.69)

97.37 (83.84 to 111.34)

–25.74 (–33.94 to
–16.03)

–1.12 (–1.35 to –0.88) –1.09 (–1.19 to –0.99)

Southern Latin
America

161.14 (150.08 to 167.35) 72.05 (65.77 to 76.66)

Caribbean

185.06 (172 to 193.88)

Central Europe

300.44 (283.07 to 308.93) 159.8 (139.02 to 178.42)

Eastern Europe

326.57 (308.96 to 336.05) 284.55 (253.63 to 310.14) –12.86 (–19.52 to –6.08) –0.6 (–1.04 to –0.15)

–1.19 (–1.49 to –0.9)

Central Asia

322.28 (301.66 to 333.6)

–0.62 (–0.76 to –0.48)

–55.29 (–57.27 to –53.1) –2.76 (–2.96 to –2.56) –2.74 (–2.95 to –2.53)

122.12 (106.01 to 139.69) –34.01 (–41.86 to
–25.76)

360.02 (330.9 to 390.24)

–2.18 (–2.27 to –2.1)

–46.81 (–52.44 to
–41.39)

11.71 (3.8 to 20.22)

–1.4 (–1.64 to –1.16)

–1.16 (–1.25 to –1.07)

–2.19 (–2.43 to –1.96) –3.34 (–3.59 to –3.09)

0.41 (0.22 to 0.6)

North Africa
309.32 (284.33 to 332.13) 219.01 (194.15 to 246.75) –29.2 (–36.92 to –21.85) –1.19 (–1.31 to –1.07) –1.71 (–1.77 to –1.64)
and Middle East
South Asia

159.72 (143 to 177.5)

149.21 (130.6 to 167.88)

Southeast Asia

117.63 (106.32 to 128.2)

112.63 (100.54 to 122.88) –4.26 (–16.26 to 7.52)

–0.13 (–0.18 to –0.08) –0.06 (–0.1 to –0.03)

East Asia

98.76 (88.2 to 109.52)

114.31 (98.93 to 128.75)

15.75 (–0.97 to 36.12)

0.52 (0.17 to 0.86)

0.12 (–0.04 to 0.28)

Oceania

183.67 (150.74 to 228.78) 198.93 (162.3 to 246.64)

8.31 (–8.61 to 29.3)

0.26 (0.19 to 0.34)

0.33 (0.18 to 0.48)

Western SubSaharan Africa

126.66 (106.08 to 154.95) 114.61 (95.9 to 132.21)

–9.51 (–31.88 to 7.22)

–0.27 (–0.39 to –0.15) –0.54 (–0.61 to –0.47)

–6.09 (–27.99 to 13.35)

–0.26 (–0.31 to –0.22) –0.64 (–0.69 to –0.6)

Eastern Sub-Sa- 99.72 (86.19 to 111.65)
haran Africa

a

93.65 (74.72 to 111.98)

–6.59 (–21.1 to 6.7)

–0.25 (–0.35 to –0.14) –0.27 (–0.39 to –0.16)

Central Sub-Sa- 130.41 (108.75 to 160.36) 116.78 (88.97 to 150.45)
haran Africa

–10.45 (–27.43 to 10.48) –0.38 (–0.53 to –0.23) –0.66 (–0.74 to –0.58)

Southern SubSaharan Africa

4.63 (–4.08 to 14.8)

89.43 (80.06 to 97.89)

93.57 (84.43 to 102.05)

0.45 (0.13 to 0.78)

–0.64 (–0.85 to –0.43)

UI: uncertainty interval.

b

AAPC: average annual percentage change.

https://publichealth.jmir.org/2024/1/e46821

XSL• FO
RenderX

JMIR Public Health Surveill 2024 | vol. 10 | e46821 | p.1339
(page number not for citation purposes)


Figure 2. The age-standardized mortality in 2019 and net drift of mortality during 1990-2019 for ischemic heart disease in 204 countries and territories. 
(A) World map of age-standardized mortality for ischemic heart disease in 2019. (B) World map of net drifts for ischemic heart disease mortality, that is, estimated annual percentage change of mortality from the age-period-cohort model. Net drift captures components of the trends attributable to calendar time and successive birth cohorts.

### Time Trends for IHD Mortality in APC Test

Globally, the overall net drift for IHD mortality was $-1.15\%$ (95% CI $-1.21\%$ to $-1.09\%$), which was greater in the female ($-1.33\%, \; 95\% \; CI \; -1.41\%$ to $-1.25\%$) than in the male group ($-1.04\%, \; 95\% \; CI \; -1.11\%$ to $-0.96\%$), with local drifts fluctuating downward across all age groups ($<0\%; \; P<.001$; Tables S2 and S3 in Multimedia Appendix 2 and Figure 3A).

The global local drift went below the net drift after the 55- to 59-year-old age group and reached nadirs in 65- to 69-year-old age group ($-1.54\%, \; 95\% \; CI \; -1.59\%$ to $1.48\%$) and >95-year-old age group ($-1.61\%, \; 95\% \; CI \; -1.79\%$ to $-1.43\%$). Among SDI quintiles, IHD mortality was substantially reduced in high-SDI (net drift $-2.89\%, \; 95\% \; CI \; -3.07\%$ to $-2.71\%$)
high-middle–SDI (net drift –2.00%, 95% CI –2.17% to –1.83) regions, especially between the ages of 40 and >95 years and 35-69 years (local drift –2.0%), respectively. However, the reductions in IHD mortality in low-SDI regions were less striking, with the death risk increasing with age, especially in the low-SDI region.

Figure 3. Time trends and age-period-cohort effects on ischemic heart disease mortality by global and sociodemographic index (SDI) quintiles, 1990-2019. (A) Net and local drifts of ischemic heart disease mortality (estimates from age-period-cohort models) for 19 age groups (5-9 to >95 years), 1990-2019. The dots and shaded areas indicate the annual percentage change of mortality (%) and the corresponding 95% CIs. The topmost black horizontal line marks “0” change. The horizontal discontinuous lines show overall annual percentage changes in mortality (net drifts), with yellow for female, gray for male, and blue for both. The continuous solid lines show annual percentage changes from 5-9 to >95 years (local drifts), in male (gray), female (yellow), and blue for both. (B) Temporal change in the relative proportion of ischemic heart disease mortality across age groups, 1990-2019. (C) Age effects are shown by the fitted longitudinal age curves of mortality (per 100,000 person-years) adjusted for period deviations. (D) Period effects are shown by the relative risk of mortality (mortality rate ratio) and computed as the ratio of age-specific rates (adjusted for age and nonlinear period effects) from 1990-1994 to 2015-2019 (2000-2004 as the referent period). (F) Cohort effects are shown by the relative risk of mortality and computed as the ratio of age-specific rates (adjusted for age and nonlinear period effects) from the 1895 cohort to the 2000 cohort, with the referent cohort set at 1945. The dots and shaded areas denote mortality rates or rate ratios and their corresponding 95% CIs.

At the national level, 43 countries and territories had increased trends (net drifts ≥0.0%) and 29 had modest reductions (–0.5 to 0.0%) in mortality from IHD (Figure 2B and Figures S3A-S3F in Multimedia Appendix 1). There were 9 nations with net drifts >1.00%, with 2 from middle-SDI (Philippines and Uzbekistan), 5 from low-middle–SDI (Lesotho, Zimbabwe, the Dominican Republic, Timor-Leste, and Kenya), and 2 from low-SDI (Mozambique and Guinea) regions. Large decreases in IHD mortality were estimated in 90 nations (net drift <–2.00%), in which the biggest decline occurred in the Republic of Korea (high SDI) with net drift –6.06% (95% CI –6.23% to –5.88%), followed by nations from 5 high-SDI (Denmark, Norway, Estonia, the Netherlands, and Ireland) and 2 high-middle–SDI (Israel and Bahrain) regions with net drift –5.00%. With significant decreases in IHD-related deaths over the past 30 years, Saint Lucia (net drift –3.21%, 95% CI –5.33% to –1.04%) and Equatorial Guinea (–2.9%, 95% CI –3.75% to –2.03%) were typical countries in middle-SDI, the Maldives (–4.03%, 95% CI –4.98% to –3.06%) and Belize (–2.18%, 95% CI –3.34% to –0.93%) in low-middle–SDI, and Afghanistan (–1.35%, 95% CI –1.43% to –1.27%) and Yemen (–1.05%, 95% CI –1.14% to –0.95%) in low-SDI regions. Temporal changes in the age distribution of death number and death percent, as an indirect marker of IHD population survival, are shown in Figures 3B and 3C and Figures S4A-S4F and S5A-S5F in Multimedia Appendix 1.

Sex variations in time trends of IHD mortality are presented in some regions and nations with local drifts. Globally, the local drift in male group resembled that in female group, reaching

https://publichealth.jmir.org/2024/1/e46821

JMIR Public Health Surveill 2024 | vol. 10 | e46821 | p.1341

(page number not for citation purposes)
the bottom in the age group of 65-69 years (female: –1.64%, 95% CI –1.71% to –1.56%; male: –1.51%, 95% CI –1.58% to –1.43%), but the reduction in IHD mortality in male group aged 45 years or younger was not as steep as in female group, which was more pronounced in low-SDI regions (Figure 3A). For high- and high-middle–SDI regions, the improvements in trends in IHD mortality were mostly similar in male and female groups (Figures S3B and S3C in Multimedia Appendix 1). It is worth noting that in American Samoa (high-middle SDI), female group had a significantly higher IHD mortality than male group. For lower-income nations (middle, low-middle, and low SDI), the time trends for IHD mortality were greater in female than male group in many countries (Figures S3D-S3F in Multimedia Appendix 1).

APC Effects on IHD Mortality
For age effect, IHD mortality increased with age, with a sharp increase in the later years of life (>60 years; Figure 3D). Similar patterns of IHD mortality trends were found across different SDI quintiles, with modestly higher mortality in male than female group, but high-SDI region showed an overall lower mortality across all age groups, with a nadir between 60 and 90 years of age. For period and cohort effects, trends of rate ratios for IHD mortality declined across successive period groups (refer to 2000-2004) and birth cohort groups (refer to 1985-2000) and showed similar directions in SDI quintiles, with noticeable improvements in high-SDI region (Figures 3E and 3F). In low-SDI regions, IHD mortality declined significantly for female group but fluctuated horizontally for male group (middle SDI and low-middle SDI) across successive periods. For cohort effects, sex differences (decline in female greater than male) were more remarkable in those born after 1945 in middle- and low-middle–SDI regions, and after 1970 in low-SDI region.

APC effects for IHD mortality in the top 5 countries and Brazil, according to the rank of gross domestic product obtained from the International Monetary Fund in 2019 [25], are shown in Figure 4. The United States, Japan, Germany, and Brazil had decreased trends for IHD mortality but with less age effects and declining period and cohort risks. China had significantly increased risks in people aged 75 years or older and those born before 1945. India also had high mortality among those aged 40 years or older, with little difference among period groups. The age, period, and cohort effects on IHD mortality across 21 GBD regions and 204 countries and territories are shown in Figures S6A-S6F, S7A-S7F, and S8A-S8F in Multimedia Appendix 1.
Risk Factors for Mortality From IHD

Across the top 26 risk factors (5 metabolic, 16 behavioral, and 5 environmental), with PAF, high systolic blood pressure (SBP), air pollution, and 3 other metabolic risks (high low-density lipoprotein [LDL] cholesterol, high fasting plasma glucose [FPG], and BMI) were the key contributors to IHD mortality in all SDI regions (Figure 5A and Figures S9A-S9H in Multimedia Appendix 1). In 2019, high SBP and LDL cholesterol were the 2 leading risk factors, contributing to 53.20% (95% CI 43.38%-62.76%) and 41.41% (95% CI 31.48%-52.05%) in global total deaths, respectively. From 1990 to 2019, the high FPG contribution to IHD mortality increased markedly in global and 5 SDI quintiles. During the same period, the high SBP risk increased slightly in 3 low-SDI regions but decreased in high- and high-middle–SDI regions, especially in the high-SDI region (Figure 5B). High LDL cholesterol and high BMI were prominent among young people, while high FPG and kidney dysfunction were prominent in older people (Figure 6).

Behavioral risks associated with increased IHD mortality included smoking, 13 dietary risks, low physical activity, and alcohol use. The proportion of deaths from IHD that was attributable to smoking, which was the third hazard in 3 low-SDI regions, varied substantially by sex: 29.96% (95% CI 28.74%-31.33%) in male compared with 12.23% (95% CI
11.27%-13.30%) in female (Figure 5), and was concentrated in the population 35-64 years of age (>35%; Figure 6). Dietary risks decreased with age, in which diets low in whole grains and legumes were prominent (Figure 6), and the risk of a high-sodium diet was more skewed toward male than female (Figure 5C). The risk of low physical activity increased with age (Figure 6). Interestingly, alcohol use had noticeable protective effects for IHD at the global and national levels (Figure 6 and Figures S9A-S9H in Multimedia Appendix 1).

Particulate matter pollution, an environmental risk, was the third-highest specific risk factor for IHD mortality in 3 low-SDI regions (Figure 5) and the third factor in several GBD regions (Eastern, Western, Central Sub-Saharan Africa, and South Asia; Figure S9A in Multimedia Appendix 1), but only the 12th in high-SDI regions (Figure 5), the risk of which increased with age (Figure 6). The older individuals were more susceptible to low temperatures, and lead exposure was more hazardous in low-SDI regions (Figure 6). The ranks, time trends, sex differences, and age trends of regional and national risk factors for IHD mortality are shown in Figures S9A-S9G, S10A-S10D, S11A-S11C, and S12A-S12B in Multimedia Appendix 1.
Figure 5. Top risk factors with population attributable fractions and time trends among the most specific Global Burden of Disease 2019 Study risks for mortality from ischemic heart disease (IHD). (A) Ranked contribution of the top 26 risk factors to the age-standardized death rate of ischemic heart disease by global and sociodemographic index (SDI) quintiles, 2019, for combined sex, female, and male. (B) Time trends of the top 10 risk factors to the age-standardized death rate of ischemic heart disease by global and SDI quintiles, 1990 to 2019, for combined sex, female, and male. (C) Time trends of the top 10 risk factors for mortality from ischemic heart disease with the ratio of male to female by global and SDI quintiles, 1990 to 2019. Risk factors are ranked according to colors (leading risk factor for age-standardized death=dark red; lowest risk factor for age-standardized death=dark blue). LDL: low-density lipoprotein.
Figure 6. The trends of population attributable fractions for 26 factors with age by global and sociodemographic index (SDI) quintiles, 2019, for both sexes combined. LDL: low-density lipoprotein.

Discussion

Overview

This report is the first to use the APC model to analyze temporal trends in IHD mortality on a global scale and across 204 countries and territories. Globally, the total number of IHD-related deaths increased over the past 30 years and continued to be the leading cause of NCD deaths in 2019. By contrast, ASMR decreased, with the biggest decline in mortality in high-SDI regions. Aging is an independent risk factor for mortality from IHD, with noticeable sex differences in period and cohort risks in lower-income countries. Although metabolic risks were the leading factors for mortality from IHD worldwide in 2019, smoking, particulate matter pollution, and dietary risks were also important risk factors, increasingly occurring at a younger age. Diets low in whole grains and legumes were
prominent dietary risks, regardless of sex, while smoking and a high-sodium diet were more frequent in the male than female group.

The trends for mortality from IHD varied across different regions and countries and were not completely commensurate with SDI levels. Overall, high-SDI regions and countries had higher mortality rate for IHD, but with an impressive annual downward trend. However, ASMR increased in Guam (high SDI) and Ukraine (high-middle SDI), whereas a significant decrease occurred in lower-income countries, for example, Rwanda, Ethiopia, and Afghanistan. In higher-income countries, hypertension, hyperlipidemia, diabetes, and obesity have gradually become an increasing threat, which are key risk factors accounting for a large proportion of the mortality rate [26-31]. However, newer modalities of disease prevention and a better-integrated health care system contributed to the greatest reduction in IHD mortality in higher-income countries. For lower-income countries, the Westernized lifestyle with increased metabolic risks and severe air pollution is causing IHD to emerge as a growing health hazard [32-36]. A previous study showed that CVD mortality was much higher in patients with diabetes in lower- than higher-income countries. This increased risk was not altered even when risk factors such as physical inactivity, smoking, BMI, and hypertension and treatments are taken into account [37] but may be related to the fact that lower-income countries are severely underresourced in the number of health care providers, technological infrastructure, and implementation of treatment [38-40]. Therefore, countries with different degrees of SDI, especially lower-income countries with more complex and challenging situations, need to formulate appropriate prevention and intervention strategies based on local health-related resources.

The marked upsurge of IHD mortality in people aged 65 years or older shows that aging is an independent risk factor for mortality from IHD [41,42]. Aging is a nonmodifiable risk factor for IHD, with cellular senescence and vascular aging accounting for atherosclerosis, which is the primary cause of IHD [43,44]. The older individual is susceptible to immune dysfunction and chronic metabolic diseases, which increase the risk of IHD mortality [45]. We found that risk due to low physical activity increased with age, which was associated with CVD mortality [46]. Furthermore, the aging of the population caused by low birth rate and longer life expectancy also indirectly leads to an increase in IHD mortality [14]. In China, IHD mortality, controlled for period and cohort effects, increased sharply after the age of 70 years, where 65.6% of the health burden were projected to occur in older adults in 2050 [47]. The age at which there is a rapid rise in mortality is also delayed in other SDI countries where death trends are increasing, such as Ukraine, the Philippines, Lesotho, and Mozambique, among others. These data suggest that prevention and treatment strategies and multidisciplinary management for IHD in the older population should be a major focus of health systems.

The period and cohort effects on IHD mortality have shown an overall downward trend, which may be attributable to the implementation of better public health policies and improvements in health care [48]. Social stability, improvement in living standards, and advances in medical technologies (medical green channels, eg, Chest Pain Center, advanced interventional techniques, medical devices, and therapeutic drugs) have contributed to the improvement in survival rates of patients with IHD [49,50]. However, unfavorable period and cohort effects continue in some countries, especially in those with lower income. From 1990 to 2019, the high SBP contribution to IHD mortality increased slightly in low-SDI regions, but the risk related to high FPG increased across 5 SDI quintiles. Hazards for IHD mortality caused by high LDL cholesterol, particulate matter pollution, high BMI, and 13 dietary risks are mainly concentrated in the younger age group. The United States (high SDI) is one example of a country that has begun to control air pollution, reducing by more than 80% the atmospheric particulate matter concentration in 2017; the Clean Air Act was passed in 1970 and amended in 1977 and 1990 [51]. China (middle SDI) and India (low-middle SDI) also showed a decline in household air pollution–attributable IHD mortality with a series of measures to decrease air pollution [52-53]. One recent study showed that greater adherence to healthy eating patterns, as recommended by the Dietary Guidelines for Americans, was consistently associated with lower risk of total and IHD-caused mortality [54]. Okinawans, 65 years of age or older, have a lower risk for metabolic diseases and mortality, which may be related to the traditional Okinawan diet consisting of 85% calories from carbohydrates and 9% from protein [55]. The above reports show that current resources are largely inadequate to manage patients with IHD in many countries, and health care strategies from countries with significant improvements in IHD mortality may serve as a model for countries at the same level.

The association between alcohol use and IHD remains an open question. On the one hand, alcohol consumption has been proven to be a leading risk factor for disease burden, but on the other hand, some researchers have suggested that low levels of alcohol consumption can have a beneficial effect on IHD outcomes [56]. Our analysis showed that alcohol use had noticeable protective effects for IHD, but the effect was small compared with the other attributable burdens in those locations. Although the results of our analysis are consistent with previous analyses [56,57], it does not prove that higher alcohol consumption is beneficial for IHD. When considered with overall health risks, alcohol use was strongly associated with cancer, injuries, and communicable disease. Factors such as sex, ethnicity, alcohol consumption, and frequency also need to be considered in future analyses [58].

Sex differences in IHD mortality trends have gradually emerged, especially in lower-income countries [39,60]. In this study, we found that risks from smoking, a diet high in sodium, and particulate matter pollution were greater in male than female individuals globally. By contrast, risks from kidney dysfunction were greater in female than male individuals in high-SDI regions, and risks from high BMI were higher in female than male individuals in low-SDI regions. The traditional risk factors linked to sex-related determinants may be related to shifting roles and relations under the double burden of employment and caregiving responsibilities, which exacerbate the burden of IHD, especially in developing countries [61-64]. However, high-SDI regions had the lowest ratio of male-to-female IHD mortality.
attributable to smoking, suggesting that smoking control programs and policies are becoming successful in high-income countries [65,66]. These indicate that country-specific strategies to incorporate sex determinants into health research are essential to the improvement in IHD mortality.

There are several limitations to this study. First, the gathering and quality of data sources are different across countries in GBD. Limitations in some low- and middle-income countries are unavoidable, such as unavailable or incomplete data, delayed and inaccurate reporting, and misclassified coding [1]. Second, although the time trends analysis was performed across about 30 years, this research was based on cross-sectional data from GBD, which was not a cohort study. Third, mortality data with different trends at the global and national level were analyzed, but subnational differences were not calculated. Moreover, the national data on health system quality, including medication use, interventional operation implementation, and access to medical care were not included in the data source in this study.

In addition, SDI as a composite indicator of income per capita is time-varying, but this study only used SDI of 2019 for the analysis of data from 1990 to 2019 due to data acquisition limitations. Finally, the mortality data for IHD presented in this report were collected before the COVID-19 pandemic. Studies have reported that COVID-19 can induce CVD and increase the risk of death in those with preexisting CVD [67,68], which may be a key focus for IHD mortality analysis in the future.

Conclusions
Although there have been reductions in IHD mortality over the past 30 years, especially in high-SDI regions and countries, IHD remains the leading cause of global mortality from NCDs. Aging is highly correlated with IHD-related death. Sex-related differences in period and cohort risks occurred across higher- and lower-income countries. Metabolic risks were the leading factors for IHD mortality worldwide in 2019, followed by smoking, particulate matter pollution, and dietary risks.

Acknowledgments
We acknowledge all members of the Institute for Health Metrics and Evaluation, the University of Washington, and all study groups involved in the Global Burden of Disease 2019 Study. This work was supported in part by the Key Science and Technology Program of the Chongqing Special Project of Social Undertaking and People’s Livelihood Guarantee (cstc2017shms-zdyfX0046), the Chongqing Technology Innovation and Application Demonstration Project (cstc2018scx-mszdX0024), and the Chongqing Science and Technology Research and Development Platform (cstc2015ypt-gcjysjzx0166) from the Chongqing Science and Technology Commission.

Data Availability
The original data sources for this study are publicly available from the Global Burden of Disease 2019 Study data [16].

Authors’ Contributions
CZ conceptualized and designed the study. Q-WZ contributed to the interpretation of the data, statistical analyses, and supporting figures. TS drafted and revised the manuscript and prepared the figures. YH, PAJ, and HW were involved in the rigorous revision of the manuscript. MT, BH, XL, and CL participated in the discussion and review of the manuscript. CZ and Q-WZ are cocorresponding authors. Q-WZ can be contacted at zhangqingweif@hotmail.com. All the authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental Figures. [DOCX File, 33335 KB - publichealth_v10i1e46821_app1.docx ]

Multimedia Appendix 2
Supplemental Tables. [DOCX File, 200 KB - publichealth_v10i1e46821_app2.docx ]

References


16. GBD results. GHDx. URL: https://vizhub.healthdata.org/gbd-results/ [accessed 2023-12-19]


25. IMF. URL: https://www.imf.org/zh/home [accessed 2023-12-19]


38. Shu et al. JMIR Public Health and Surveillance 2024 | vol. 10 | e46821 | p.1350 (page number not for citation purposes)


Abbreviations

AAMR: all-age mortality rate
APC: age-period-cohort
ASMR: age-standardized mortality rate
CVD: cardiovascular disease
FPG: fasting plasma glucose
GBD: Global Burden of Disease 2019 Study
IHD: ischemic heart disease
LDL: low-density lipoprotein
NCD: noncommunicable disease
PAF: population attributable fraction
SBP: systolic blood pressure
SDI: sociodemographic index
UI: uncertainty interval

Edited by A Mavragani; submitted 27.02.23; peer-reviewed by P Wang, W Wang; comments to author 22.08.23; revised version received 05.10.23; accepted 22.11.23; published 24.01.24.

Please cite as:
URL: https://publichealth.jmir.org/2024/1/e46821
doi:10.2196/46821
PMID:38265846

©Tingting Shu, Ming Tang, Bo He, Xiaozhu Liu, Yu Han, Chang Liu, Pedro A Jose, Hongyong Wang, Qing-Wei Zhang, Chunyu Zeng. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 24.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Depression and Anxiety Symptoms Among Cisgender Gay and Bisexual Men During the Onset of the COVID-19 Pandemic: Time Series Analysis of a US National Cohort Study

Chloe Mirzayi, PhD; Drew Westmoreland, PhD; Matthew Stief, PhD; Christian Grov, PhD
CUNY Graduate School of Public Health and Health Policy, New York, NY, United States

Corresponding Author:
Christian Grov, PhD
CUNY Graduate School of Public Health and Health Policy
55 W 125 Street
New York, NY, 10027
United States
Phone: 1 6463640254
Email: cgrov@sph.cuny.edu

Abstract

Background: The onset of the COVID-19 pandemic in the United States in March 2020 caused a dramatic change in the way many people lived. Few aspects of daily life were left undisturbed by the pandemic’s onset as well as the accompanying policies to control the spread of the disease. Previous research has found that the pandemic may have significantly impacted the mental health of lesbian, gay, bisexual, transgender, and queer (LGBTQ) individuals—potentially more so than other individuals. However, the pandemic did not affect all areas of the United States at the same time, and there may be regional variation in the impact of the onset of the pandemic on depressive symptoms among LGBTQ individuals.

Objective: To assess regional variation of the impact of the pandemic, we conducted a time series analysis stratified by US geographic region to examine symptoms of depression and anxiety among a sample of primarily cisgender gay and bisexual men before and after the onset of the COVID-19 pandemic in the United States.

Methods: In total, 5007 participants completed assessments as part of the Together 5000 study, an ongoing prospective cohort study. Depressive and anxiety symptoms were measured using the Patient Health Questionnaire-4. Patient Health Questionnaire-4 scores were graphed as a function of days from March 15, 2020. Locally estimated scatterplot smoothing trend lines were applied. A sieve-bootstrap Mann-Kendall test for monotonic trend was conducted to assess the presence and direction of trends in the scatterplots. We then compared the observed trends to those observed for 1 year prior (2018-2019) to the pandemic onset using data collected from the same sample.

Results: Significant positive trends were detected for the Northeast (P=.03) and Midwest (P=.01) regions of the United States in the 2020 assessment, indicating that symptoms of anxiety and depression were increasing in the sample in these regions immediately prior to and during the onset of the pandemic. In contrast, these trends were not present in data from the 2018 to 2019 assessment window.

Conclusions: Symptoms of anxiety and depression increased among the study population in the Northeast and Midwest during the beginning months of the COVID-19 pandemic, but similar increase was not observed in the South and West regions. These trends were also not found for any region in the 2018 to 2019 assessment window. This may indicate region-specific trends in anxiety and depression, potentially driven by the burden of the pandemic and policies that varied from region to region. Future studies should consider geographic variation in COVID-19 spread and policies as well as explore potential mechanisms by which this could influence the mental health of LGBTQ individuals.

(JMIR Public Health Surveill 2024;10:e47048) doi:10.2196/47048

KEYWORDS
COVID-19 pandemic; lesbian, gay, bisexual, transgender, and queer; LGBTQ; mental health; depression; anxiety; time series; gay and bisexual men; cisgender gay; pandemic; gay
Introduction

The onset of the COVID-19 pandemic in the United States in early 2020 represented a nearly unprecedented disruption in daily life for many Americans. Widespread disruptions were present in nearly every facet of daily life, including in workplaces, schools, and social settings. Mandatory business closures and recommended social distancing were implemented with some heterogeneity in the United States. Many states in the Northeast, Midwest, and West implemented stay-at-home orders in March 2020 prior to Southern states, where the pandemic later took hold in the summer [1].

Several large studies have linked the onset of the pandemic to a significant increase in the symptoms of depression and anxiety in US adults [2-5]. Other studies have found that lesbian, gay, bisexual, transgender, and queer (LGBTQ) people may be particularly vulnerable to developing symptoms of depression and anxiety relative to non-LGBTQ populations [6]. For instance, a study of cannabis users conducted during the pandemic found that the LGBTQ cannabis users were more likely to report symptoms of depression and anxiety [7]. Similarly, a national longitudinal cohort study of LGB adults found increased anxiety and depression symptoms among participants after the beginning of the pandemic [8]. Additionally, a study of LGBTQ college students found that nearly half had unmet mental health needs during the first few months of the pandemic [6]. These findings may be due to being confined in uncomfortable home situations or isolation from friends, which may have disparate impacts on LGBTQ individuals compared to others [9]. However, other studies have found decreased anxiety symptoms and no differences in depressive symptoms before and after the start of the pandemic [10].

Recently, researchers have proposed potential mechanisms by which the onset of the COVID-19 pandemic may have disproportionately affected LGBTQ people. A web-based study in Germany during the pandemic found that depressive symptoms were mediated by loneliness [11], while a study of Mexican LGBTQ people found that perceived social support also acted as a mediator [12]. Taken together, it appears that a lack of access to supportive friends and other members of the LGBTQ community, and the accompanying loneliness, may be one mechanism through which the COVID-19 pandemic, and associated lockdowns, had a greater impact on LGBTQ individuals compared to others.

Despite several studies finding that LGBTQ people were more likely to experience symptoms of depression and anxiety with the onset of the COVID-19 pandemic, we have been unable to identify any studies that examined trends in symptoms and anxiety over time for the period. Time series studies have been used to examine trends in mental health leading up to and immediately following the 2016 US presidential election [13], in mental health referrals during the pandemic [14,15], and in suicide mortality also during the COVID-19 pandemic [16]. However, to our knowledge, no study has examined symptoms of depression and anxiety in an LGBTQ population during the onset of the COVID-19 pandemic in the United States using a time series approach.

The aim of this study was to examine the effects of the onset of the first wave of the COVID-19 pandemic on a population of primarily gay and bisexual men, but also transgender men and transgender women as well as nonbinary people, using time series analysis. Given the heterogeneity of the burden of cases and in COVID-19 policies, analyses were stratified by US census region. We hypothesized that regions that were more impacted by the early onset of the COVID-19 pandemic or had more stringent pandemic control policies or recommendations (eg, the Northeast) would show an increase in symptoms of depression and anxiety as the pandemic progressed.

Methods

Study Population

Participants were members of Together 5000, a US national cohort of primarily cisgender gay and bisexual men at risk for HIV recruited from sexual networking applications. Members of the cohort complete yearly web-based surveys as well as at-home HIV testing beginning in 2017 and running through the onset of the pandemic. The yearly web-based surveys assessed a variety of health-related domains such as socioeconomic factors, drug and alcohol use, sexual behaviors, food and housing insecurity, and mental health measures. A detailed profile of the study procedures and participants has been published elsewhere [17].

Measures

Data Collection

Data for this study were primarily taken from participants’ 24-month assessment. These assessments began on November 20, 2019 (~116 days), and ran through August 15, 2020 (~153 days)—with some participants completing their assessment prior to the onset of the COVID-19 pandemic in the United States and others completing it after. We defined the onset as March 15, 2020, which roughly coincides with the beginning of the week in which many local and state governments announced mandatory closure of nonessential businesses and schools [18-20]. To establish whether a similar trend would be observed in prior yearly assessments, we compared 24-month results to participants’ 12-month assessment (all of which were conducted prior to the pandemic). Survey completion dates were calculated relative to March 15, 2019, and ranged from November 7, 2018 (~127 days), to June 24, 2019 (~101 days).

Demographics

On the baseline survey, race and ethnicity information was assessed using a list of common racial and ethnic categories for participants to select all that apply. Participants were also given the ability to specify their own race or ethnicity. These options were recoded to a 4-category variable representing Black, Hispanic, White, and multiracial or others. To assess current gender identity, participants were asked their assigned sex at birth as well as their current gender identity. Participants had the option of writing in their own gender identity. The variable
was then recoded into 5 categories: cisgender man, cisgender woman, transgender man, transgender woman, or others or nonbinary. Finally, participants were asked to select the best option for their sexual orientation: gay, queer, or homosexual; bisexual; straight or heterosexual; or others. This variable was not recoded prior to analysis.

Region
Participants self-reported their zip codes at baseline and were able to provide updated zip codes during the 12- and 24-month assessments. These zip codes were coded into state of residence, which in turn were categorized according to US census regions [21].

Patient Health Questionnaire-4
The 4-item Likert-type Patient Health Questionnaire-4 (PHQ-4) was used to assess symptoms of anxiety and depression. The Patient Health Questionnaire-2, a measure of depression, and Generalized Anxiety Disorder-2, a measure of anxiety, constitute the PHQ-4. The reliability and validity of the PHQ-4 in the general population have been previously established [22]. The PHQ-4 is scored from 0 to 12, with an increasing score indicating increasingly severe anxiety and depressive symptomology. The continuous measure was used for analyses.

Statistical Analysis
PHQ-4 scores were stratified by US census region of residence and then plotted as a function of the survey completion date for participants completing the 24-month assessment, which ran from November 20, 2019, to August 15, 2020. A locally estimated scatterplot smoothing trend line with 95% CI was applied. A sieve-bootstrap Mann-Kendall test for monotonic trends was conducted to assess the presence and direction of trends in the scores. All steps of this analysis were then repeated for the 12-month assessment for comparison.

Analyses were conducted in R (version 4.1.0; R Foundation for Statistical Computing). Plots and trend lines were created using the ggplot2 package [23], while the test for trend was conducted using the funtimes package [24].

Ethical Considerations
All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional review board (Institutional Review Board of the City University of New York Graduate School of Public Health and Health Policy, IRB file #2017-0893) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All participants provided informed consent for data collection at each assessment. Participant data were deidentified and stored in an encrypted closed network. Participants received US $25 Amazon gift cards for completing study assessments.

Results
In total, 5007 participants completed the 24-month assessment survey and resided in 1 of the 4 US census regions. An additional 22 participants completed the survey but resided in a US territory and were not included in these analyses. Of these 5007 participants, 784 (15.6%) resided in the Northeast, 780 (15.6%) resided in the Midwest, 2285 (45.6%) resided in the South, and 1158 (23.1%) resided in the West.

Demographically, 507 (10.1%) participants were Black, 1213 (24.2%) were Hispanic, 2677 (53.5%) were White, and 610 (12.2%) were another race or multiracial. The mean age of the sample at the 24-month assessment was 32.9 (SD 7.5) years. The sample was primarily cisgender men with 4885 (97.6%) identifying as such, with the remaining 122 (2.4%) participants identifying as transgender men, transgender women, or nonbinary. A substantial proportion of the sample identified as gay, queer, or homosexual (n=4291, 85.7%) or bisexual (n=660, 13.2%), with the remainder identifying as others. Full demographics as well as PHQ-4 scores stratified by demographic variables are presented in Table 1.

The mean PHQ-4 score at the 24-month assessment was 4.0 (SD 3.0), while the mean PHQ-4 score at the 12-month assessment was 3.7 (SD 2.4). A paired t-test revealed this difference to be statistically significant (t_{4653}=2.99; P=.019), indicating an overall increase in PHQ-4 scores between the 12-month assessment and the 24-month assessment.

Survey completion dates for the 24-month assessment were calculated relative to the onset of the COVID-19 pandemic, with 0 representing March 15, 2020. These dates ranged from November 20, 2019, to August 15, 2020. Individual participant Generalized Anxiety Disorder–Patient Health Questionnaire scores were then plotted as a function of survey completion dates stratified by US census region.

The locally estimated scatterplot smoothing trend line in PHQ-4 scores for the Northeast revealed a trend downward in the days immediately prior to the onset of the pandemic with an upturn in scores following the onset (Figure 1). This trend was significant (τ=0.12; P=.03). A small, steady increase in scores was also observed among participants in the Midwest, beginning prior to the pandemic onset and continuing through to the summer (τ=0.14; P=.01). No trends are immediately apparent in the trend lines for the South and West, and tests for trends were not statistically significant (Table 2).

To establish whether a similar trend would be observed in other yearly assessments, the same analysis was conducted for the 12-month assessment (Figure 2). Inspection of the figure reveals a decrease in scores for participants in the Northeast after March 15, 2019, followed by a sharp upturn at the end. Midwest participant scores decreased markedly shortly after March 15, 2019. A small, nonsignificant oscillation in scores was observed among Western participants. No clear trend was observed in scores among Southern participants. All tests for monotonic trends were nonsignificant for 2019 (all P>.05; Table 2), indicating, compared to 2020, there was no observed trend in PHQ-4 scores over time.
Table 1. Demographics of a prospective US national cohort study of primarily cisgender gay and bisexual men from 2019 to 2020 stratified by US census region with PHQ-4<sup>a</sup> (anxiety and depressive symptoms) scores.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Northeast (n=784)</th>
<th>Midwest (n=780)</th>
<th>South (n=2285)</th>
<th>West (n=1158)</th>
<th>Total (N=5007)</th>
<th>PHQ-4 score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race or ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>77 (9.8)</td>
<td>55 (7.1)</td>
<td>330 (14.4)</td>
<td>45 (3.9)</td>
<td>507 (10.1)</td>
<td>2.8 (3.3)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>158 (20.2)</td>
<td>99 (12.7)</td>
<td>580 (25.4)</td>
<td>376 (32.5)</td>
<td>1213 (24.2)</td>
<td>3.1 (3.4)</td>
</tr>
<tr>
<td>White</td>
<td>419 (53.4)</td>
<td>544 (69.7)</td>
<td>1146 (50.2)</td>
<td>568 (49.1)</td>
<td>2677 (53.5)</td>
<td>3.7 (3.5)</td>
</tr>
<tr>
<td>Others or multiracial</td>
<td>130 (16.6)</td>
<td>82 (10.5)</td>
<td>229 (10)</td>
<td>169 (14.6)</td>
<td>610 (12.2)</td>
<td>3.5 (3.6)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisgender men</td>
<td>758 (96.7)</td>
<td>762 (97.7)</td>
<td>2234 (97.8)</td>
<td>1131 (97.7)</td>
<td>4885 (97.6)</td>
<td>3.4 (3.5)</td>
</tr>
<tr>
<td>Transgender men</td>
<td>9 (1.1)</td>
<td>5 (0.6)</td>
<td>18 (0.8)</td>
<td>7 (0.6)</td>
<td>39 (0.8)</td>
<td>5.7 (4.1)</td>
</tr>
<tr>
<td>Transgender women</td>
<td>1 (0.1)</td>
<td>3 (0.4)</td>
<td>10 (0.4)</td>
<td>7 (0.6)</td>
<td>21 (0.4)</td>
<td>1.8 (2.0)</td>
</tr>
<tr>
<td>Others or nonbinary</td>
<td>16 (2)</td>
<td>10 (1.3)</td>
<td>23 (1)</td>
<td>13 (1.1)</td>
<td>62 (1.2)</td>
<td>5.4 (4.1)</td>
</tr>
<tr>
<td>Sexual orientation, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gay, queer, or homosexual</td>
<td>607 (88.9)</td>
<td>678 (86.9)</td>
<td>1926 (84.3)</td>
<td>990 (85.5)</td>
<td>4291 (85.7)</td>
<td>3.5 (3.5)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>81 (10.3)</td>
<td>96 (12.3)</td>
<td>328 (14.4)</td>
<td>155 (13.4)</td>
<td>660 (13.2)</td>
<td>3.3 (3.5)</td>
</tr>
<tr>
<td>Straight or heterosexual</td>
<td>0 (0)</td>
<td>2 (0.3)</td>
<td>7 (0.3)</td>
<td>4 (0.3)</td>
<td>13 (0.3)</td>
<td>3.5 (4.1)</td>
</tr>
<tr>
<td>Others</td>
<td>6 (0.8)</td>
<td>4 (0.5)</td>
<td>24 (1.1)</td>
<td>9 (0.8)</td>
<td>43 (0.9)</td>
<td>4.2 (3.9)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>32.6 (7.5)</td>
<td>33.5 (8.0)</td>
<td>32.5 (8.1)</td>
<td>33.5 (7.1)</td>
<td>32.9 (7.9)</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>PHQ-4: Patient Health Questionnaire-4.
<sup>b</sup>N/A: not applicable.

Figure 1. US census region–stratified trends in PHQ-4 scores over time before and after the onset of the COVID-19 pandemic of a prospective US national cohort study of primarily cisgender gay and bisexual men in 2020. PHQ-4: Patient Health Questionnaire-4.
Table 2. The sieve-bootstrap Mann-Kendall test for monotonic trend of Patient Health Questionnaire-4 (depression and anxiety symptoms) of a prospective US national cohort study of primarily cisgender gay and bisexual men from 2019 to 2020. Tests were conducted for each regional time series for each year. Positive $\tau$ indicates an increasing monotonic trend.

<table>
<thead>
<tr>
<th>Region</th>
<th>2019 $\tau$</th>
<th>2019 P value</th>
<th>2020 $\tau$</th>
<th>2020 P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>0.05</td>
<td>.44</td>
<td>0.12</td>
<td>.03</td>
</tr>
<tr>
<td>Midwest</td>
<td>0.00</td>
<td>.96</td>
<td>0.14</td>
<td>.01</td>
</tr>
<tr>
<td>South</td>
<td>0.01</td>
<td>.83</td>
<td>0.00</td>
<td>.96</td>
</tr>
<tr>
<td>West</td>
<td>0.07</td>
<td>.26</td>
<td>0.03</td>
<td>.50</td>
</tr>
</tbody>
</table>

Figure 2. US census region–stratified trends in PHQ-4 scores over time before and 1 year prior to the onset of the COVID-19 pandemic of a prospective US national cohort study of primarily cisgender gay and bisexual men in 2019. PHQ-4: Patient Health Questionnaire-4.

Discussion

Principal Findings

The region-stratified time series analysis of symptoms of depression and anxiety in the days leading up to and following the onset of the COVID-19 pandemic in the United States revealed significant positive trends among participants in the Northeast and Midwest. These trends were not observed in the South or West, nor were they present the previous year.

The disparate impact of the COVID-19 pandemic on the Northeast [25] as well as heterogeneous COVID-19 policies across the United States may explain the observed trends. Other studies have found that lockdown policies, along with adherence to them, may have resulted in increased depressive symptoms [26].

Comparison to Prior Work

The findings of significant positive trends in symptoms of depression and anxiety during the onset of the COVID-19 pandemic were consistent with similar studies that found an increased prevalence of depression and anxiety among the general population as well as LGBTQ people during the COVID-19 pandemic [2,6]. However, to our knowledge, this is the first time-series analysis of symptoms of anxiety and depression conducted in an LGBTQ population with data before and after the onset of the COVID-19 pandemic in the United States as well as the first study to identify a regional trend in this population.

Other time series studies have shown the impact of the COVID-19 pandemic on related mental health constructs. A UK study found a significant decrease in referrals to mental health services during the start of the pandemic, although this study was not focused specifically on sexual minorities [14]. Another study of the general population found a decrease in emergency department visits for psychiatric issues [27]. The decrease in mental health service use combined with increased symptoms could indicate a clear, unmet need for mental health care during the beginning of the COVID-19 pandemic. It is important to consider that the pandemic may have had differential effects on sexual minority individuals based on stigma and vulnerabilities. For instance, young adults who had to return to their parents’ homes as a result of the pandemic were found to have faced greater psychological distress [28]. Future studies could explore whether the observed regional effects were modified by changes in living arrangement. Another potential effect modifier could be social support that other studies have found as an important protective effect for sexual minority populations against depressive symptoms [29].
Limitations

There are some limitations to this study. Survey completion dates were used in calculating days relative to March 15, 2020, and March 15, 2019. It is possible for participants to have started the survey, completed the PHQ-4, and then not completed the survey in its entirety until several days later; however, we do not believe this would have tremendously impacted responses in a meaningful way nor were there enough participants who started before March 15, 2020, and completed it after March 15, 2020, to have had a discernible impact on aggregate results. There may also be other important factors that affect the mental health of sexual minority populations that were not included in this study such as internalized homonegativity [30], social connectedness [31], and minority stress [32]. Finally, the study population was primarily cisgender gay and bisexual men and therefore does not represent the full spectrum of LGBTQ individuals.

Conclusions

We found evidence of increased depression and anxiety symptoms among a sample of primary cisgender gay and bisexual men during the initial onset of the COVID-19 pandemic in 2020 in the Midwest and Northeast regions of the United States. This trend was not similar among the same study population the previous year nor among participants in the South or West regions of the United States.

Acknowledgments

The authors thank the additional members of the Together 5000 study team: David Pantalone, Sarit A Golub, Viraj V Patel, Gregorio Millett, Don Hoover, Pedro Carneiro, Javier Lopez-Rios, Alexa D’Angelo, Fatima Zohra, Corey Morrison, and Michelle Dearolf. The authors also thank the program staff at the National Institutes for Health (NIH): Gerald Sharp, Sonia Lee, and Michael Stirrett and the members of our scientific advisory board: Michael Camacho, Demetre Daskalakis, Sabina Hirshfield, Jeremiah Johnson, Claude Mellins, and Milo Santos. While the NIH financially supported this research, the content is the responsibility of the authors and does not necessarily reflect the official views of the NIH. Together 5000 was funded by the NIH (UH3 AI 133675; principal investigator: CG). Other forms of support include the City University of New York Institute for Implementation Science in Population Health, the Einstein–Rockefeller–City University of New York Center for AIDS Research (P30 AI124414).

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

CM conceived the study, developed the methodology, conducted statistical analyses, created tables and figures, and wrote and edited the paper. DW conceived the study, advised on the methodology, and aided in editing the paper. MS administered the Together 5000 study and aided in the conceptualization of the study as well as editing the paper. CG supervised the Together 5000 study, acquired funding for the study, aided in study conceptualization, and edited the paper. All authors reviewed the final paper.

Conflicts of Interest

None declared.

References


Abbreviations

LGBTQ: lesbian, gay, bisexual, transgender, and queer
PHQ-4: Patient Health Questionnaire-4
Global Trends in the Incidence of Anxiety Disorders From 1990 to 2019: Joinpoint and Age-Period-Cohort Analysis Study

Huiru Cao1*, MD; Yang Wu2*, MD; Hui Yin3, MD; Yanqi Sun4, MD; Hui Yuan5, PhD; Mengjun Tao2, MD

1Department of Gastroenterology, The First Affiliated Hospital of Wannan Medical College, Wuhu, China
2Department of Health Management Center, The First Affiliated Hospital of Wannan Medical College, Wuhu, China
3Department of Hospital Infection, The First Affiliated Hospital of Wannan Medical College, Wuhu, China
4Department of Prevention and Health Care, People's Hospital of Rizhao, Rizhao, China
5School of Public Health, Wannan Medical College, Wuhu, China

*these authors contributed equally

Corresponding Author:
Mengjun Tao, MD
Department of Health Management Center
The First Affiliated Hospital of Wannan Medical College
No. 2, Zheshan West Road
Wuhu, 241000
China
Phone: 86 17398374118
Email: taomengjun@yjsyy.com

Abstract

Background: Anxiety disorders (ADs) are the most common mental illness with high prevalence, chronicity, and comorbidity. Despite rapid economic and cultural development, the global incidence of ADs continues to increase, with predominance in male individuals.

Objective: To address the above issues, we analyzed the dynamic trends of the global incidence and disease burden of ADs from 1990 to 2019 and their different effects on age, period, and birth cohort and predicted the future trend of AD incidence.

Methods: The data were obtained from the Global Burden of Disease study in 2019. A joinpoint regression model was used to calculate the annual percent change in AD incidence, and age-period-cohort analysis was used to estimate the independent effects of age, period, and cohort. Nordpred age-period-cohort analysis was used to predict the incidence of ADs from 2020 to 2044.

Results: The age-standardized incidence rate of ADs increased by 1.06% for both sexes, and the age-standardized disability-adjusted life-year (DALY) rate (ASDR) decreased by 0.12%. Joinpoint regression indicated that increments in average annual percent changes in the age-standardized incidence rate (0.068 vs 0.012) and ASDR (0.035 vs –0.015) for ADs globally were higher among male individuals than female individuals. The age-period-cohort analyses revealed that the relative risk (RR) of the incidence and DALYs of ADs among people of different sexes increased with age in adolescence and middle age and then decreased. For the period effect, the RR of incidence decreased, whereas the RR of DALYs increased in both sexes. Moreover, the RR of the incidence gradually increased and DALYs slowly decreased with birth year for both male and female individuals. New cases of ADs in male individuals are predicted to increase in the coming 25 years.

Conclusions: This study provided the changing trend of the global incidence and disease burden of ADs in the past 3 decades, indicating that early prevention and effective control cannot be ignored. We analyzed the age-period-cohort effect of potential trends in ADs and predicted future incidence trends. The results suggest that we should take active intervention measures, focusing on high-risk groups and developing effective management and control policies to reduce the global burden of disease.

(JMIR Public Health Surveill 2024;10:e49609) doi:10.2196/49609

KEYWORDS
age-period-cohort analysis; anxiety disorders; incidence; disability-adjusted life-years; DALYs; joinpoint regression model; prediction
Introduction

Anxiety disorders (ADs) form the most common group of mental disorders, which are characterized by excessive and enduring fear, anxiety, or avoidance of perceived threats and can also include panic attacks [1], with generalized ADs being the most common AD seen in primary care [2]. This is based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and the International Classification of Diseases, 11th Edition (ICD-11) diagnostic systems, which categorize ADs based on key symptoms at 3 levels such as cognitive, emotional, and somatic, including separation anxiety, selective mutism, specific phobias, social AD, panic disorder, agoraphobia, and generalized AD [1,3]. The lifetime prevalence rates of AD have reached 34% [4], and ADs generally start in childhood or adolescence [5,6]. Most patients with ADs often have comorbid ADs and other mental disorders, especially depression [7] and somatic disorders such as heart disease, hyperthyroidism, asthma, and epilepsy [8-10]. Importantly, ADs are 2 times more prevalent among female individuals than male individuals [11]. The high prevalence, chronicity, and comorbidity of ADs led the World Health Organization to rank ADs as the ninth leading health-related cause of disability [12]. ADs not only seriously affect patients’ daily functioning and quality of life but also impose a heavy burden on society and account for 3.3% of the global burden of disease (GBD) [13,14].

According to the large-scale World Mental Health Surveys, the prevalence of ADs was highest in high-income countries, especially in Australia, the United States, and European countries [15,16]. Estimated from the 2010 cost model, over €74 billion (US $81.64 billion) was largely due to indirect costs such as disability for 30 European countries [13]. Even if there is insufficient evidence to show that the prevalence of ADs is increasing, the burden of disease cannot be ignored. ADs are still an important global public health problem.

At present, there are relatively few studies on the global trend in the prevalence of ADs, and they use only traditional descriptive analysis of age-specific incidence or mortality data at different times, which cannot eliminate or control for the interaction among age, period, and cohort factors. The joinpoint regression (JPR) model is mainly used to analyze the time trend of incidence and mortality and to analyze the burden of disease, which can better reflect the change in the epidemic trend and its impact [17,18]. The age-period-cohort model improved the traditional descriptive analysis method to estimate the risk of disease incidence or mortality and its trend while adjusting for age, period, and cohort [18,19].

Therefore, this study analyzed data from the GBD 2019 study. The JPR model was used to explore the temporal trends in the incidence and disability-adjusted life-year (DALY) rates of ADs globally and to explore the net age, period, and cohort effects with the age-period-cohort model. Meanwhile, the prediction of the incidence of ADs in the next 25 years based on these data can provide a scientific basis for the evaluation of AD prevention and treatment to reduce the disease burden caused by ADs.

Methods

Data Sources

Data on AD trends during 1990-2019 were retrieved from the World Health Organization GBD estimates. GBD data are derived from private and public organizations worldwide that collect data through surveys, reports, scientific literature, censuses, and other methods. The study has provided improved standardized methods and a comprehensive assessment of incidence and DALYs for 369 diseases and injuries and 87 risk factors in 204 countries and territories. Further details of the general methodologies for the GBD have been described elsewhere [20-22], and data and the protocol for the 2019 GBD can be accessed through the Global Health Data Exchange GBD Results Tool. All GBD 2019 analyses adhered to the Guidelines for Accurate and Transparent Health Estimates Reporting [23].

Ethical Considerations

This study used data from the GBD Study 2019, which was approved by the institutional review board of the University of Washington. Original data were collected with informed consent from study participants or with a waiver from the institutional review board. As this was a secondary analysis of publicly available data, no further review by an institutional review board was required following the data use agreement of The Institute for Health Metrics and Evaluation.

Statistical Analysis

Overall Temporal Trends Analysis of Global

Data on the global and regional incidence of ADs from 1990 to 2019 were extracted from the GHDx database, including crude incidence (CIR), DALYs, age-standardized incidence, and age-standardized DALYs of ADs in the all-age group. Data are reported as estimates with 95% uncertainty intervals (UIs). Age-standardized incidence rates (ASIRs) and age-standardized DALY rates (ASDRs) were used to assess the temporal trend of the global incidence of ADs from 1990 to 2019. The ASIR and ASDR in the GBD database were standardized based on 2019 global population data [20].

JPR Analysis

JPR model analysis was applied to estimate the trends in AD prevalence from 1990 to 2019. As proposed by Kim et al [24], joinpoint analysis can automatically divide longitudinal variations into different segments by segmented regression and identify the segment trends with statistical significance. Regression fitting was performed on the natural logarithm of the prevalence and mortality rate in different segments, and then the annual percent changes (APCs) and their 95% CIs were calculated for each period. The global trend was described by the average annual percent changes (AAPCs). The APCs and AAPCs were considered statistically significant by nonoverlapping 95% CIs and P<.05 compared to the null hypothesis of no variation. We applied the JPR program (version 4.8.0.1) from the Statistical Research and Applications Branch of the Surveillance Research Program of the US National Cancer Institute.
Age-Period-Cohort Analysis

The age-period-cohort model was used to estimate the relative risk (RR) of the population in a given year and the cumulative health risk since birth. The model allowed analysis of the independent effects of age, period, and cohort on temporal trends in incidence and DALY rate due to ADs and other forms of ADs. The age-period-cohort model provides a useful parametric framework that can complement standard nonparametric descriptive methods. This analysis model has been used and recognized in many previous studies [18,19]. In this model, the data collected were stratified into age groups of 5 consecutive years and periods of 5 consecutive years. The incidence rate and DALY rate of AD in 5 consecutive age groups (0-5 years), 5 consecutive periods (1990-1994 to 2015-2019), and 5 consecutive birth cohorts (1990-2015) were recorded. Age-period-cohort analysis using the intrinsic estimation method provided coefficient estimates for age, period, and cohort effects. These coefficients are converted to exponential values \[\text{exp}(\text{coefficient})=\text{ecoefficient}\], representing the RR of incident AD and AD-related DALY rates for a specific age, period, or birth cohort relative to the mean for all ages, periods, or birth cohorts [25]. Age-period-cohort analysis was performed using STATA software (version 15.0; StataCorp).

Incidence Rate Predicted

This study predicted the incidence rate of ADs from 2020 to 2044 by running a Nordpred age-period-cohort analysis by sex

Figure 1. Trends in the sex-specific incidence and DALY rate of ADs globally during 1990-2019. (A) ASIR of ADs and (B) ASDR of ADs. AD: anxiety disorder; ASDR: age-standardized DALY rate; ASIR: age-standardized incidence rate; DALY: disability-adjusted life-year.

JPR Analysis

The joinpoint model was applied to divide the temporal trends in AD incidence into several segments and estimate the APCs by sex. As displayed in Table 1 and Figure 2A, the ASIR of ADs among male individuals declined first (\(\text{APCs}_{1990-1999}=-0.12\%\)), significantly increased (\(\text{APCs}_{2005-2010}=0.67\%\)), and then increased thereafter (\(\text{APCs}_{2010-2014}=0.03\%\) and \(\text{APCs}_{2014-2019}=0.35\%\)). In contrast, the ASIR underwent 3 increases (\(\text{APCs}_{1995-2000}=0.64\%\), \(\text{APCs}_{2005-2010}=0.19\%\), and \(\text{APCs}_{2015-2019}=0.25\%\)) and 3 declines (\(\text{APCs}_{1990-1995}=-0.06\%\), \(\text{APCs}_{2000-2005}=-0.47\%\), and \(\text{APCs}_{2010-2015}=-0.43\%\)) among female individuals, as shown in Figure 2B. Moreover, the trend in the ASDR was consistent with the change in the ASIR (Table 1; Figure 2C and D). Over the entire study period, the ASIRs of ADs were 0.068% (95% CI 0.35-1.0) among male individuals and 0.012% (95% CI 0.8-1.0) among female individuals (Table 1; Figure 2A and B).

Results

Global Trends in the ASIRs and ASDRs of ADs From 1990 to 2019

The ASIR of ADs showed a slightly fluctuating upward trend, increasing from 579.30 (95% UI 467.81-702.90) per 100,000 in 1990 to 585.45 (95% UI 474.21-709.53) per 100,000 in 2019, with an increase of 1.06%. The growth rate of the ASIR was higher in male individuals than female individuals, although the incidence rate for female individuals is about 1.4 times higher than that of male individuals. Among them, the ASIR increased by 1.92% in male individuals and 0.43% in female individuals (Figure 1A).

The trend of ASDR for ADs was relatively stable from 360.55 (95% UI 250.82-492.46) per 100,000 in 1990 to 360.12 (95% UI 248.60-494.44) per 100,000 in 2019, decreased by 0.12%. The ASDR increased by 0.90% in male individuals and decreased by 0.6% in female individuals (Figure 1B).
Figure 2. Joinpoint regression analysis of the sex-specific age-standardized incidence and disability-adjusted life years rate for ADs in global from 1990 to 2019. (A) Age-standardized incidence rate (ASIR) for males; (B) Age-standardized incidence rate (ASIR) for females; (C) Age-standardized DALYs rate (ASDR) for males; (D) Age-standardized DALYs rate (ASDR) for females. For a higher-resolution version of this figure, see Multimedia Appendix 1.
Table 1. Joinpoint regression analysis of the sex-specific age-standardized incidence and DALY* rate for anxiety disorders globally from 1990 to 2019.

<table>
<thead>
<tr>
<th>Categories, sex, and period</th>
<th>APC(^b) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incidence</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
</tr>
<tr>
<td>1990-1999</td>
<td>(-0.123 (–0.153 to –0.094))(^c)</td>
</tr>
<tr>
<td>1999-2005</td>
<td>(-0.363 (–0.435 to –0.292))(^c)</td>
</tr>
<tr>
<td>2005-2010</td>
<td>0.675 (0.572 to 0.777)(^c)</td>
</tr>
<tr>
<td>2010-2014</td>
<td>0.033 (–0.128 to –0.194)</td>
</tr>
<tr>
<td>2014-2019</td>
<td>0.353 (0.280 to 0.425)(^c)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
</tr>
<tr>
<td>1990-1995</td>
<td>(-0.060 (–0.095 to –0.025))(^c)</td>
</tr>
<tr>
<td>1995-2000</td>
<td>0.643 (0.593 to 0.693)(^c)</td>
</tr>
<tr>
<td>2000-2005</td>
<td>(-0.469 (–0.518 to –0.419))(^c)</td>
</tr>
<tr>
<td>2005-2010</td>
<td>0.189 (0.139 to 0.238)(^c)</td>
</tr>
<tr>
<td>2010-2015</td>
<td>(-0.475 (–0.376 to –1.860))(^c)</td>
</tr>
<tr>
<td>2015-2019</td>
<td>0.246 (0.196 to 0.296)(^c)</td>
</tr>
<tr>
<td><strong>AAPC(^d) of incidence</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>0.068 (0.035 to 0.100)(^c)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>0.012 (–0.006 to 0.030)</td>
</tr>
<tr>
<td><strong>DALY rates</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
</tr>
<tr>
<td>1990-1999</td>
<td>(-0.158 (–0.183 to –0.132))(^c)</td>
</tr>
<tr>
<td>1999-2005</td>
<td>(-0.387 (–0.491 to –0.283))(^c)</td>
</tr>
<tr>
<td>2005-2010</td>
<td>0.688 (0.583, 0.792)(^c)</td>
</tr>
<tr>
<td>2010-2014</td>
<td>(-0.128 (–0.292 to 0.037))</td>
</tr>
<tr>
<td>2014-2019</td>
<td>0.326 (0.252 to 0.400)(^c)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
</tr>
<tr>
<td>1990-1996</td>
<td>(-0.035 (–0.073 to 0.003))</td>
</tr>
<tr>
<td>1996-2000</td>
<td>1.027 (0.914 to 1.140)(^c)</td>
</tr>
<tr>
<td>2000-2005</td>
<td>(-0.470 (–0.540 to –0.400))(^c)</td>
</tr>
<tr>
<td>2005-2010</td>
<td>0.121 (0.050 to 0.192)(^c)</td>
</tr>
<tr>
<td>2010-2014</td>
<td>(-0.714 (–0.825 to –0.603))(^c)</td>
</tr>
<tr>
<td>2014-2019</td>
<td>0.063 (0.013 to 0.113)(^c)</td>
</tr>
<tr>
<td><strong>AAPC of DALY rates</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td>0.035 (0.001 to 0.070)(^c)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>(-0.015 (–0.042 to –0.013))</td>
</tr>
</tbody>
</table>

\(^a\)DALY: disability-adjusted life-year.

\(^b\)APC: annual percent change.

\(^c\)Changes that are statistically significant.
dAAPC: average annual percent change.

**Age-Period-Cohort Analysis**

**Age Effect**

Table 2 and Figure 3A and B show the longitudinal age curve of the incidence and DALY rates of ADs. After adjusting for period effects, in the reference cohort, the incidence of ADs showed 2 upward and downward trends with age in the reference cohort, similar to an M-shaped curve. Among both male and female individuals, individuals aged 10-14 years had the highest incidence of ADs. The second peak occurred at ages 40-44 years for male individuals and at ages 35-39 years for female individuals. In addition, the DALY rate of ADs was highest at 15-19 years of age for male individuals and 20-24 years of age for female individuals, and then the DALY rate decreased with age.
Table 2. RRs of the incidence and disability-adjusted life-year (DALY) rates of anxiety disorders globally from 1990 to 2019 due to age, period, and cohort effects.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Incidence, RR (95% CI)</th>
<th>DALY rates, RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-4</td>
<td>0.198 (0.180-0.218)</td>
<td>0.214 (0.198-0.232)</td>
</tr>
<tr>
<td>5-9</td>
<td>1.204 (1.153-1.257)</td>
<td>1.323 (1.274-1.374)</td>
</tr>
<tr>
<td>10-14</td>
<td>1.626 (1.565-1.688)</td>
<td>1.838 (1.779-1.899)</td>
</tr>
<tr>
<td>15-19</td>
<td>1.405 (1.325-1.460)</td>
<td>1.675 (1.622-1.730)</td>
</tr>
<tr>
<td>20-24</td>
<td>1.324 (1.273-1.376)</td>
<td>1.615 (1.564-1.667)</td>
</tr>
<tr>
<td>25-29</td>
<td>1.397 (1.345-1.452)</td>
<td>1.679 (1.627-1.732)</td>
</tr>
<tr>
<td>30-34</td>
<td>1.476 (1.421-1.533)</td>
<td>1.752 (1.698-1.809)</td>
</tr>
<tr>
<td>35-39</td>
<td>1.558 (1.499-1.618)</td>
<td>1.828 (1.770-1.888)</td>
</tr>
<tr>
<td>40-44</td>
<td>1.584 (1.523-1.646)</td>
<td>1.805 (1.746-1.866)</td>
</tr>
<tr>
<td>45-49</td>
<td>1.554 (1.493-1.617)</td>
<td>1.684 (1.626-1.745)</td>
</tr>
<tr>
<td>50-54</td>
<td>1.523 (1.462-1.585)</td>
<td>1.559 (1.502-1.618)</td>
</tr>
<tr>
<td>55-59</td>
<td>1.484 (1.425-1.547)</td>
<td>1.424 (1.369-1.480)</td>
</tr>
<tr>
<td>60-64</td>
<td>1.369 (1.313-1.428)</td>
<td>1.256 (1.205-1.308)</td>
</tr>
<tr>
<td>65-69</td>
<td>1.177 (1.126-1.230)</td>
<td>1.056 (1.010-1.103)</td>
</tr>
<tr>
<td>70-74</td>
<td>0.982 (0.937-1.029)</td>
<td>0.845 (0.806-0.886)</td>
</tr>
<tr>
<td>75-79</td>
<td>0.782 (0.744-0.823)</td>
<td>0.662 (0.609-0.656)</td>
</tr>
<tr>
<td>80-84</td>
<td>0.595 (0.562-0.630)</td>
<td>0.446 (0.420-0.474)</td>
</tr>
<tr>
<td>85-89</td>
<td>0.422 (0.395-0.451)</td>
<td>0.323 (0.302-0.347)</td>
</tr>
<tr>
<td>90-94</td>
<td>0.250 (0.228-0.273)</td>
<td>0.195 (0.177-0.215)</td>
</tr>
<tr>
<td><strong>Period</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>1.020 (1.000-1.041)</td>
<td>1.039 (1.021-1.058)</td>
</tr>
<tr>
<td>1999</td>
<td>1.001 (0.981-1.022)</td>
<td>1.047 (1.028-1.065)</td>
</tr>
<tr>
<td>2004</td>
<td>0.979 (0.960-0.999)</td>
<td>1.010 (0.993-1.028)</td>
</tr>
<tr>
<td>2009</td>
<td>0.996 (0.977-1.017)</td>
<td>0.996 (0.979-1.014)</td>
</tr>
<tr>
<td>2014</td>
<td>0.995 (0.975-1.016)</td>
<td>0.961 (0.944-0.978)</td>
</tr>
<tr>
<td>2019</td>
<td>1.008 (0.987-1.029)</td>
<td>0.951 (0.933-0.969)</td>
</tr>
<tr>
<td><strong>Cohort</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>0.919 (0.749-1.128)</td>
<td>0.800 (0.642-0.997)</td>
</tr>
<tr>
<td>1905</td>
<td>0.927 (0.816-1.052)</td>
<td>0.806 (0.704-0.924)</td>
</tr>
<tr>
<td>1910</td>
<td>0.939 (0.855-1.032)</td>
<td>0.823 (0.743-0.911)</td>
</tr>
<tr>
<td>1915</td>
<td>0.953 (0.883-1.028)</td>
<td>0.849 (0.783-0.920)</td>
</tr>
<tr>
<td>1920</td>
<td>0.964 (0.904-1.028)</td>
<td>0.867 (0.811-0.928)</td>
</tr>
<tr>
<td>1925</td>
<td>0.964 (0.910-1.020)</td>
<td>0.886 (0.836-0.940)</td>
</tr>
<tr>
<td>Year</td>
<td>Incidence, RR (95% CI)</td>
<td>DALY rates, RR (95% CI)</td>
</tr>
<tr>
<td>------</td>
<td>------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>1930</td>
<td>0.968 (0.918-1.021)</td>
<td>0.904 (0.856-0.955)</td>
</tr>
<tr>
<td>1935</td>
<td>0.969 (0.921-1.020)</td>
<td>0.919 (0.873-0.968)</td>
</tr>
<tr>
<td>1940</td>
<td>0.972 (0.926-1.021)</td>
<td>0.933 (0.889-0.980)</td>
</tr>
<tr>
<td>1945</td>
<td>0.979 (0.933-1.026)</td>
<td>0.954 (0.911-0.999)</td>
</tr>
<tr>
<td>1950</td>
<td>0.981 (0.936-1.027)</td>
<td>0.970 (0.929-1.013)</td>
</tr>
<tr>
<td>1955</td>
<td>0.993 (0.949-1.039)</td>
<td>0.992 (0.952-1.034)</td>
</tr>
<tr>
<td>1960</td>
<td>1.004 (0.961-1.049)</td>
<td>1.014 (0.975-1.054)</td>
</tr>
<tr>
<td>1965</td>
<td>1.000 (0.958-1.044)</td>
<td>1.025 (0.988-1.064)</td>
</tr>
<tr>
<td>1970</td>
<td>1.009 (0.968-1.053)</td>
<td>1.043 (1.007-1.081)</td>
</tr>
<tr>
<td>1975</td>
<td>1.032 (0.991-1.075)</td>
<td>1.071 (1.036-1.108)</td>
</tr>
<tr>
<td>1980</td>
<td>1.044 (1.004-1.086)</td>
<td>1.097 (1.063-1.133)</td>
</tr>
<tr>
<td>1985</td>
<td>1.048 (1.009-1.089)</td>
<td>1.113 (1.079-1.148)</td>
</tr>
<tr>
<td>1990</td>
<td>1.068 (1.026-1.112)</td>
<td>1.139 (1.102-1.177)</td>
</tr>
<tr>
<td>1995</td>
<td>1.078 (1.032-1.126)</td>
<td>1.164 (1.123-1.207)</td>
</tr>
<tr>
<td>2000</td>
<td>1.070 (1.018-1.124)</td>
<td>1.183 (1.135-1.232)</td>
</tr>
<tr>
<td>2005</td>
<td>1.050 (0.989-1.114)</td>
<td>1.193 (1.135-1.253)</td>
</tr>
<tr>
<td>2010</td>
<td>1.042 (0.957-1.133)</td>
<td>1.213 (1.131-1.302)</td>
</tr>
<tr>
<td>2015</td>
<td>1.055 (0.835-1.332)</td>
<td>1.259 (1.040-1.524)</td>
</tr>
<tr>
<td></td>
<td>Deviance</td>
<td>AIC</td>
</tr>
<tr>
<td>0.073</td>
<td>8.597</td>
<td>−317.081</td>
</tr>
<tr>
<td>0.067</td>
<td>8.804</td>
<td>−317.523</td>
</tr>
<tr>
<td>0.042</td>
<td>8.117</td>
<td>−319.184</td>
</tr>
<tr>
<td>0.062</td>
<td>8.656</td>
<td>−317.839</td>
</tr>
</tbody>
</table>

**Notes:**
- RR: relative risk.
- P < 0.001.
- P < 0.05.
- P < 0.01.
- AIC: Akaike information criterion.
- BIC: Bayesian information criterion.
**Figure 3.** RRs of the incidence and disability-adjusted life-year (DALY) rate of anxiety disorders globally from 1990 to 2019 due to age, period, and cohort effects. (A) Age effects on incidence, (B) age effects on DALY rate, (C) period effects on incidence, (D) period effects on DALY rate, (E) cohort effects on incidence, and (F) cohort effects on DALY rate. RR: relative risk.

**Period Effect**
Table 2 and Figure 3C and D show the estimated period effects by sex during the whole study period. Regarding the incidence, the period effects showed a downward trend from 1990-1994 to 2000-2004 for male individuals, with the period RR decreasing from 1.02 to 0.98 and then slightly increasing in 2015-2019. However, the period RR slightly increased from 1990-1994 to 1995-1999 for female individuals, and then it decreased. Regarding the DALY rate, the period RR showed similar increasing patterns for both sexes (increased by 13.63% for male individuals and 11.70% for female individuals).

**Cohort Effect**
The risk of ADs increased slowly with the year of birth for both sexes (Table 2; Figure 3E). Specifically, from the 1990-1994 to 2015-2019 birth cohorts, the RR of AD incidence substantially increased by 57.34% for female individuals. However, the cohort RR slightly increased by 17.32% from 1990-1994 to 1995-1999 for male individuals and then slightly
fluctuated. Regarding the DALY rate, the period RR showed similar decreasing patterns for both sexes (decreased by 40.94% for male individuals and 42.64% for female individuals), as shown in Table 2 and Figure 3F.

Predicted Trends of ADs in 2020-2044
We predicted that the future ASIR of ADs would be relatively stable, but the trend would be the opposite for male individuals compared to female individuals. However, the incidence among male individuals will increase, and the incidence among female individuals will gradually decrease (Figure 4).

Discussion
Principal Findings
ADs are some of the most common mental disorders, and their prevalence and disease burden are affected by population growth trends, socioeconomic level, the natural environment, and other factors [5]. Our research analyzed the long-term trend in the incidence and DALY rate of ADs globally; evaluated the potential effects of age, period, and birth cohort; and predicted the prevalence trends in the next 25 years. This study provides guidance for the development of effective prevention and control policies.

In 1990, the CIR of ADs was 581.81 per 100,000, and it increased to 592.20 per 100,000 in 2019 (an increase of 1.79%). The DALY rate in 1990 was 348.81 per 100,000, and it increased to 370.61 per 100,000 in 2019 (an increase of 6.25%). This trend shows that ADs are still a global public health issue that cannot be ignored and that we need to strengthen the prevention and timely diagnosis and treatment of affected populations.

Our research showed that compared with the increase in the CIR and DALY rate, the increase in the ASIR of ADs was smaller, and the ASDR was slightly lower from 1990 to 2019. JPR model analysis showed that the trends of ASIR and ASDR were consistent in both sexes. Generally, the incidence of ADs and their DALYs should be very similar because DALYs are years of life lost plus years lived with disability, and years of life lost only accounts for a very small proportion of ADs in terms of DALYs because ADs are not a fatal disease. Therefore, it was not surprising that the ASDR would be very similar to the ASIR. Notably, the ASIR and ASDR decreased substantially between 2000 and 2005 for both sexes. This downward trend may be due to the rapid development of the global economy and culture, the great improvement of medical standards, and the declining birth rate [20].

In terms of incidence, the ASIR was higher among female individuals than among male individuals in 2019. This was consistent with the results of previous studies [11,26]. In 2015, a systematic review of prevalence studies showed that female individuals were twice as likely as male individuals to have ADs [4], which indicated that more female individuals are affected by ADs than male individuals worldwide. A strong body of evidence implicates sex hormone fluctuations in female individuals as the major biological factor driving sex differences in anxiety risk [11,26,27]. The ASDR among female individuals
was also higher than that among male individuals, which is in line with a higher prevalence among female individuals than among male individuals. In addition, mental health knowledge has increased considerably, and stigma has decreased in countries that allow male individuals to talk about their feelings, seek help, and overcome the “rule”: male individuals drink and female individuals become depressed or anxious [28]. However, the ASIR of ADs was relatively stable (AAPCs 0.012%) and the ASDR decreased (AAPCs –0.015%) among female individuals, while both the ASIR (AAPCs 0.068%) and ASDR (AAPCs 0.035%) increased among male individuals. This may reveal a potentially increasing burden of ADs among male individuals.

Age-period-cohort model analysis showed that the upward trend in AD incidence was affected by age, period, and cohort. Age is an important demographic risk factor, and adolescents and middle-aged individuals are at high risk of ADs, similar to an M-shaped curve. The period effect showed that the overall incidence rate showed a downward trend, while the DALY rate showed an upward trend. The incidence of ADs increased, while the DALY rate decreased by year of birth.

Regarding the age effect, the RR of AD incidence was the highest in the age group of 10-14 years for both sexes, and the RR of the DALY rate was highest in the age group of 15-19 years among male individuals and in the age group of 20-24 years among female individuals. These findings were supported by a previous review [29]. Studies have shown that the high incidence of ADs among adolescents is closely related to genetic factors [30,31], higher vulnerability caused by hormonal processes, and adverse childhood experiences (serious illness, parental separation and emotional maltreatment, and physical and sexual abuse) [32-35]. Furthermore, the risk of ADs incidence among male individuals aged 40-44 years and female individuals aged 35-39 years was similar to that among adolescents, and then the risk decreased with age. The high risk of middle-aged people may stem from occupational pressure, family relationships, economic burdens, physical status, and other factors [36-38].

The period effect is usually caused by a series of complex historical events and environmental factors. We observed upward trends in the risk of developing ADs over time among male individuals after 2004. Notably, the incidence of ADs among male individuals increased rapidly during 2005-2010 and continued to increase until 2019 according to joinpoint analysis. This may be due to environmental changes and a more rapid pace of life [39,40]. During the study period, the world experienced a bird flu pandemic, rising oil prices, repeated terrorist attacks, and more severe natural disasters [41-44]. However, the period effect showed that the incidence of ADs among female individuals decreased over time after 1999, which was consistent with the reduction in incidence from 2000 to 2005 in the JPR analysis. The decline in incidence among female individuals may be due to higher levels of education, higher socioeconomic status, and a relative decline in fertility [38,45-47]. Compared with the incidence rate, the RR of the DALY rate increased over time in both sexes. This trend may be explained by the higher lifetime prevalence of ADs and the low and limited treatment rates in the early stage, which makes most patients with early-onset ADs develop a persistent course [48,49]. Long-term persistence or repeated attacks aggravate the degree of the disease and are prone to comorbidities and physical diseases, which increase the disability rate of these patients in the later stage [15]. Studies have found a lack of associations between sociodemographic characteristics and AD persistence, such as socioeconomic status and level of education, and inconclusive results for other socioeconomic factors, such as sex and age, although these characteristics were repeatedly found to be associated with the onset and prevalence of ADs [5].

The cohort effect on the incidence of ADs revealed continuous upward trends in later birth cohorts for both male and female individuals. The possible reason was that the later birth cohorts faced greater pressures, such as environmental degradation, dwindling resources, and economic depression. In contrast, the DALY rate of ADs decreased with birth year. The reason may be that the latest birth cohort was better educated, had better health care, and was more aware of health and disease prevention than the earliest birth cohort.

The predicted trend of AD incidence in our study over the next 25 years is consistent with the current trend, with a downward trend among female individuals and an upward trend among male individuals. This trend calls for us to pay more attention to the risk of ADs, especially among male individuals, and formulate effective prevention and treatment policies to reduce the incidence of ADs and thus reduce the disease burden.

In summary, between 1990 and 2019, the incidence of ADs among male and female individuals worldwide was dynamic. Overall, the global incidence rate among male individuals increased significantly while that among female individuals decreased significantly. According to age-period-cohort analysis, the incidence of the age effect showed an M-shaped curve. That is, adolescents and middle-aged people had the highest incidence. In addition, the incidence of ADs among male individuals will continue to increase soon. In this regard, we should pay more attention to the prevalence of ADs among male individuals, provide greater intervention and treatment for female individuals, and identify adolescent and middle-aged risk groups to support the timely initiation of intervention measures and effectively reduce the disease burden of ADs. At the same time, the COVID-19 pandemic has heightened anxiety, increasing the incidence of ADs and their disease burden [50]. Therefore, multiple interventions should be used aggressively to alleviate anxiety and promote mental health during the COVID-19 pandemic [51].

Limitations

This study has limitations. First, GBD data are modeled at a higher setting and are derived from the analysis of limited raw data. Nevertheless, the results of this study rely heavily on the outcome data of the GBD study 2019. Second, the age-period-cohort model is the mathematical method used, and it does not fully reflect the reality of the situation. In addition, the GBD data do not include groups that are transgender. Therefore, the results should be interpreted with caution.
Conclusions
In summary, ADs are becoming more common in both male and female individuals, and the disease burden in male individuals has increased slightly, indicating that early prevention and effective control cannot be ignored. The incidence of the age effect showed an M-shaped curve, and the incidence of ADs in male individuals will continue to increase soon. This suggests that we should take active intervention measures, focusing on high-risk groups, and develop effective management and control policies to reduce GBD.

Acknowledgments
We would like to thank the Institute for Health Metrics and Evaluation staff and its collaborators who prepared these publicly available data. This study was supported by the Science-Technology Foundation for Middle-aged and Young Scientist of Wannan Medical College (WKS2022F03).

Data Availability
The data sets supporting the conclusions of this article are available in the Global Burden of Disease Data Tool repository.

Authors' Contributions
HC, YW, and MT contributed to the conception and design of the study. YS organized the database. H Yin and YS performed the statistical analysis. HC and YW wrote the manuscript. H Yuan critically revised the manuscript for important academic content. MT supervised the research and is the corresponding author. All authors approved the final version to be published and can certify that no other individuals not listed as authors have made substantial contributions to the paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Joinpoint regression analysis of the sex-specific age-standardized incidence and DALY rate for ADs globally from 1990 to 2019. (A) ASIR for male individuals; (B) ASIR for female individuals; (C) ASDR for male individuals; (D) ASDR for female individuals. AD: anxiety disorder; DALY: disability-adjusted life-year; APC: average percentage change; ASDR: age-standardized DALY rate; ASIR: age-standardized incidence rate.

References


Abbreviations

AAPC: average annual percent change
AD: anxiety disorder
APC: annual percent change
ASDR: age-standardized disability-adjusted life-year rate
ASIR: age-standardized incidence rate
CIR: crude incidence
DALY: disability-adjusted life-years
DSM-5: Diagnostic and Statistical Manual of Mental Disorder, Fifth Edition
GBD: Global Burden of Disease
JPR: joinpoint regression
RR: relative risk
UI: uncertainty interval

Edited by A Mavragani, T Sanchez; submitted 03.06.23; peer-reviewed by YL Chen, G Mayer; comments to author 03.08.23; revised version received 25.08.23; accepted 14.12.23; published 29.01.24.

Please cite as:
Cao H, Wu Y, Yin H, Sun Y, Yuan H, Tao M
Global Trends in the Incidence of Anxiety Disorders From 1990 to 2019: Joinpoint and Age-Period-Cohort Analysis Study
JMIR Public Health Surveill 2024;10:e49609
URL: https://publichealth.jmir.org/2024/1/e49609
doi:10.2196/49609
PMID:38285497

©Huiru Cao, Yang Wu, Hui Yin, Yanqi Sun, Hui Yuan, Mengjun Tao. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 29.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Relationship Between Lipoprotein(a), Renal Function Indicators, and Chronic Kidney Disease: Evidence From a Large Prospective Cohort Study

Yingxin Liu¹, MPH; Ruoting Wang¹, MSc; Shuai Li¹, PhD; Changfa Zhang¹, MD; Gregory Y H Lip²,³, MD; Lehana Thabane⁴, PhD; Guowei Li¹,⁴,⁵, PhD

¹Center for Clinical Epidemiology and Methodology, Guangdong Second Provincial General Hospital, Guangzhou, China
²Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, Liverpool, United Kingdom
³Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
⁴Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
⁵Father Sean O’Sullivan Research Centre, St Joseph’s Healthcare Hamilton, Hamilton, ON, Canada

Corresponding Author:
Guowei Li, PhD
Center for Clinical Epidemiology and Methodology, Guangdong Second Provincial General Hospital
No. 466 Xingang Middle Road, Haizhu district
Guangzhou, 510317
China
Phone: 86 020 32640264
Email: lig28@mcmaster.ca

Abstract

Background: Chronic kidney disease (CKD) poses a significant global public health challenge. While lipoprotein(a) (Lp[a]) has been established as a significant factor in cardiovascular disease, its connection to CKD risk remains a topic of debate. Existing evidence indicates diverse risks of kidney disease among individuals with various renal function indicators, even when within the normal range.

Objective: This study aims to investigate the joint associations between different renal function indicators and Lp(a) regarding the risks of incident CKD in the general population.

Methods: The analysis involved a cohort of 329,415 participants without prior CKD who were enrolled in the UK Biobank between 2006 and 2010. The participants, with an average age of 56 (SD 8.1) years, included 154,298/329,415 (46.84%) males. At baseline, Lp(a) levels were measured using an immunoturbidimetric assay and classified into 2 groups: low (<75 nmol/L) and high (≥75 nmol/L). To assess participants’ baseline renal function, we used the baseline urine albumin-to-creatinine ratio (UACR) and estimated glomerular filtration rate (eGFR). The relationship between Lp(a), renal function indicators, and the risk of CKD was evaluated using multivariable Cox regression models. These models were adjusted for various factors, including sociodemographic variables, lifestyle factors, comorbidities, and laboratory measures.

Results: A total of 6003 incident CKD events were documented over a median follow-up period of 12.5 years. The association between elevated Lp(a) levels and CKD risk did not achieve statistical significance among all participants, with a hazard ratio (HR) of 1.05 and a 95% CI ranging from 0.98 to 1.13 (P=.16). However, a notable interaction was identified between Lp(a) and UACR in relation to CKD risk (P for interaction=.04), whereas no significant interaction was observed between Lp(a) and eGFR (P for interaction=.96). When compared with the reference group with low Lp(a) and low-normal UACR (<75 nmol/L), the group with high Lp(a) and low-normal UACR exhibited a nonsignificant association with CKD risk (HR 0.98, 95% CI 0.90-1.08; P=.74). By contrast, both the low Lp(a) and high-normal UACR (≥10 mg/g) group (HR 1.16, 95% CI 1.08-1.24; P<.001) and the high Lp(a) and high-normal UACR group (HR 1.32, 95% CI 1.19-1.46; P<.001) demonstrated significant associations with increased CKD risks. In individuals with high-normal UACR, elevated Lp(a) was linked to a significant increase in CKD risk, with an HR of 1.14 and a 95% CI ranging from 1.03 to 1.26 (P=.01). Subgroup analyses and sensitivity analyses consistently produced results that were largely in line with the main findings.

Conclusions: The analysis revealed a significant interaction between Lp(a) and UACR in relation to CKD risk. This implies that Lp(a) may act as a risk factor for CKD even when considering UACR. Our findings have the potential to provide valuable
insights into the assessment and prevention of CKD, emphasizing the combined impact of Lp(a) and UACR from a public health perspective within the general population. This could contribute to enhancing public awareness regarding the management of Lp(a) for the prevention of CKD.

*(JMIR Public Health Surveill 2024;10:e50415) doi:10.2196/50415*

**KEYWORDS**
lipoprotein(a); chronic kidney disease; renal function; urinary albumin-to-creatinine ratio; glomerular filtration rate

**Introduction**
Chronic kidney disease (CKD), a significant contributor to cardiovascular disease (CVD) and mortality, affected approximately 697.5 million individuals worldwide, with a prevalence of 9.1% in 2017 [1,2]. In public health and clinical research, CKD is typically diagnosed using the estimated glomerular filtration rate (eGFR) and urine tests to detect the presence of albumin or protein or a combination of both [3]. As per recommendations [3], a diagnosis of CKD can be made when the eGFR is <60 mL/min/1.73 m² and is combined with albuminuria, usually defined as a urinary albumin-to-creatinine ratio (UACR) ≥30 mg/g.

Lipoprotein(a) (Lp(a)) is a particle resembling low-density lipoprotein (LDL), comprising a large glycoprotein apolipoprotein(a) bound to an apolipoprotein B100 molecule [4]. The adverse impact of Lp(a) on CVD has been robustly substantiated by epidemiological, experimental, and genetic studies [5,6]. By contrast, the association between Lp(a) and the risk of CKD is still a topic of debate, with inconsistent findings reported in both observational studies and Mendelian randomization studies [7-14]. As a result, Lp(a) has not yet been incorporated into current guidelines or public health policies for CKD prevention, detection, management, or surveillance [3,15]. A deeper understanding of the role of Lp(a) in CKD could enhance public initiatives for the prevention and risk management of CKD. Therefore, further explorations are needed to investigate the relationship between Lp(a) and CKD risk.

Remarkably, no previous studies have investigated the association between the risk of CKD and Lp(a) in conjunction with eGFR or UACR, despite both measures being commonly used as indicators of renal function [8-10]. Heterogeneous risks of CKD have been observed in participants with different baseline eGFR or UACR measures, even when these 2 indicators were within the normal range [16-19]. For example, evidence has suggested that a UACR value in the high-normal range (10-30 mg/g) is significantly associated with the progression of CKD and renal failure when compared with the low-normal group (<10 mg/g) used as a reference [17,18,20]. While some previous studies generally adjusted for renal function indicators in their regression models, the association between Lp(a) and CKD risk was estimated based on an average level of renal function indicators. Therefore, the inconsistent associations between Lp(a) and CKD risk may, at least in part, depend on participants’ varying renal function levels across different studies.

In this study, our objective was to investigate the combined associations of Lp(a) and renal function indicators in relation to CKD risk among participants without a history of CKD from the UK Biobank cohort study. Exploring the potential interplay between Lp(a) and renal function may provide new evidence for assessing and preventing CKD risk in the general population from a public health perspective. This research could contribute to raising public awareness about the importance of managing Lp(a) for CKD prevention.

**Methods**

**Study Population**
Information about the UK Biobank study has been extensively documented in prior literature and is available on the official website [21,22]. In summary, the UK Biobank is a comprehensive cohort study that encompasses biological and medical data from approximately half a million residents in the United Kingdom since 2006. Enrolled participants provided written informed consent, and the study received approval from the North West Multi-Centre Research Ethics Committee.

A total of 502,411 participants were included in our analyses. Exclusions were made for participants with missing data on eGFR (n=33,141), UACR (n=13,408), or Lp(a) (n=81,863), as well as those with renal dysfunction at baseline, including a diagnosis of CKD, an eGFR<60 mL/min/1.73 m², or a UACR≥30 mg/g (n=44,584). Consequently, the final analysis included 329,415 participants. The participant selection process is depicted in Figure S1 in Multimedia Appendix 1.

**Outcomes**
Within the UK Biobank, incident disease status and death information were determined through linkage with hospital in-patient data, cancer registry records, and death registry records. Our primary outcome focused on event-free survival time to the first moderate to severe CKD event. CKD events included stages 3-5 and end-stage renal disease (ESRD), identified by ICD-10 (10th revision of the International Statistical Classification of Diseases and Related Health Problems) codes N18.0, N18.3, N18.4, and N18.5, ascertained from hospital in-patient records in either the primary or the secondary position. The secondary outcomes in our study encompassed the individual components of CKD, specifically CKD stage 3, CKD stage 4, and CKD stage 5 as well as ESRD. All participants were monitored from the date of recruitment (spanning from 2006 to 2010) to the occurrence of a CKD diagnosis, death, or the conclusion of the follow-up period (September 30, 2021, for England; July 31, 2021, for Scotland; and February 28, 2018, for Wales), whichever transpired first.
Exposures

Serum Lp(a) levels were assessed using an immunoturbidimetric assay (Beckman Coulter AU5800; Randox Laboratories). In accordance with guidelines, Lp(a) was categorized into 2 groups: low (<75 nmol/L) and high (≥75 nmol/L) [4,9,23]. Serum creatinine (mmol/L) and urinary creatinine (mmol/L) were determined through enzymatic analyses (Beckman Coulter AU5800). The eGFR was computed using the serum creatinine–based Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation [24]. UACR was calculated as the ratio of urinary albumin (in mg/L) to urinary creatinine, with the former measured by an immunoturbidimetric assay (Beckman Coulter AU5400). As per recommendations, we classified both eGFR (low-normal, <90 mL/min/1.73 m² and high-normal, ≥90 mL/min/1.73 m²) and UACR (low-normal, <10 mg/g and high-normal, ≥10 mg/g) into 2 groups [3].

Other Independent Variables

Additional baseline independent variables considered comprised sociodemographic factors, lifestyle details, comorbidities, medication use, and laboratory samples. Sociodemographic factors included age (in years), sex (male or female), Townsend Deprivation Index (TDI), ethnicity (White, Mixed, Asian, Black, Chinese, and others), college degree or higher (yes or no), and residential area (urban or rural).

Lifestyle variables were BMI, smoking status (never, previous, or current smoker), alcohol drinking status (never, previous, or current drinker), regular vitamin supplement consumption (yes or no), mineral supplement use (yes or no), and coffee intake (yes or no). Comorbidities comprised a previous history of cancer, nonhypertensive CVD, depression, diabetes mellitus (DM), hypertension, high cholesterol, and use of drugs (anti-diabetic drugs, anti-hypertensive drugs, or cholesterol-lowering drugs). Laboratory samples included high-density lipoprotein (HDL)-cholesterol (HDL-C), LDL-cholesterol (LDL-C), triglycerides (TGs), C-reactive protein (CRP), glycated hemoglobin (HbA₁c), and urate. Table S1 in Multimedia Appendix 1 provides details on the aforementioned variables.

Data on sociodemographic factors and lifestyles were obtained through participant self-reports at baseline interviews. Information on baseline comorbidities and drug usage was gathered from participant self-reports, hospital in-patient records at baseline, and the relevant treatment/medication received. Laboratory samples, including blood and urine, were collected during participant recruitment.

Statistical Analyses

Baseline characteristics of the included participants were presented as mean (SD) for continuous variables or percentage (percentage) for categorical variables. Chi-square and independent t tests (2-tailed) were performed to compare categorical and continuous variables, respectively, stratified by low and high Lp(a) groups.

Effect modification analyses were conducted to assess whether the impact of the high Lp(a) group on CKD risk varied within the strata of eGFR or UACR. We observed significant modifications by UACR (P=0.02 for relative excess risk due to interaction), indicating that the association between Lp(a) and CKD risk was influenced by UACR levels, while no significant modifications were found by eGFR (P=0.12 for relative excess risk due to interaction; refer to Tables S2 and S3 in Multimedia Appendix 1). When considering Lp(a), UACR, and eGFR as continuous variables, a significant interaction was observed only between Lp(a) and UACR (P=0.03), while no significant interaction was found between Lp(a) and eGFR (P=0.27) concerning CKD risk (refer to Table S4 in Multimedia Appendix 1). Subsequently, we delved into exploring the joint effect of Lp(a) and UACR on the risk of CKD through further analyses categorizing participants into 4 groups: low Lp(a) and low-normal UACR, low Lp(a) and high-normal UACR, high Lp(a) and low-normal UACR, and high Lp(a) and high-normal UACR.

Multivariable Cox proportional hazards models were used to explore the joint associations between Lp(a), UACR, and the risk of CKD, using the low Lp(a) and low-normal UACR group as the reference. The fully adjusted models included covariate adjustments for age, sex, BMI, TDI, college degree, ethnicity, area, smoking and drinking status, regular intake of coffee, vitamin and mineral supplements, personal medical history of cancer, CVD, depression, DM, hypertension, high cholesterol, drugs for DM, hypertension, high cholesterol, systolic blood pressure, HDL-C, LDL-C, TGs, HbA₁c, CRP, urate, and eGFR. The variables included in the models were selected based on clinical expertise, prevailing research practices, and statistical knowledge [7-12]. The results were presented as hazard ratios (HRs) along with their corresponding 95% CIs. Additionally, a parsimonious model was used, adjusting only for age, sex, BMI, comorbidities, use of drugs, and eGFR, to evaluate the consistency of results with those from the fully adjusted models.

Subgroup analyses were performed to investigate the relationship between Lp(a), UACR, and CKD risk, stratified by sex (male vs female), age (<65 vs ≥65 years), medical history of DM (yes vs no), medical history of hypertension (yes vs no), and medical history of high cholesterol (yes vs no). In the subgroup analysis involving participants with DM, additional adjustment was made for their DM duration, considering the close association between DM duration and kidney function [25].

Several sensitivity analyses were conducted to assess the robustness of the main results. Initially, considering the potential correlation between LDL-C and Lp(a), we conducted the same analyses in multivariable models, excluding LDL-C to mitigate potential multicollinearity. Additionally, a sensitivity analysis was performed by further adjusting for moderate-to-vigorous physical activity and sleep patterns. Sleep pattern was defined based on a previous study using the UK Biobank cohort and incorporating 5 sleep behaviors: chronotype, duration, insomnia, snoring, and excessive daytime sleepiness [26]. Furthermore, we adjusted for family history of kidney diseases and the use of drugs for kidney diseases, including angiotensin-converting enzyme inhibitors and angiotensin receptor blockers, in another sensitivity analysis. Although comorbidities, including obesity,
DM, and hypertension at baseline, were adjusted for in the models, it is possible that some participants without these comorbidities at baseline developed them during the 12.5-year follow-up. To address this, we conducted a sensitivity analysis by excluding participants who were free of these comorbidities at baseline but developed obesity, DM, or hypertension during the follow-up. This was done to minimize the dynamic impact of these comorbidities on the association between Lp(a), UACR, and CKD risk. We conducted a Fine-Gray competing risk analysis, treating all-cause mortality as a competing event for CKD [27]. To address reverse causation, we repeated the Cox regression analyses after excluding CKD events that occurred within the first year and the first 3 years of follow-up. Additionally, as another sensitivity analysis, we used multiple imputation techniques for missing data (seed=12345) to assess the robustness of the main findings. According to the guideline [28], participants with normal indicators (eGFR≥60 mL/min/1.73 m² and UACR<30 mg/g) can be further diagnosed with CKD stage 1 or 2 if they exhibit other markers of kidney damage (eg, hematuria, electrolyte abnormalities, or structural abnormalities detected by imaging such as polycystic or dysplastic kidneys). Unfortunately, due to the unavailability of specific disease markers, we were unable to fully identify participants who should be diagnosed with CKD stage 1 or 2 at baseline. Consequently, we conducted another sensitivity analysis by including participants with a baseline eGFR≥90 mL/min/1.73 m² in the analyses to exclude those with suspected CKD stage 1 or 2.

All tests were 2-sided with a significance level of .05. Statistical analyses were performed using SAS software version 9.4 (SAS Institute Inc.) and R software version 4.1.2 (R Foundation).

**Ethical Considerations**

The UK Biobank study was approved by the North West Multicenter Research Ethics Committee (reference number: 16/NW/0274). All participants provided written consent before enrollment. The present analysis has received an exemption from the Research Ethics Committee of Guangdong Second Provincial General Hospital (reference number: 2022-KY-KZ-119-01) because it was a secondary analysis based on open data according to current regulations.

**Results**

A total of 329,415 participants (mean age 56 years; n=175,117, 53.15% females) without prior CKD were included in the analyses. Among them, 258,388 (78.44%) had low Lp(a), while 71,027 (21.56%) had high Lp(a). Table 1 and Table S5 in Multimedia Appendix 1 provide descriptions and comparisons of baseline characteristics. Participants with high Lp(a) exhibited higher BMI, HDL-C, LDL-C, and HbA₁c levels, and were more likely to have a previous history of CVD and high cholesterol compared with the low Lp(a) group.
Table 1. Baseline characteristics of included participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (n=329,415)</th>
<th>Low Lp(a) (n=258,388)</th>
<th>High Lp(a) (n=71,027)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex, n (%)</td>
<td>154,298 (46.84)</td>
<td>120,890 (46.79)</td>
<td>33,408 (47.04)</td>
<td>.20</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>56.3 (8.09)</td>
<td>56.3 (8.09)</td>
<td>56.1 (8.13)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>27.3 (4.46)</td>
<td>27.3 (4.65)</td>
<td>27.4 (4.71)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Townsend Deprivation Index, mean (SD)</td>
<td>−1.35 (3.07)</td>
<td>−1.37 (3.05)</td>
<td>−1.29 (3.12)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>College degree or higher, n (%)</td>
<td>108,671 (32.99)</td>
<td>85,368 (33.04)</td>
<td>23,303 (32.81)</td>
<td>.25</td>
</tr>
<tr>
<td>Urban area, n (%)</td>
<td>280,042 (85.01)</td>
<td>219,554 (84.97)</td>
<td>60,488 (85.16)</td>
<td>.08</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>White</td>
<td>309,923 (94.08)</td>
<td>244,101 (94.47)</td>
<td>65,822 (92.67)</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>2016 (0.61)</td>
<td>1574 (0.61)</td>
<td>442 (0.62)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>6536 (1.98)</td>
<td>5212 (2.02)</td>
<td>1324 (1.86)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>5315 (1.61)</td>
<td>3038 (1.18)</td>
<td>2277 (3.21)</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>1112 (0.34)</td>
<td>985 (0.38)</td>
<td>127 (0.18)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>3027 (0.92)</td>
<td>2326 (0.90)</td>
<td>701 (0.99)</td>
<td></td>
</tr>
<tr>
<td>Moderate-to-vigorous physical activity (metabolic equivalent of task, minutes/week), n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.48</td>
</tr>
<tr>
<td>0</td>
<td>35,993 (10.93)</td>
<td>28,260 (10.94)</td>
<td>7733 (10.89)</td>
<td></td>
</tr>
<tr>
<td>1-599</td>
<td>68,698 (20.85)</td>
<td>53,928 (20.87)</td>
<td>14,770 (20.79)</td>
<td></td>
</tr>
<tr>
<td>600-1199</td>
<td>45,633 (13.85)</td>
<td>35,869 (13.88)</td>
<td>9764 (13.75)</td>
<td></td>
</tr>
<tr>
<td>≥1200</td>
<td>117,130 (35.56)</td>
<td>91,704 (35.49)</td>
<td>25,426 (35.80)</td>
<td></td>
</tr>
<tr>
<td>Smoking status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.06</td>
</tr>
<tr>
<td>Never</td>
<td>181,057 (54.96)</td>
<td>141,843 (54.90)</td>
<td>39,214 (55.21)</td>
<td></td>
</tr>
<tr>
<td>Previous</td>
<td>112,575 (34.17)</td>
<td>88,571 (34.28)</td>
<td>24,004 (33.80)</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>34,217 (10.39)</td>
<td>26,773 (10.36)</td>
<td>7444 (10.48)</td>
<td></td>
</tr>
<tr>
<td>Drinking status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>Never</td>
<td>14,099 (4.28)</td>
<td>10,924 (4.23)</td>
<td>3175 (4.47)</td>
<td></td>
</tr>
<tr>
<td>Previous</td>
<td>11,196 (3.40)</td>
<td>8776 (3.40)</td>
<td>2420 (3.41)</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>303,380 (92.10)</td>
<td>238,125 (92.16)</td>
<td>65,255 (91.87)</td>
<td></td>
</tr>
<tr>
<td>Coffee intake, n (%)</td>
<td>256,104 (77.75)</td>
<td>201,200 (77.87)</td>
<td>54,904 (77.30)</td>
<td>.001</td>
</tr>
<tr>
<td>Vitamin supplement, n (%)</td>
<td>103,367 (31.38)</td>
<td>80,797 (31.27)</td>
<td>22,570 (31.78)</td>
<td>.009</td>
</tr>
<tr>
<td>Mineral supplement, n (%)</td>
<td>140,300 (42.59)</td>
<td>110,173 (42.64)</td>
<td>30,127 (42.42)</td>
<td>.28</td>
</tr>
<tr>
<td>Sleep pattern, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.84</td>
</tr>
<tr>
<td>Poor</td>
<td>6302 (1.91)</td>
<td>4926 (1.91)</td>
<td>1376 (1.94)</td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>106,559 (32.35)</td>
<td>83,610 (32.36)</td>
<td>22,949 (32.31)</td>
<td></td>
</tr>
<tr>
<td>Healthy</td>
<td>159,283 (48.35)</td>
<td>124,911 (48.34)</td>
<td>34,372 (48.39)</td>
<td></td>
</tr>
<tr>
<td>Cancer, n (%)</td>
<td>36,744 (11.15)</td>
<td>29,065 (11.25)</td>
<td>7679 (10.81)</td>
<td>.001</td>
</tr>
<tr>
<td>Cardiovascular disease, n (%)</td>
<td>41,465 (12.59)</td>
<td>32,133 (12.44)</td>
<td>9535 (13.42)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Depression, n (%)</td>
<td>15,254 (4.63)</td>
<td>4282 (1.66)</td>
<td>19,536 (27.51)</td>
<td>.32</td>
</tr>
<tr>
<td>Diabetes mellitus, n (%)</td>
<td>25,129 (7.63)</td>
<td>19,721 (7.63)</td>
<td>5408 (7.61)</td>
<td>.97</td>
</tr>
<tr>
<td>Course of diabetes, mean (SD)</td>
<td>9.69 (13.2)</td>
<td>9.74 (13.2)</td>
<td>9.49 (13.0)</td>
<td>.37</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>87,697 (26.62)</td>
<td>68,541 (26.53)</td>
<td>19,156 (26.97)</td>
<td>.06</td>
</tr>
</tbody>
</table>
There were 6003 incident CKD events recorded, with a median follow-up of 12.5 years and a total of 4,011,201 person-years. Table 2 (also see Multimedia Appendix 2) presents the results for the independent associations between Lp(a), UACR, and CKD risk. The high Lp(a) group showed a nonsignificant association with an elevated CKD risk (HR 1.05, 95% CI 0.98-1.13; \( P = .16 \)) compared with the low Lp(a) group. By contrast, the high-normal UACR group was significantly associated with a 20% increased risk of CKD compared with the low-normal UACR group (HR 1.20, 95% CI 1.13-1.27; \( P < .001 \)). Similar results were observed when treating Lp(a) and UACR as continuous variables (see Table S4 in Multimedia Appendix 1).
Table 2. Associations between Lp(a)\(^a\), UACR\(^b\), and risks of chronic kidney disease.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Number of cases/total participants</th>
<th>HR(^c) (95% CI); P value(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent associations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lp(a)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low group</td>
<td>4642/258,388</td>
<td>Reference</td>
</tr>
<tr>
<td>High group</td>
<td>1361/71,027</td>
<td>1.05 (0.98-1.13); .16</td>
</tr>
<tr>
<td><strong>UACR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-normal group</td>
<td>3247/182,740</td>
<td>Reference</td>
</tr>
<tr>
<td>High-normal group</td>
<td>2756/146,675</td>
<td>1.20 (1.13-1.27); &lt; .001</td>
</tr>
<tr>
<td><strong>Joint associations between baseline Lp(a)(^e) and UACR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>2527/142,923</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>720/39,817</td>
<td>0.98 (0.90-1.08); .74</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>2115/15,465</td>
<td>1.16 (1.08-1.24); &lt; .001</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>641/31,210</td>
<td>1.32 (1.19-1.46); &lt; .001</td>
</tr>
</tbody>
</table>

\(^a\)Lp(a): lipoprotein a.
\(^b\)UACR: urine albumin-creatinine ratio.
\(^c\)HR: hazard ratio.
\(^d\)Fully adjusted model adjusted for age, sex, body mass index, Townsend Deprivation Index, college degree, ethnicity, area, smoking and drinking status, regular intake of coffee, vitamin and mineral supplement, personal medical history of cancer, cardiovascular disease, depression, diabetes, hypertension, high cholesterol, drugs for diabetes, hypertension, high cholesterol, systolic blood pressure, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides, glycated hemoglobin (HbA\(_1c\)), C-reactive protein, urate, and estimated glomerular filtration rate.
\(^e\)Baseline Lp(a) was categorized into low (<75 nmol/L) and high (≥75 nmol/L) groups. Baseline UACR within the normal range was classified into low-normal (0-9.9 mg/g) and high-normal (10-29.9 mg/g) groups.

The joint associations between Lp(a) and UACR, as demonstrated through the 4 groups generated from their cross-categorization, with low Lp(a) and low-normal UACR as the reference group, are shown in Table 2. Among participants with low-normal UACR, a nonsignificant association was observed between high Lp(a) and CKD risk (HR 0.98, 95% CI 0.90-1.08; P=.74). When compared with the low Lp(a) and low-normal UACR group, both the low Lp(a) and high-normal UACR and high Lp(a) and high-normal UACR groups were significantly associated with increased risks of CKD, with HRs of 1.16 (95% CI 1.08-1.24; P<.001) and 1.32 (95% CI 1.19-1.46; P<.001), respectively. Table S2 in Multimedia Appendix 1 shows that a significant association between Lp(a) and increased CKD risk was only observed in the high-normal UACR group (HR 1.14, 95% CI 1.03-1.26; P=.01), but not in the low-normal UACR group (HR 0.98, 95% CI 0.90-1.08; P=.74). Similar results from the parsimonious model were found to support our main findings (Table S6 in Multimedia Appendix 1).

When treating UACR as a continuous variable, increased HRs regarding the relationship between the high Lp(a) group and CKD risk were observed as UACR elevated (Figure 1), indicating that an increased UACR modifies the propensity of high Lp(a) toward CKD risk when compared with low Lp(a).
Figure 1. HRs for the high Lp(a) group regarding the risk of chronic kidney disease at different levels of UACR. CKD: chronic kidney disease; HR: hazard ratio; lower 95%: lower limit of 95% CI; Lp(a): lipoprotein(a); UACR: urine albumin-to-creatinine ratio; upper 95%: upper limit of 95% CI.

Figure 2 shows the joint associations between Lp(a) and UACR with the risk of CKD in different subgroups. Results similar to the main findings were observed within different strata of sex, previous history of DM, and high cholesterol. Among participants aged ≥65 years or without a previous history of hypertension, both the low Lp(a) and high-normal UACR (P=.08 for participants aged ≥65 years and P=.07 for participants without hypertension) and high Lp(a) and high-normal UACR (P=.10 for participants aged ≥65 years and P=.14 for participants without hypertension) groups were nonsignificantly associated with increased risks of CKD. Sensitivity analyses yielded largely similar results to the main findings (Figure 3 and Tables S7 and S8 in Multimedia Appendix 1).
Figure 2. Stratified analyses of joint associations between baseline Lp(a) and UACR regarding the risk of chronic kidney disease. *Baseline Lp(a) was categorized into low (<75 nmol/L) and high (≥75 nmol/L) groups. Baseline UACR within the normal range was classified into low-normal (0-9.9 mg/g) and high-normal (10-29.9 mg/g) groups. HR: hazard ratio; Lp(a): lipoprotein(a); UACR: urine albumin-to-creatinine ratio.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>No. of cases/total participants</th>
<th>HR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>By sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR*</td>
<td>1149/59,779</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>319/16,754</td>
<td>0.99 (0.86-1.14)</td>
<td>.87</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>1345/77,719</td>
<td>1.16 (1.04-1.28)</td>
<td>.01</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR*</td>
<td>350/20,805</td>
<td>1.26 (1.10-1.44)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>1278/63,144</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>401/23,063</td>
<td>0.98 (0.87-1.11)</td>
<td>.79</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>770/37,746</td>
<td>1.16 (1.05-1.28)</td>
<td>.01</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>246/10,345</td>
<td>1.36 (1.19-1.52)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>By age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;65 years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>1426/119,053</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>426/33,357</td>
<td>0.99 (0.87-1.12)</td>
<td>.87</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>1156/92,374</td>
<td>1.25 (1.14-1.37)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>373/25,121</td>
<td>1.50 (1.32-1.72)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≥65 years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>1101/23,870</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>254/64,466</td>
<td>0.98 (0.82-1.10)</td>
<td>.90</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>986/23,091</td>
<td>1.10 (0.99-1.21)</td>
<td>.08</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>268/60,896</td>
<td>1.14 (0.97-1.33)</td>
<td>.10</td>
</tr>
<tr>
<td>By diabetes mellitus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With diabetes mellitus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>505/10,503</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>146/29,61</td>
<td>1.03 (0.83-1.26)</td>
<td>.92</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>476/92,16</td>
<td>1.15 (1.04-1.28)</td>
<td>.02</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>147/24,57</td>
<td>1.57 (1.28-1.94)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Without diabetes mellitus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>1767/114,922</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>484/31,922</td>
<td>0.95 (0.86-1.06)</td>
<td>.35</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>1429/91,894</td>
<td>1.20 (1.11-1.30)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>421/24,836</td>
<td>1.26 (1.14-1.43)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>By hypertension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With hypertension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>1524/76,444</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>55/12,405</td>
<td>0.97 (0.87-1.08)</td>
<td>.83</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>1708/85,008</td>
<td>1.12 (1.02-1.23)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>517/17,065</td>
<td>1.37 (1.23-1.53)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Without hypertension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>602/66,479</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>156/18,352</td>
<td>0.96 (0.80-1.17)</td>
<td>.67</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>407/49,457</td>
<td>1.16 (0.99-1.35)</td>
<td>.07</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>124/13,225</td>
<td>1.20 (0.95-1.51)</td>
<td>.14</td>
</tr>
<tr>
<td>By high cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With high cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>1033/93,996</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>307/7444</td>
<td>0.95 (0.80-1.10)</td>
<td>.29</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>862/19,123</td>
<td>1.13 (1.01-1.25)</td>
<td>.03</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>261/6602</td>
<td>1.24 (1.05-1.45)</td>
<td>.01</td>
</tr>
<tr>
<td>Without high cholesterol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Lp(a) and low-normal UACR</td>
<td>1494/116,927</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High Lp(a) and low-normal UACR</td>
<td>413/32,433</td>
<td>1.01 (0.89-1.14)</td>
<td>.90</td>
</tr>
<tr>
<td>Low Lp(a) and high-normal UACR</td>
<td>1253/96,342</td>
<td>1.26 (1.15-1.37)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High Lp(a) and high-normal UACR</td>
<td>380/25,607</td>
<td>1.44 (1.26-1.64)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Figure 3. Further analyses of joint associations between baseline Lp(a) and UACR regarding the risk of CKD. *Baseline Lp(a) was categorized into low (<75 nmol/L) and high (≥75 nmol/L) groups. Baseline UACR within the normal range was classified into low-normal (0-9.9 mg/g) and high-normal (10-29.9 mg/g) groups. ACEI: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; CKD: chronic kidney disease; DM: diabetes mellitus; Lp(a): lipoprotein(a); UACR: urine albumin-to-creatinine ratio.

Table S9 in Multimedia Appendix 1 shows results for secondary outcomes (5615 with CKD stage 3, 393 with CKD stage 4, and 252 with CKD stage 5 and ESRD), with similar findings to the primary outcome in general.

Discussion

Principal Findings

In this study based on data from a prospective cohort, our principal findings are as follows: (1) high Lp(a) was nonsignificantly associated with increased CKD risk among all participants; (2) there was a significant interaction between Lp(a) and UACR but not between Lp(a) and eGFR; (3) when taking low Lp(a) and low-normal UACR as the reference, the high-normal UACR groups with any Lp(a) level were significantly associated with an increased risk of CKD, with no significant risks observed in the high Lp(a) and low-normal UACR group; (4) among those with high-normal UACR, high Lp(a) was associated with a significant increase in CKD risk; and (5) when treating UACR as a continuous variable, increased HRs regarding the relationship between Lp(a) and CKD risk were observed as UACR elevated.

Our study shows that the association between Lp(a) and CKD risk was nonsignificant among the general population without previous CKD in both categorical and continuous forms. Indeed, the relationship between Lp(a) and CKD risk has been explored, with inconsistent findings reported. One Chinese cohort study, including 6257 adults, showed that elevated Lp(a) was significantly associated with an increased risk of reduced renal function [10]. Nevertheless, the DiaGene study, including participants with type 2 DM, found that neither the high baseline Lp(a) group (also defined as ≥75 nmol/L) nor 2 related Lp(a) single-nucleotide polymorphisms were significantly related to the risk of incident nephropathy [9]. Inconsistent findings have also been reported in several Mendelian randomization studies. For instance, the study by Zheng et al [14] supported the causal role of Lp(a) in CKD development, while another study did not detect a significant association between Lp(a) and the risk of nephropathy [13]. These discrepant findings may partly be due to the heterogeneity of the target populations, outcome definitions, and statistical analyses. Although some previous studies adjusted for renal function indicators as covariates in their models [7-10,12], the relationship between Lp(a) and CKD risk was indeed assessed based on the average level of renal function in their populations. Thus, previous studies found different associations between Lp(a) and the risk of CKD, probably depending on the various average levels of renal function at baseline.

By contrast, we found a significant interaction between Lp(a) and UACR, and more specifically, a synergistic effect between Lp(a) and UACR on the risk of CKD. Interestingly, when the UACR value exceeded 10 mg/g approximately, the relationship between high Lp(a) and CKD risk became significant (Figure 1 and Table S2 in Multimedia Appendix 1), supporting the modification of UACR to the association between Lp(a) and CKD risk. Of note, Lp(a) together with other lipid and lipoprotein concentrations such as LDL particles had been reported to associate with UACR elsewhere, even among participants with a normal UACR range [29-31]. Therefore, it was plausible that UACR could act downstream on the pathway from Lp(a) to CKD. To address this concern, we performed a post hoc mediation analysis using a generalized linear model with 100 bootstrapping times [32,33]. The total effect of Lp(a) on CKD risk was 0.199 (95% CI 0.007-0.399), with the average direct effect of 0.195 (95% CI 0.008-0.373) and the average causal mediation effect through UACR of 0.004 (95% CI –0.005 to 0.014), indicating a nonsignificant mediation effect. This could further support the joint relationship between Lp(a) and UACR regarding CKD risk from a public health perspective.

Nevertheless, why the relationship between Lp(a) and CKD risk depended on UACR remained largely unknown. A possible
mechanism might be that elevated Lp(a) could be linked to increased synthesis of proteins in the liver due to proteinuria [34], where high UACR could be closely associated with proteinuria. There appeared to be an interaction between Lp(a) and hypertension regarding CVD outcomes [35], where hypertension had been reported to be significantly associated with elevated normal UACR [36]. Nevertheless, no significant subgroup effect by the existence of hypertension was found in our subgroup analysis (Figure 2). Of note, evidence from basic science might help explain the interaction. Lp(a) could enhance the expression of adhesion molecules in endothelial cells and aggravate normal endothelial function [37,38]. UACR serves as a common marker of endothelial and kidney function [39], and is reported to be an earlier and greater marker for some kidney outcomes than eGFR [40-42]; therefore, the detrimental effect of Lp(a) on the progression of CKD may be only significantly observed in those with high-normal UACR who had impaired renal function of clearance and self-recovery. However, the underlying mechanisms of the effects of Lp(a) and UACR on CKD risk have not been extensively elucidated, necessitating additional population studies for further exploration and clarification.

In participants aged over 65 years, high-normal UACR groups showed a nonsignificant association with CKD risk (Figure 2), consistent with findings from a large collaborative meta-analysis [19]. Moreover, the relationship between Lp(a) and CKD risk in patients with DM has been extensively investigated, yielding inconsistent results [11]. In our study, among participants with a history of DM, a significant association between Lp(a) and CKD risk was observed in the high-normal UACR group, but not in the low-normal UACR group (Figure 2). Hence, once more, examining the interaction between Lp(a) and UACR concerning CKD risk in participants with DM could contribute to understanding the prior inconclusive results in public health studies, where diverse baseline renal function existed across populations. Nonetheless, it is crucial to approach the results of our subgroup analyses with caution, considering their exploratory nature and their role in generating hypotheses.

CKD has emerged as a significant public health concern, characterized by a widespread prevalence and a substantial global disease burden [43]. The imperative now is to urgently improve early detection and preventive measures for CKD, given the considerable costs associated with therapy and the elevated mortality rates linked to advanced CKD stages [43]. Providing clarity on the role of Lp(a) could contribute significantly to public initiatives focused on preventing and managing the risks associated with CKD. Nevertheless, it is noteworthy that Lp(a) remains unaddressed in existing public health policies or guidelines for CKD [3,15]. It is widely acknowledged that serum Lp(a) levels are predominantly genetically determined, showing no significant associations with environmental or lifestyle factors [44,45]. While certain medications, such as muvalaplin, have shown effectiveness in lowering Lp(a) levels, their safety, tolerability, and cost-effectiveness still lack comprehensive clarity [46-48]. Addressing these uncertainties calls for more extensive and prolonged clinical trials in the future. Hence, there is a particular need for public health interventions, focusing on improving the management of both Lp(a) and UACR, as well as enhancing the accessibility of kidney health therapies. Furthermore, there is a necessity for quantitative assessment to determine the CKD risk attributed to Lp(a) in diverse populations. Such evaluations could strengthen the case for considering Lp(a) in the screening and management protocols for CKD. Similarly, should our findings be externally validated, the incorporation of Lp(a) into strategies could aid in targeting populations at a heightened risk of CKD, thereby bolstering efforts in CKD prevention.

**Strengths**

Our study carries several strengths. First, we leveraged data from a nationwide cohort, providing a substantial amount of information for our analyses. The application of rigorous methodology underpins the validity and robustness of our results. Notably, this study represents the first attempt to explore the connection between Lp(a), UACR, and the risk of CKD. By doing so, we aimed to elucidate the intricate relationship between Lp(a) and CKD risk within the general population, offering valuable insights for CKD risk assessment and prevention. This perspective, grounded in public health, sheds light on the role of Lp(a) and a renal function indicator in shaping strategies for CKD prevention.

**Limitations**

Several limitations should be acknowledged. First, owing to the observational design of our study, it is crucial to recognize that potential bias or confounding effects could not be entirely mitigated, and establishing a causal relationship between Lp(a) and CKD risk was beyond the scope of our investigation. Second, because both Lp(a) and UACR data were gathered at baseline, our analyses could not explore whether Lp(a) serves as a marker for UACR in the context of CKD risk. In addition, our study lacked the capacity to analyze changes in Lp(a) and UACR concerning the risk of CKD. Furthermore, Lp(a) measurements were conducted using a widely available immunoassay method, potentially introducing measurement errors due to the heterogeneity of isofrom size when compared with the gold-standard method used by the Northwest Lipid Metabolism Diabetes Research Laboratory [49]. Similarly, there might be measurement errors for UACR, calculated as the ratio between urinary albumin and creatinine. The assays used for urinary albumin and creatinine involved an immunoturbidimetric method and enzymatic analysis, respectively [50]. As CKD outcomes were determined using records from hospital in-patient data, cancer registries, and death registries, there is a possibility that CKD events were inadequately estimated due to underdiagnosis, misdiagnosis, or incorrect coding, the extent of which remains unknown. In addition, it is important to note that the UK Biobank, from which our participants were drawn, had over 90% (309,923/329,415, 94.08%) of individuals identifying as White. As a result, the generalizability of our study results to other racial groups may be compromised.

**Conclusions**

A notable interaction was identified between Lp(a) and UACR concerning the risk of CKD. This implies that Lp(a) may serve as a risk factor for CKD even when considering the influence of UACR. Our results offer valuable insights into the assessment

https://publichealth.jmir.org/2024/1/e50415
and prevention of CKD, emphasizing the combined role of Lp(a) and UACR from a public health perspective within the general population. This perspective can contribute to enhancing public awareness regarding the management of Lp(a) for the prevention of CKD.

Acknowledgments
This research was conducted using the UK Biobank Resource under Application Number 63844. We thank Drs Jingyi Zhang (Guangdong Second Provincial General Hospital, Guangzhou, China), Harriette Van Spall, and Hertzel Gerstein (McMaster University, Hamilton, Canada) for their help with the manuscript. This research was supported by the Science Foundation of Guangdong Second Provincial General Hospital (number: YY2018-002, recipient: GL).

Data Availability Statement
The data can be obtained by sending an application to the UK Biobank [22].

Authors’ Contributions
YL and GL participated in the design and conception of the study. YL and GL were responsible for data collection. YL performed the data analysis, table designs, interpretation, and drafted the manuscript. RW, SL, CZ, GYHL, LT, and GL provided professional support and revised the manuscript critically. All authors read and gave their approval of the final version. GL is the study guarantor.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary materials for Relationship Between Lipoprotein(a), Renal Function Indicators, and Chronic Kidney Disease: Evidence From a Large Prospective Cohort Study.

Multimedia Appendix 2
The Strengthening the Reporting of Observational studies in Epidemiology (STROBE) checklist.

References


22. UK Biobank. URL: https://www.ukbiobank.ac.uk [accessed 2023-06-30]


Abbreviations

CKD: chronic kidney disease
CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration
CRP: C-reactive protein
CVD: cardiovascular disease
DM: diabetes mellitus
eGFR: estimated glomerular filtration rate
ESRD: end-stage renal disease
HbA1c: glycated hemoglobin
HDL: high-density lipoprotein
HR: hazard ratio
ICD-10: 10th revision of the International Statistical Classification of Diseases and Related Health Problems
LDL: low-density lipoprotein
Lp(a): lipoprotein(a)
TDI: Townsend Deprivation Index
TG: triglyceride
UACR: urine albumin-to-creatinine ratio
Impact of the COVID-19 Pandemic on the Personal Networks and Neurological Outcomes of People With Multiple Sclerosis: Cross-Sectional and Longitudinal Case-Control Study

Claire Riley1*, MD; Shruthi Venkatesh2*, BS; Amar Dhand3, MD, DPhil; Nandini Doshi2, BS; Katelyn Kavak4, MS; Elle Levit5, MD; Christopher Perrone6, MD; Bianca Weinstock-Guttman4, MD; Erin Longbrake5, MD, PhD; Philip De Jager1, MD, PhD; Zongqi Xia2, MD, PhD

1Columbia University Irving Medical Center, New York, NY, United States
2University of Pittsburgh, Pittsburgh, PA, United States
3Brigham and Women's Hospital, Boston, MA, United States
4SUNY at Buffalo, Buffalo, NY, United States
5Yale University, New Haven, CT, United States
6University of Pennsylvania, Philadelphia, PA, United States
*these authors contributed equally

Corresponding Author:
Zongqi Xia, MD, PhD
University of Pittsburgh
3501 Fifth Ave
BST3-10.044
Pittsburgh, PA, 15260
United States
Phone: 1 412 383 5377
Email: zxia1@post.harvard.edu

Abstract

Background: The coronavirus disease 2019 (COVID-19) pandemic has negatively affected the social fabric.

Objective: We evaluated the associations between personal social networks and neurological function in people with multiple sclerosis (pwMS) and controls in the prepandemic and pandemic periods.

Methods: During the early pandemic (March-December 2020), 8 cohorts of pwMS and controls completed a questionnaire quantifying the structure and composition of their personal social networks, including the health behaviors of network members. Participants from 3 of the 8 cohorts had additionally completed the questionnaire before the pandemic (2017-2019). We assessed neurological function using 3 interrelated patient-reported outcomes: Patient Determined Disease Steps (PDDS), Multiple Sclerosis Rating Scale-Revised (MSRS-R), and Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function. We identified the network features associated with neurological function using paired 2-tailed $t$ tests and covariate-adjusted regressions.

Results: In the cross-sectional analysis of the pandemic data from 1130 pwMS and 1250 controls during the pandemic, having a higher percentage of network members with a perceived negative health influence was associated with worse disability in pwMS (MSRS-R: $\beta=2.181$, 95% CI 1.082-3.279; $P<.001$) and poor physical function in controls (PROMIS Physical Function: $\beta=-5.707$, 95% CI $-7.405$ to $-4.010$; $P<.001$). In the longitudinal analysis of 230 pwMS and 136 controls, the networks of all participants contracted, given an increase in constraint (pwMS-prepandemic: mean 52.24, SD 15.81; pwMS-pandemic: mean 56.77, SD 18.91; $P=.006$. Controls-prepandemic: mean 48.07, SD 13.36; controls-pandemic: mean 53.99, SD 16.31; $P=.001$) and a decrease in network size (pwMS-prepandemic: mean 8.02, SD 5.70; pwMS-pandemic: mean 6.63, SD 4.16; $P=.003$. Controls-prepandemic: mean 8.18, SD 4.05; controls-pandemic: mean 6.44, SD 3.92; $P<.001$), effective size (pwMS-prepandemic: mean 3.30, SD 1.59; pwMS-pandemic: mean 2.90, SD 1.50; $P=.007$. Controls-prepandemic: mean 3.85, SD 1.56; controls-pandemic: mean 3.40, SD 1.55; $P=.01$), and maximum degree (pwMS-prepandemic: mean 4.78, SD 1.86; pwMS-pandemic: mean 4.32, SD 1.92; $P=.01$. Controls-prepandemic: mean 5.38, SD 1.94; controls-pandemic: mean 4.55, SD 2.06; $P<.001$). These network changes were not
associated with worsening function. The percentage of kin in the networks of pwMS increased (mean 46.06%, SD 30.34% to mean 54.36%, SD 30.16%; \( P = .003 \)) during the pandemic, a change that was not seen in controls.

**Conclusions:** Our findings suggest that high perceived negative health influence in the network was associated with worse function in all participants during the pandemic. The networks of all participants became tighter knit, and the percentage of kin in the networks of pwMS increased during the pandemic. Despite these perturbations in social connections, network changes from the pre-pandemic to the pandemic period were not associated with worsening function in all participants, suggesting possible resilience.

*(JMIR Public Health Surveill 2024;10:e45429) doi:10.2196/45429*

**KEYWORDS**
neurology; neurodegenerative disease; multiple sclerosis; personal networks; COVID-19

**Introduction**

**Background**

Multiple sclerosis (MS) is a chronic autoimmune disease affecting the central nervous system, leading to neurodegeneration and neurological disability [1,2]. Despite notable advancement in elucidating the factors influencing MS susceptibility, the mechanisms underlying disability accumulation are less well defined [3]. In addition to genetic predisposition, modifiable environmental factors such as personal networks play a key role in shaping health outcomes for people living with neurological diseases [4].

Personal social network features can affect the health outcomes of people living with neurological diseases, including stroke, traumatic brain injury, and MS [5-8]. In people with MS (pwMS), advantageous personal network structures (e.g., larger network size and more diffuse connections among network members) are associated with better language function and larger regional brain volume [9]. Conversely, adverse personal networks are associated with increased social isolation and loneliness, which may in turn exert direct biological effects by altering inflammation, recovery from injury, and resilience against neurodegeneration [10-18].

Our prior research implicates personal social networks as a potentially modifiable environmental contributor to neurological disability in pwMS [19,20]. Importantly, the vulnerability of pwMS to social isolation and loneliness exacerbated during the coronavirus disease 2019 (COVID-19) pandemic [21,22]. Physical distancing measures, stay-at-home orders, and travel restrictions imposed during the pandemic led to the contraction of personal social networks, increased social isolation and loneliness, excessive reliance on virtual communication, and alterations in the frequency and nature of social interactions [23-28]. Understanding how personal social network structure (e.g., network size and density) and composition (e.g., demographics and health behaviors of the network members) changed during the pandemic and assessing the impact of these changes on disability accumulation in MS could potentially inform novel interventions that may improve the quality of life and health outcomes for pwMS.

**Objectives**

In this study, we compared the personal social networks of pwMS and controls during the COVID-19 pandemic and identified network features associated with increased disability accumulation. Further, we examined changes in personal social networks due to the COVID-19 pandemic in a subset of individuals with prepandemic data.

**Methods**

**Study Cohort**

During the pandemic, the multicenter Multiple Sclerosis Resilience to COVID-19 (MSReCOV) Collaborative recruited new pwMS and healthy controls through the University of Pittsburgh Medical Center (UPMC), Columbia University Irving Medical Center (CUIMC), University of Buffalo Medical Center, University of Pennsylvania, and Yale University [29-32]. In addition, pwMS and controls were recruited from 2 existing clinic-based cohorts (one at UPMC and another at CUIMC) and from a national cohort of first-degree relatives of pwMS (Genes and Environment in Multiple Sclerosis [GEMS] study) [33-38]. The inclusion criteria were adults aged \( \geq \) 18 years either with or without a neurologist-confirmed diagnosis of MS.

For the cross-sectional analysis, we included pwMS and healthy controls from the MSReCOV study as well as 3 previously established cohorts (clinic cohorts at UPMC and CUIMC as well as the GEMS cohort). We deployed a modified personal network (PERSNET) questionnaire (Multimedia Appendix 1) between March 2020 and December 2020 to assess demographic, clinical, and personal social network features through a secure web-based platform (REDCap [Research Electronic Data Capture]) [39].

For the longitudinal analysis, we leveraged PERSNET data collected before the COVID-19 pandemic (2017-2019) from the 3 existing cohorts (clinic cohorts at UPMC and CUIMC as well as the GEMS cohort) for comparison with the early pandemic data in the same participants (Figure 1A) [20].
Figure 1. Overview of study design. (A) Schematic illustration of study populations and survey data collection using REDCap (Research Electronic Data Capture). (B) Personal social network features as exposure, as illustrated by a representative network of a hypothetical participant. (C) Patient-reported outcomes of neurological disability or physical function.

Ethical Considerations
The institutional review board of each enrolling site approved the study (UPMC: STUDY21100060 and STUDY20040160; CUIMC: AAAS9668; University of Buffalo Medical Center: MOD00008107; University of Pennsylvania: 843454; and Yale University: HIC 200027987). All participants provided informed consent. Participants completed the survey anonymously through a secure privacy-compliant platform [39]. Participation was voluntary. To ensure confidentiality, deidentified data were used for analysis. The study did not provide compensation for survey completion.

Personal Network Metrics
We deployed an updated version of the PERSNET survey, adapted from the foundational General Social Survey [5-7,19,20,40]. In the PERSNET survey, participants identified individuals in their personal social network with whom they discuss personal matters or socialize or from whom they derive social support. These individuals could have any professional or personal relationship with the participant (e.g., coworker, parent, sibling, spouse, or child). We assessed the structure and composition of each participant’s network [41,42].

Network structure includes 6 quantitative features: size (number of individuals in the network, excluding the index person), density (sum of ties, excluding the index person’s ties, divided by all possible ties), constraint (a more granular density that assesses the extent to which the index person is connected to individuals who are connected to one another), effective size (number of nonredundant network members), and maximum degree and mean degree (highest and average number of ties, respectively, belonging to a network member).

Network composition quantifies the demographic characteristics and health behaviors of network members. Network demographics include the percentage of kin (percentage of network members who are family), SD of age (age range of network members), diversity of sex (proportion of sexes from 0 to 1, where 0 indicates a single sex and 1 indicates an equal ratio of men and women), and diversity of race (similar proportion of represented races, where 0 indicates a single race). Network health behaviors include the percentage of network members who smoke, consume alcohol, exhibit poor dietary habits, lead sedentary lifestyles, and exert perceived negative health influence. Other network composition features include the frequency, duration, and the living distance of network member from the index person, which quantify the depth of the relationships. Compositional features account for network size.

Neurological Outcomes
To assess the status of neurological function, we used 3 interrelated patient-reported outcomes (PROs). The Patient Determined Disease Steps (PDDS) scale indicates the extent of gait impairment and correlates with the clinician-determined Expanded Disability Status Scale (EDSS). PDDS scores range from 0 to 8, where 0 corresponds to normal, and 8 indicates bedbound status [43]. The Multiple Sclerosis Rating Scale-Revised (MSRS-R) assesses the global neurological symptom burden, including walking, function in the extremities, vision, speech, swallowing, cognition, sensation, bladder function, and bowel function [44,45]. Each symptom domain score ranges from 0 to 4, where 0 indicates no impairment, and 4 indicates severe impairment. Higher cumulative MSRS-R scores (0-32) indicate greater neurological symptom burden and worse neurological function. Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function (version 1.2) is a generalizable measure of physical function also validated for pwMS [46,47]. PROMIS Physical Function is reported as a normally distributed T-score on a scale ranging from 0 to 100, where 50 represents the average for the US population and higher scores indicate better function. Whereas PDDS and MSRS-R are specific for pwMS, PROMIS is applicable to both pwMS and controls. All participants completed PRO assessment when completing the PERSNET survey during the pandemic, whereas a subset also completed the PROs before the pandemic.

Covariates
We considered the following confounding factors that could potentially influence neurological function: age, sex, race, ethnicity, disease duration, employment, education, occupation, income, marital status, and cohabitant status. Race was categorized as African or African American, American Indian or Alaska Native or Native Hawaiian or other Pacific Islander, Asian, White, multiracial, or other. Ethnicity was categorized as Hispanic (or Latinx) and non-Hispanic. Because of the
relatively small number of racial and ethnic minority participants in our cohorts, race and ethnicity were dichotomized as non-Hispanic White versus otherwise (encompassing individuals of Hispanic and non-European descent) in subsequent analyses. Disease duration was defined as the time from the first self-reported neurological symptom onset to the time of the most recent PRO assessment. Employment status was categorized as employed for wages, self-employed, out of work and looking for work, out of work but not currently looking for work, homemaker, student, military, retired, or unable to work. Education level was classified based on the highest level of education achieved: some high school or less, high school graduate, some college, associate degree, bachelor’s degree, or graduate degree. Occupation was categorized as business owner, executive or manager, professional, sales or clerical worker, service worker, or other. Annual household income level included the following brackets: ≤US $19,999, US $20,000 to US $34,999, US $35,000 to US $49,999, US $50,000 to US $64,999, US $65,000 to US $79,999, US $80,000 to US $94,999, US $95,000 to US $109,999, US $110,000 to US $124,999, and ≥US $125,000. Marital status was categorized as married or unmarried. Cohabitation status was categorized as living alone versus otherwise. In the subset with prepandemic data, the time elapsed between prepandemic and pandemic PERSNET assessments ranged from 1 to 4 years. Some covariates were missing because certain questions (e.g., employment status, income) in the PERSNET survey were made optional to reduce patient discomfort when completing the questionnaire.

To select the informative covariates, we examined the correlation between these features and PROs in univariate analyses (Figure S1 in Multimedia Appendix 2). We included features that met the following predefined criteria as covariates for downstream analyses: feature presence in >70% of pwMS, a Pearson correlation coefficient of ≥0.1, and a nominal statistical significance (P < .05) in association with all 3 PROs. We identified age, disease duration, employment, and income as meeting the criteria and adjusted these covariates in regression models involving pwMS. For analyses involving controls, we adjusted for age, employment, and income but not disease duration because it does not apply to controls. For the longitudinal analyses, we further adjusted for the time elapsed between prepandemic and pandemic PERSNET assessments as well as study cohort as additional covariates for consistency with our previous analysis [19,20].

Statistical Analysis
We performed two types of analyses: (1) cross-sectional comparison between pwMS and controls during the COVID-19 pandemic and (2) longitudinal analysis of pwMS and controls during the COVID-19 pandemic when compared to their prepandemic baseline.

For the cross-sectional analysis, we first compared the personal networks of pwMS and controls using paired 2-tailed t tests. Next, we examined the association between network features (structure and composition) and PROs in pwMS in covariate-adjusted regression models. Given that the PROMIS Physical Function measure is generalizable across health and disease, we further assessed the association between network features and PROMIS Physical Function scores in controls for comparison. In a joint analysis of the pandemic data that include both pwMS and controls, we performed a moderation analysis to assess whether having an MS diagnosis influenced the association between network features and PROMIS Physical Function scores [48-50].

For the longitudinal analysis, we examined the within-subject differences in network features in pwMS and controls during the pandemic when compared to the most proximal prepandemic baseline (i.e., the closest value before the pandemic as baseline) using paired t tests. Next, we assessed the association of change in network features (i.e., pandemic value minus prepandemic baseline) in relation to the latest available PROs (during the pandemic) in pwMS using covariate-adjusted regressions and an omnibus test [20,38]. For the omnibus test, we combined the P values derived from the covariate-adjusted regressions for each PRO. Using the Fisher combined probability test, we calculated the chi-squared statistic and compared the observed values with the expected empirical distribution. To further interrogate these relationships, we generated a quantile-quantile (Q-Q) plot of the observed versus expected P values of the associations between the longitudinal changes in each network feature (pandemic value minus prepandemic baseline) and each PRO (the latest available score during the pandemic). The 95% CIs of the Q-Q plot were obtained from the empirical P value distribution generated by 10,000 permutations of the null hypothesis. We used 10,000 permutations, given the large number of network features and the sample size [19,20]. We performed similar longitudinal analysis in controls using PROMIS Physical Function for comparison. Raw P values were adjusted by Bonferroni correction for multiple comparisons. All statistical analyses were performed using R software (version 3.6.0) [51].

Code Availability
The code for this project is available on GitHub [52].

Results
Participant Characteristics
The cross-sectional analysis included 1130 pwMS (age: mean 50.7, SD 12.1 y) and 1250 controls (age: mean 44.3, SD 12.1 y; Table 1). Participants were predominantly women (pwMS: 925/1130, 81.86%; controls: 960/1250, 76.80%) and non-Hispanic White (pwMS: 1043/1130, 92.30%; controls: 1208/1250, 96.64%), pwMS were less likely to be employed (544/1130, 48.14%) than controls (876/1250, 70.08%). The disability burden among pwMS was mild to moderate (PDDS score: mean 1.85, SD 2.12; MSRS-R score: mean 7.55, SD 5.49; PROMIS Physical Function score: mean 46.4, SD 10.82).
Table 1. Characteristics of the cross-sectional and longitudinal cohort participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Cross-sectional cohorts&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Longitudinal cohorts&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pwMS&lt;sup&gt;c&lt;/sup&gt; (N=1130)</td>
<td>Controls (N=1250)</td>
<td></td>
</tr>
<tr>
<td>Age (y), mean (SD)</td>
<td>50.7 (12.1)</td>
<td>44.4 (12.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td>.009</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>925 (81.9)</td>
<td>960 (76.8)</td>
<td></td>
</tr>
<tr>
<td>Man</td>
<td>203 (18.0)</td>
<td>286 (22.9)</td>
<td></td>
</tr>
<tr>
<td>Nonbinary intersex</td>
<td>2 (0.2)</td>
<td>4 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Race, n (%)</td>
<td>&lt;.001</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>African or African American</td>
<td>46 (4.1)</td>
<td>12 (1)</td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native or Native Hawaiian or other Pacific Islander</td>
<td>10 (0.9)</td>
<td>6 (0.5)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>8 (0.7)</td>
<td>8 (0.6)</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>White</td>
<td>1043 (92.7)</td>
<td>1208 (96.7)</td>
<td>214 (93.0)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>9 (0.8)</td>
<td>11 (0.9)</td>
<td>4 (1.7)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (0.4)</td>
<td>3 (0.2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Not sure</td>
<td>5 (0.4)</td>
<td>1 (0.1)</td>
<td>3 (1.3)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td>.26</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latinx</td>
<td>38 (3.4)</td>
<td>31 (2.5)</td>
<td>6 (2.6)</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>1068 (94.9)</td>
<td>1201 (96.3)</td>
<td>220 (95.7)</td>
</tr>
<tr>
<td>Not sure</td>
<td>19 (1.7)</td>
<td>15 (1.2)</td>
<td>4 (1.7)</td>
</tr>
<tr>
<td>Education, n (%)</td>
<td>&lt;.001</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>High school graduate</td>
<td>70 (6.2)</td>
<td>35 (2.8)</td>
<td>12 (5.3)</td>
</tr>
<tr>
<td>Some college</td>
<td>154 (13.7)</td>
<td>106 (8.5)</td>
<td>30 (13.3)</td>
</tr>
<tr>
<td>Associate degree</td>
<td>111 (9.9)</td>
<td>70 (5.6)</td>
<td>20 (8.8)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>363 (32.4)</td>
<td>441 (35.3)</td>
<td>75 (33.3)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>418 (37.3)</td>
<td>595 (47.6)</td>
<td>88 (39.1)</td>
</tr>
<tr>
<td>Employment&lt;sup&gt;e&lt;/sup&gt;, n (%)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Employed for wages</td>
<td>544 (48.1)</td>
<td>876 (70.1)</td>
<td>98 (43.9)</td>
</tr>
<tr>
<td>Homemaker</td>
<td>63 (5.6)</td>
<td>77 (6.2)</td>
<td>11 (4.9)</td>
</tr>
<tr>
<td>Out of work and looking for work</td>
<td>17 (1.5)</td>
<td>27 (2.2)</td>
<td>8 (3.5)</td>
</tr>
<tr>
<td>Out of work but not currently looking for work</td>
<td>29 (2.6)</td>
<td>15 (1.2)</td>
<td>6 (2.7)</td>
</tr>
<tr>
<td>Retired</td>
<td>187 (16.5)</td>
<td>98 (7.8)</td>
<td>34 (15)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>63 (5.6)</td>
<td>97 (7.8)</td>
<td>14 (6.2)</td>
</tr>
<tr>
<td>Student</td>
<td>12 (1.1)</td>
<td>18 (1.4)</td>
<td>3 (1.3)</td>
</tr>
<tr>
<td>Unable to work</td>
<td>190 (16.8)</td>
<td>39 (3.1)</td>
<td>49 (21.7)</td>
</tr>
<tr>
<td>Military</td>
<td>2 (0.2)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Annual household income (US $), n (%)</td>
<td>&lt;.001</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>≤19,999</td>
<td>74 (7.6)</td>
<td>37 (3.1)</td>
<td>19 (12.6)</td>
</tr>
<tr>
<td>20,000-34,999</td>
<td>85 (8.7)</td>
<td>56 (4.7)</td>
<td>15 (9.9)</td>
</tr>
<tr>
<td>35,000-49,999</td>
<td>97 (10)</td>
<td>75 (6.3)</td>
<td>23 (15.2)</td>
</tr>
<tr>
<td>50,000-64,999</td>
<td>92 (9.5)</td>
<td>106 (8.9)</td>
<td>14 (9.3)</td>
</tr>
</tbody>
</table>
The longitudinal analysis included 230 pwMS and 136 controls. Similar to the cross-sectional analysis, controls were younger (age: mean 50.4, SD 11.7 y for pwMS and mean 43.1, SD 12.6 y for controls), had higher levels of education (163/225, 72.4% of pwMS and 103/114, 90.4% of controls completed college) and employment (pwMS: 98/223, 43.9%; controls: 81/114, 71.1%), and had higher annual household income than pwMS.

### Cross-Sectional Analysis of the Pandemic Period

First, we compared the network features of pwMS and controls during the COVID-19 pandemic (Table 2). In this unadjusted analysis, pwMS had higher density, higher constraint, smaller effective size, a higher percentage of kin, a lower percentage of people known for <6 years, lower percentage of network members who live >15 miles (>24 km) away, and lower percentage of network members who drink in their social networks when compared to controls. pwMS and controls reported similarly high percentages of their network members with a perceived negative health influence (pwMS: mean 36.16%, SD 30.43%; controls: mean 36.34%, SD 30.16%).

### Table 2: Comparison of Demographics and Network Characteristics Between pwMS and Controls During the Pandemic Period

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Cross-sectional cohorts&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Longitudinal cohorts&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pwMS&lt;sup&gt;c&lt;/sup&gt; (N=1130)</td>
<td>Controls (N=1250)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65,000-79,999</td>
<td>79 (8.1)</td>
<td>120 (10.1)</td>
</tr>
<tr>
<td>80,000-94,999</td>
<td>76 (7.8)</td>
<td>94 (7.9)</td>
</tr>
<tr>
<td>95,000-109,999</td>
<td>80 (8.2)</td>
<td>127 (10.7)</td>
</tr>
<tr>
<td>110,000-124,999</td>
<td>85 (8.7)</td>
<td>112 (9.4)</td>
</tr>
<tr>
<td>≥125,000</td>
<td>303 (31.2)</td>
<td>461 (38.8)</td>
</tr>
<tr>
<td>Married, n (%)</td>
<td>786 (69.9)</td>
<td>841 (67.4)</td>
</tr>
<tr>
<td>Live alone, n (%)</td>
<td>156 (14.8)</td>
<td>186 (15)</td>
</tr>
<tr>
<td>Occupation, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business owner</td>
<td>35 (6.2)</td>
<td>23 (2.4)</td>
</tr>
<tr>
<td>Executive or manager</td>
<td>100 (17.6)</td>
<td>157 (16.2)</td>
</tr>
<tr>
<td>Laborer or unskilled worker</td>
<td>7 (1.2)</td>
<td>4 (0.4)</td>
</tr>
<tr>
<td>Machine operator, inspector, or bus or cab driver</td>
<td>1 (0.2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Mechanic, electrician, or skilled worker</td>
<td>6 (1.1)</td>
<td>14 (1.4)</td>
</tr>
<tr>
<td>Other</td>
<td>80 (14.1)</td>
<td>129 (13.3)</td>
</tr>
<tr>
<td>Professional</td>
<td>282 (49.6)</td>
<td>545 (56.1)</td>
</tr>
<tr>
<td>Sales or clerical worker</td>
<td>54 (9.5)</td>
<td>84 (8.6)</td>
</tr>
<tr>
<td>Service worker</td>
<td>3 (0.5)</td>
<td>16 (1.6)</td>
</tr>
<tr>
<td>PDDS&lt;sup&gt;f&lt;/sup&gt; score, mean (SD)</td>
<td>1.9 (2.1)</td>
<td>N/A&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>MSRS-R&lt;sup&gt;h&lt;/sup&gt; score, mean (SD)</td>
<td>7.6 (5.5)</td>
<td>N/A</td>
</tr>
<tr>
<td>PROMIS&lt;sup&gt;i&lt;/sup&gt; Physical Function T-score, mean (SD)</td>
<td>46.4 (10.8)</td>
<td>56.3 (9.0)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Cross-sectional cohorts include cohorts 1 and 2 (University of Pittsburgh Medical Center: clinic-based cohort and Multiple Sclerosis Resilience to COVID-19 [MSReCOV] collaborative recruitment), cohorts 3 and 4 (Columbia University Irving Medical Center: clinic-based cohort and MSReCOV collaborative recruitment), cohort 5 (Yale University), cohort 6 (University of Buffalo Medical Center), cohort 7 (University of Pennsylvania), and cohort 8 (Genes and Environment in Multiple Sclerosis [GEMS] cohort).

<sup>b</sup>Longitudinal cohorts include cohort 1 (University of Pittsburgh Medical Center: clinic-based cohort), cohort 3 (Columbia University Irving Medical Center: clinic-based cohort), and cohort 8 (GEMS cohort).

<sup>c</sup>PwMS: people with multiple sclerosis.

<sup>d</sup>Italized P values met the significance threshold (P<.05).

<sup>e</sup>The sample size for certain questionnaire response may be smaller than the overall cohort size due to the optional completion of certain questions (e.g., employment status) to minimize participant discomfort.

<sup>f</sup>PDDS: Patient Determined Disease Steps.

<sup>g</sup>N/A: not applicable.

<sup>h</sup>MSRS-R: Multiple Sclerosis Rating Scale-Revised.

<sup>i</sup>PROMIS: Patient-Reported Outcomes Measurement Information System.
Table 2. Comparison of personal social network features in people with multiple sclerosis (MS) and controls during the COVID-19 pandemic.

<table>
<thead>
<tr>
<th>Features</th>
<th>pwMS a (N=1130), mean (SD)</th>
<th>Controls (N=1250), mean (SD)</th>
<th>P value b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network structure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network size</td>
<td>6.79 (4.24)</td>
<td>6.78 (3.90)</td>
<td>.95</td>
</tr>
<tr>
<td>Density</td>
<td>0.76 (0.25)</td>
<td>0.72 (0.24)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Constraint</td>
<td>56.88 (19.07)</td>
<td>54.30 (17.03)</td>
<td>.001</td>
</tr>
<tr>
<td>Effective size</td>
<td>2.97 (1.62)</td>
<td>3.38 (1.65)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Maximum degree</td>
<td>4.38 (2.07)</td>
<td>4.54 (1.99)</td>
<td>.06</td>
</tr>
<tr>
<td>Mean degree</td>
<td>3.48 (1.77)</td>
<td>3.43 (1.59)</td>
<td>.49</td>
</tr>
<tr>
<td><strong>Network composition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of kin</td>
<td>55.89 (29.50)</td>
<td>50.71 (28.84)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SD of age</td>
<td>12.65 (6.80)</td>
<td>11.96 (5.82)</td>
<td>.03</td>
</tr>
<tr>
<td>Diversity of sex</td>
<td>67.26 (39.72)</td>
<td>70.80 (30.26)</td>
<td>.02</td>
</tr>
<tr>
<td>Diversity of race</td>
<td>6.00 (16.42)</td>
<td>7.32 (17.03)</td>
<td>.06</td>
</tr>
<tr>
<td>Percentage of network members contacted weekly or less</td>
<td>15.89 (20.90)</td>
<td>17.67 (21.15)</td>
<td>.04</td>
</tr>
<tr>
<td>Percentage of network members known for &lt;6 years</td>
<td>11.81 (19.61)</td>
<td>17.57 (23.65)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Percentage of network members who live &gt;15 miles (&gt;24 km) away</td>
<td>32.45 (27.82)</td>
<td>38.93 (26.63)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Percentage of network members who drink</td>
<td>12.43 (24.03)</td>
<td>16.63 (26.55)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Percentage of network members who smoke</td>
<td>8.87 (18.14)</td>
<td>7.17 (15.85)</td>
<td>.02</td>
</tr>
<tr>
<td>Percentage of network members who are nonexercisers</td>
<td>35.28 (31.60)</td>
<td>39.06 (30.43)</td>
<td>.004</td>
</tr>
<tr>
<td>Percentage of network members who have a bad diet</td>
<td>23.04 (27.61)</td>
<td>22.75 (28.08)</td>
<td>.81</td>
</tr>
<tr>
<td>Percentage of network members with a negative health influence</td>
<td>36.16 (30.43)</td>
<td>36.34 (30.16)</td>
<td>.89</td>
</tr>
</tbody>
</table>

a pwMS: people with multiple sclerosis.

b Italicized P values met the significance threshold (P<.002; α=.05, corrected for 18 comparisons).

Next, we examined the association between each structural and compositional network feature in relation to PDDS, MSRS-R, and PROMIS Physical Function scores in pwMS during the pandemic, after adjusting for age, disease duration, employment, and income in linear regression models (Table 3). pwMS who had a higher percentage of network members with a perceived negative health influence had higher MSRS-R scores, indicating greater MS symptom burden and disability (β=2.181, 95% CI 1.082-3.279; P<.001). None of the specific network health behavior features (e.g., smoking, alcohol use, bad diet, and sedentary lifestyle) that might be construed as a negative health influence showed a statistically significant association with MSRS-R scores. No other network feature had a significant association with any of the PROs after Bonferroni correction.
Table 3. Cross-sectional analysis of personal social network features in relation to patient-reported outcomes in people with multiple sclerosis during the COVID-19 pandemic.

<table>
<thead>
<tr>
<th>Features</th>
<th>PDSSa</th>
<th>Sample size, n</th>
<th>βc (95% CI)</th>
<th>P valuec</th>
<th>MSRS-Rb</th>
<th>Sample size, n</th>
<th>βc (95% CI)</th>
<th>P valuec</th>
<th>PROMIS Physical Function</th>
<th>Sample size, n</th>
<th>βc (95% CI)</th>
<th>P valuec</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>737</td>
<td>−0.017 (−0.050 to 0.016)</td>
<td>.31</td>
<td>.46</td>
<td>728</td>
<td>0.165 (0.001 to 0.330)</td>
<td>.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>704</td>
<td>−0.051 (−0.569 to 0.467)</td>
<td>.85</td>
<td>.87</td>
<td>695</td>
<td>−0.369 (−2.933 to 2.194)</td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constraint</td>
<td>704</td>
<td>−0.001 (−0.009 to 0.006)</td>
<td>.68</td>
<td>.63</td>
<td>695</td>
<td>−0.006 (−0.041 to 0.029)</td>
<td>.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effective size</td>
<td>704</td>
<td>0.006 (−0.079 to 0.091)</td>
<td>.90</td>
<td>.69</td>
<td>695</td>
<td>0.150 (−0.270 to 0.569)</td>
<td>.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum degree</td>
<td>704</td>
<td>0.018 (−0.050 to 0.085)</td>
<td>.60</td>
<td>.82</td>
<td>695</td>
<td>0.017 (−0.316 to 0.350)</td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean degree</td>
<td>704</td>
<td>0.012 (−0.066 to 0.089)</td>
<td>.77</td>
<td>.44</td>
<td>695</td>
<td>−0.049 (−0.433 to 0.335)</td>
<td>.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Network composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of kin</td>
<td>713</td>
<td>0.191 (−0.250 to 0.632)</td>
<td>.40</td>
<td>.29</td>
<td>704</td>
<td>−1.356 (−3.541 to 0.830)</td>
<td>.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD of age</td>
<td>534</td>
<td>−0.010 (−0.032 to 0.012)</td>
<td>.36</td>
<td>.66</td>
<td>526</td>
<td>0.025 (−0.082 to 0.132)</td>
<td>.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversity of sex</td>
<td>707</td>
<td>0.148 (−0.277 to 0.574)</td>
<td>.49</td>
<td>.66</td>
<td>698</td>
<td>−1.143 (−3.249 to 0.962)</td>
<td>.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversity of race</td>
<td>701</td>
<td>0.545 (−0.403 to 1.492)</td>
<td>.26</td>
<td>.39</td>
<td>692</td>
<td>−1.506 (−6.195 to 3.183)</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of network members contacted weekly or less</td>
<td>713</td>
<td>0.394 (−0.245 to 1.034)</td>
<td>.23</td>
<td>.47</td>
<td>704</td>
<td>−1.130 (−4.306 to 2.046)</td>
<td>.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of network members known for &lt;6 years</td>
<td>713</td>
<td>−0.215 (−0.909 to 0.478)</td>
<td>.54</td>
<td>.96</td>
<td>704</td>
<td>1.326 (−2.11 to 4.763)</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of network members who live &gt;15 miles (&gt;24 km) away</td>
<td>713</td>
<td>−0.199 (−0.667 to 0.269)</td>
<td>.40</td>
<td>.77</td>
<td>704</td>
<td>2.306 (−0.008 to 4.619)</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of network members who drink</td>
<td>713</td>
<td>−0.326 (−0.920 to 0.268)</td>
<td>.28</td>
<td>.61</td>
<td>704</td>
<td>2.979 (−0.003 to 5.960)</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of network members who smoke</td>
<td>713</td>
<td>−0.139 (−0.882 to 0.603)</td>
<td>.71</td>
<td>.28</td>
<td>704</td>
<td>−2.155 (−5.827 to 1.518)</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of network members who are nonexercisers</td>
<td>713</td>
<td>−0.069 (−0.486 to 0.348)</td>
<td>.75</td>
<td>.15</td>
<td>704</td>
<td>−0.281 (−2.351 to 1.788)</td>
<td>.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of network members who have a bad diet</td>
<td>713</td>
<td>0.048 (−0.437 to 0.534)</td>
<td>.85</td>
<td>.049</td>
<td>704</td>
<td>2.794 (−5.189 to −0.400)</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of network members with a negative health influence</td>
<td>713</td>
<td>−0.281 (−0.714 to 0.153)</td>
<td>.20</td>
<td>.049</td>
<td>704</td>
<td>−0.650 (−2.795 to 1.494)</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aPDSS: Patient Determined Disease Steps.
bMSRS-R: Multiple Sclerosis Rating Scale-Revised.
cPROMIS: Patient-Reported Outcomes Measurement Information System.
dThe smaller sample size in this analysis when compared to the overall cohort size was due to the requirement for participants who had complete data elements for network feature, neurological outcome, and all covariates.
Adjusted for potential confounders (as identified in Figure S1 in Multimedia Appendix 2), including age, disease duration, employment, and income.

Italicized $P$ value met the significance threshold ($P<.00092; \alpha=.05$, corrected for 54 comparisons).

For comparison, we examined the association between each network feature and PROMIS Physical Function scores during the pandemic in controls (Table S2 in Multimedia Appendix 2). We adjusted for age, employment, and income, whereas disease duration was not applicable for controls. We found that a lower maximum degree ($\beta=0.383$, 95% CI 0.141-0.626; $P=.002$), a higher percentage of network members who smoke ($\beta=-4.846$, 95% CI $-7.919$ to $-1.774$; $P=.002$), and a higher percentage of network members with a perceived negative health influence ($\beta=-5.707$, 95% CI $-7.405$ to $-4.010$; $P<.001$) were associated with lower PROMIS Physical Function scores (worse physical function) in controls. Given that the percentage of network members with a perceived negative health influence was associated with MSRS-R scores in pwMS and with PROMIS Physical Function scores in controls, respectively, this important compositional network feature that assesses the perception of negative health influences in the personal social network may contribute to physical impairment in both pwMS and controls.

Moderation Analysis Assessing the Role of Having an MS Diagnosis

We performed a moderation analysis to examine whether having an MS diagnosis influences the strength and direction of the association between network features and PROMIS Physical Function score during the COVID-19 pandemic (Figure 2A, Table S4 in Multimedia Appendix 2). In this joint analysis combining pwMS and controls, an MS diagnosis moderated the direction and strength of the association between several network features (diversity of race, percentage of network members who live $>15$ miles [$>24$ km] away, percentage of network members who drink, and percentage of network members with a perceived negative health influence) and PROMIS Physical Function score. For diversity of race, the moderating effect of an MS diagnosis was marginally significant ($\beta=-5.878$, 95% CI $-10.889$ to $-0.868$; $P=.02$) such that this network feature was not significantly associated with physical function in either pwMS (slope=$-2.968$, 95% CI $-6.893$ to 0.956; $P=.14$) or controls (slope=$2.910$, 95% CI $-0.224$ to 6.044; $P=.07$; Figure 2B). Having a higher percentage of network members who live $>15$ miles ($>24$ km) away was associated with worse physical function in pwMS (slope=$3.890$, 95% CI 1.852-5.928; $P<.001$) but not in controls (slope=$-0.772$, 95% CI $-2.765$ to 1.221; $P=.45$; Figure 2C), given the moderating effect of an MS diagnosis. Similarly, having a higher percentage of network members who drink was associated with worse physical function in pwMS (slope=$3.716$, 95% CI 1.392-6.039; $P=.002$) but again not in controls (slope=$-0.982$, 95% CI $-2.966$ to 1.003; $P=.33$; Figure 2D). For percentage of network members with a perceived negative health influence, both groups exhibited the same direction of association with physical function, but pwMS had a smaller magnitude of the association (slope=$-2.174$, 95% CI $-4.019$ to $-0.329$; $P=.02$) than controls (slope=$-6.601$, 95% CI $-8.494$ to $-4.708$; $P<.001$; Figure 2E).
Figure 2. Role of multiple sclerosis (MS) diagnosis in moderating the association between personal social network features and Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function scores in people with MS (pwMS) and controls. (A) In a moderation analysis, we examined the association between each network feature and PROMIS Physical Function after combining the data from pwMS and controls and after adjusting for age, employment, and income as covariates and further investigated whether having an MS diagnosis moderated the direction or strength of this association. An MS diagnosis moderated the direction of the association between (B) diversity of race, (C) percentage who live >15 miles (>24 km) away, and (D) percentage who drink and PROMIS Physical Function, as well as the strength of the association between (E) percentage of people with a perceived negative health influence and PROMIS Physical Function. Values on x-axis indicate the 25th, 50th, 75th percentile.
Longitudinal Analysis Comparing the Pandemic Period With the Prepandemic Baseline

We conducted a longitudinal analysis using a subset of pwMS and controls who completed the PERSNET survey both before and during the pandemic. During the pandemic, both pwMS and controls reported a lower percentage of people contacted weekly when compared to the prepandemic baseline (decrease from mean 23.48%, SD 24.95% to mean 14.89%, SD 24.95% for pwMS, \( P < .001 \); decrease from mean 30.34%, SD 26.35% to mean 18.78%, SD 22.2% for controls, \( P < .001 \)), reflecting the widespread social isolation during the pandemic (Table 4). When compared to the prepandemic baseline in both pwMS and controls, there was a reduction in network size (pwMS-prepandemic: mean 8.02, SD 5.70; pwMS-pandemic: mean 6.63, SD 4.16; \( P = .003 \); Controls-prepandemic: mean 8.18, SD 4.05; controls-pandemic: mean 6.44, SD 3.92; \( P < .001 \)), effective size (pwMS-prepandemic: mean 3.30, SD 1.59; pwMS-pandemic: mean 2.90, SD 1.50; \( P = .007 \); Controls-prepandemic: mean 3.85, SD 1.56; controls-pandemic: mean 3.40, SD 1.55; \( P = .01 \)), and maximum degree (pwMS-prepandemic: mean 4.78, SD 1.86; pwMS-pandemic: mean 4.32, SD 1.92; \( P = .01 \)). Controls-prepandemic: mean 5.38, SD 1.94; controls-pandemic: mean 4.55, SD 2.06; \( P < .001 \)), as well as an increase in constraint (pwMS-prepandemic: mean 52.24, SD 15.81; pwMS-pandemic: mean 56.77, SD 18.91; \( P = .006 \); Controls-prepandemic: mean 48.07, SD 13.36; controls-pandemic: mean 53.99, SD 16.31; \( P = .001 \)). These findings indicate contraction in personal social networks for both pwMS and controls during the pandemic period. There was an increase in the percentage of kin (from mean 46.06%, SD 29.34% to mean 54.36%, SD 30.16%; \( P = .003 \)) in the networks of pwMS during the pandemic, which was not seen in controls.

Finally, we examined whether changes in network features due to the pandemic (i.e., pandemic value minus the most proximal prepandemic baseline) in pwMS were associated with the PROs during the pandemic (Figure 3). We found no significant association between changes in network features and the latest available pandemic PROs. As a confirmation of these findings, there was no difference between the observed and expected distribution of the \( P \) values of association between changes in network features and pandemic PROs in the permuted omnibus test (PDDS: \( P = .88 \), MSRS-R: \( P = .29 \), and PROMIS Physical Function: \( P = .28 \)).
Table 4. Personal social network features in people with multiple sclerosis (MS) and controls during the COVID-19 pandemic when compared to the within-subject prepandemic baseline.

<table>
<thead>
<tr>
<th>Features</th>
<th>Controls: prepandemic baseline (N=136), mean (SD)</th>
<th>Controls: during the pandemic (N=136), mean (SD)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>pwMS&lt;sup&gt;b&lt;/sup&gt;: prepandemic baseline (N=230), mean (SD)</th>
<th>pwMS: during the pandemic (N=230), mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Network structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network size</td>
<td>8.18 (4.05)</td>
<td>6.44 (3.92)</td>
<td>&lt;.001</td>
<td>8.02 (5.70)</td>
<td>6.63 (4.16)</td>
<td>.003</td>
</tr>
<tr>
<td>Density</td>
<td>0.69 (0.24)</td>
<td>0.72 (0.24)</td>
<td>.24</td>
<td>0.74 (0.25)</td>
<td>0.77 (0.24)</td>
<td>.14</td>
</tr>
<tr>
<td>Constraint</td>
<td>48.07 (13.36)</td>
<td>53.99 (16.31)</td>
<td>.001</td>
<td>52.24 (15.81)</td>
<td>56.77 (18.91)</td>
<td>.006</td>
</tr>
<tr>
<td>Effective size</td>
<td>3.85 (1.56)</td>
<td>3.40 (1.55)</td>
<td>.01</td>
<td>3.30 (1.59)</td>
<td>2.90 (1.50)</td>
<td>.007</td>
</tr>
<tr>
<td>Maximum degree</td>
<td>5.38 (1.94)</td>
<td>4.55 (2.06)</td>
<td>&lt;.001</td>
<td>4.78 (1.86)</td>
<td>4.32 (1.92)</td>
<td>.01</td>
</tr>
<tr>
<td>Mean degree</td>
<td>3.95 (1.71)</td>
<td>3.47 (1.72)</td>
<td>.02</td>
<td>3.70 (1.64)</td>
<td>3.47 (1.66)</td>
<td>.14</td>
</tr>
<tr>
<td><strong>Network composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of kin</td>
<td>47.69 (22.86)</td>
<td>50.27 (26.87)</td>
<td>.38</td>
<td>46.06 (29.34)</td>
<td>54.36 (30.16)</td>
<td>.003</td>
</tr>
<tr>
<td>SD of age</td>
<td>12.67 (5.20)</td>
<td>11.60 (6.26)</td>
<td>.15</td>
<td>12.69 (6.05)</td>
<td>13.74 (6.69)</td>
<td>.23</td>
</tr>
<tr>
<td>Diversity of sex</td>
<td>77.61 (24.00)</td>
<td>70.31 (30.73)</td>
<td>.03</td>
<td>57.14 (39.82)</td>
<td>52.17 (57.90)</td>
<td>.29</td>
</tr>
<tr>
<td>Diversity of race</td>
<td>7.97 (18.55)</td>
<td>8.31 (18.43)</td>
<td>.87</td>
<td>5.66 (15.54)</td>
<td>7.13 (20.47)</td>
<td>.39</td>
</tr>
<tr>
<td>Percentage of network members contacted weekly or less</td>
<td>30.34 (26.35)</td>
<td>18.78 (22.23)</td>
<td>&lt;.001</td>
<td>23.48 (24.95)</td>
<td>14.89 (20.95)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Percentage of network members known for &lt;6 years</td>
<td>21.22 (23.58)</td>
<td>18.31 (21.95)</td>
<td>.27</td>
<td>13.26 (19.83)</td>
<td>10.41 (18.39)</td>
<td>.12</td>
</tr>
<tr>
<td>Percentage of network members who live &gt;15 miles (&gt;24 km) away</td>
<td>41.37 (28.04)</td>
<td>38.88 (28.42)</td>
<td>.45</td>
<td>31.48 (25.57)</td>
<td>33.76 (29.36)</td>
<td>.38</td>
</tr>
<tr>
<td>Percentage of network members who drink</td>
<td>17.85 (27.04)</td>
<td>19.13 (27.58)</td>
<td>.70</td>
<td>8.85 (18.87)</td>
<td>10.32 (21.66)</td>
<td>.51</td>
</tr>
<tr>
<td>Percentage of network members who smoke</td>
<td>12.70 (24.28)</td>
<td>8.09 (18.57)</td>
<td>.06</td>
<td>12.72 (22.43)</td>
<td>9.13 (17.90)</td>
<td>.06</td>
</tr>
<tr>
<td>Percentage of network members who are nonexercisers</td>
<td>41.93 (27.91)</td>
<td>39.86 (30.42)</td>
<td>.55</td>
<td>39.33 (30.94)</td>
<td>31.88 (33.13)</td>
<td>.01</td>
</tr>
<tr>
<td>Percentage of network members who have a bad diet</td>
<td>31.32 (28.73)</td>
<td>25.74 (27.77)</td>
<td>.11</td>
<td>24.95 (29.57)</td>
<td>24.87 (30.21)</td>
<td>.98</td>
</tr>
<tr>
<td>Percentage of network members who have a negative health influence</td>
<td>36.29 (24.75)</td>
<td>35.59 (25.14)</td>
<td>.82</td>
<td>36.77 (29.71)</td>
<td>38.02 (28.61)</td>
<td>.70</td>
</tr>
</tbody>
</table>

<sup>a</sup>Italicized P values met the significance threshold (P<.05).

<sup>b</sup>pwMS: people with multiple sclerosis.
Figure 3. Quantile-quantile plots demonstrating longitudinal changes in personal social network features (due to the COVID-19 pandemic) in relation to patient-reported outcomes during the pandemic in people with multiple sclerosis. Comparison of observed and expected associations between the difference in each quantitative structural and compositional personal social network feature (pandemic value minus prepandemic baseline) in relation to each of the 3 measures of neurological and physical function status—(A) Patient Determined Disease Steps (PDDS), (B) Multiple Sclerosis Rating Scale-Revised (MSRS-R), and (C) Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function—during the pandemic in people with multiple sclerosis after adjusting for age, disease duration, employment, income, study cohort, and time lapse between prepandemic and pandemic personal network questionnaire (PERSNET) survey assessments. Expected \( P \) values (–log10 \( P \) value) were plotted on the x-axis and observed \( P \) values (–log10 \( P \) value) on the y-axis. The gray area encompasses the 95% CIs. Points outside the gray area were considered statistically significant without adjustment for multiple comparisons.

Discussion

Principal Findings

Our quantitative assessment of the personal social network environment of pwMS during the early COVID-19 pandemic has several key findings. First, a higher percentage of network members with a perceived negative health influence was associated with greater neurological symptom burden in pwMS and worse physical function in controls during the pandemic, validating our prepandemic findings and suggesting a shared contribution of this important social environmental feature toward physical function in pwMS and controls. The magnitude of this association was larger in controls than in pwMS. Second, the personal social networks of pwMS and controls both experienced contraction during the COVID-19 pandemic when compared to the prepandemic baseline, and the personal networks of pwMS comprised a higher percentage of kin than those of controls during the pandemic. The percentage of people contacted weekly or less also decreased for both pwMS and controls, suggesting increased social isolation overall. Finally, changes in personal network features related to the COVID-19 pandemic (when compared to the prepandemic baseline) in pwMS were not associated with worsening disability during the pandemic, suggesting an element of neurological resilience despite the significant perturbation in social environment and connections.

Our study design has several novel aspects. First, this is the first known direct comparison of the personal social networks of pwMS and controls during the height of the COVID-19 pandemic when public health measures enacted to reduce contagion were widespread. Second, this is the first effort to longitudinally quantify changes in the personal networks of pwMS and controls due to the pandemic. Third, this study evaluated the differential impact of having an MS diagnosis on the associations between personal social network features and clinical outcomes in terms of both strength and directionality. Finally, this largest cross-sectional and longitudinal quantitative examination of the association of personal networks in relation to neurological and physical functions (not just in pwMS) explored the potential of neurological resilience secondary to social perturbation in the setting of the COVID-19 pandemic.

The finding of the association between a higher percentage of network members with a perceived negative health influence and worse neurological disability in pwMS and worse physical function in controls during the COVID-19 pandemic validated our prior findings from the prepandemic period [20]. These findings are unlikely to be spurious, given the relatively high proportion of participants (both pwMS and controls) whose network members have a perceived negative health influence. We hypothesize that having a higher percentage of network members with a perceived negative health influence in one’s social network could indicate low perceived social support and negative illness perception, which are both associated with worse psychosocial and health outcomes [53-59]. Moreover, this may decrease an individual’s likelihood of engaging in healthy behaviors (e.g., minimal alcohol consumption, abstinence from smoking, regular exercise, healthy diet, and medication adherence) that may reduce overall comorbidities, MS-related disease activity, and neurological disability accumulation.

During the pandemic, the personal social networks of both pwMS and controls contracted when compared to the prepandemic period, as indicated by the decrease in network size, effective size, and maximum degree. Although changes in the density of the personal social networks were minimal, constraint increased in both pwMS and controls, providing...
further evidence for network contraction. The contraction of personal social networks was driven by the pruning of weak ties in the social network. In our analysis, the strength of the ties was quantified by assessing the frequency of interaction (percentage contacted weekly or less), the duration of contact (percentage known for <6 y), and the proximity (percentage who live >15 miles [<24 km] away) of network members. Although there was no change in the duration of contact or the proximity of network members for both pwMS and controls, there was a statistically significant decrease in the frequency of contact (percentage contacted weekly or less) in both groups. The reduction in the network size and frequency of contact suggests that both pwMS and control participants interacted less with weak ties in their network based on our method of assessment of tie strength. In the context of the early COVID-19 pandemic, it is conceivable that all participants kept their social circles tight knit and interacted predominantly with close contacts to minimize potential COVID-19 exposure.

The previously reported association between tightly knit personal social networks and worse physical function was not observed in this study [20]. We postulate several explanations. First, the contraction of social networks in pwMS as well as controls during the pandemic when compared to the prepandemic baseline is likely attributable to public health measures aiming to prevent the spread of COVID-19. Personal social networks contracted and became tighter knit among pwMS during the pandemic, whereas there was greater diversity (e.g., higher effective network size and lower constraint) in the personal networks of pwMS before the pandemic. In our prior cross-sectional study, the direction of the association between network features and physical function could not be fully determined [20]. It is possible that pwMS with better neurological and physical function might have had more diverse personal social networks before the pandemic, but the overall contraction of social networks during the pandemic among pwMS might have masked the association between tightly knit networks and worse physical function. Future longitudinal analyses examining the change in network features from the pandemic period to the postpandemic period may provide insight if the network diversity reverts to the prepandemic baseline versus if the new baseline persists. Second, because the contraction of personal social networks during the pandemic was a relatively recent (as of the data collection) effect and the pandemic, there may not have been sufficient time for the onset of observable changes in neurological function. An assessment of neurological function at future time points could inform the long-term impact of social network changes due to the pandemic. Finally, the lack of an association between tight-knit personal networks and physical function suggests potential resilience of pwMS in coping with social changes attributable to the pandemic [57-60]. Although having tight-knit networks is typically associated with negative health outcomes, tight-knit networks may be advantageous in the context of minimizing COVID-19 exposures. In pwMS, these tight-knit networks saw an increase in the percentage of kin, which was not seen in controls. Having the support of kin could potentially help preserve neurological function in pwMS (e.g., family members help patients to get to clinic or rehabilitation visits or increase medication adherence). The differential impact of the COVID-19 pandemic and social isolation policies may have led pwMS to minimize social contact and restrict themselves to their tight-knit circle, given their higher risk of severe manifestations of COVID-19 [29-32].

The moderation analysis examined the interacting effect of having an MS diagnosis on the association between personal social networks and PROs in a joint analysis of pwMS and controls. An MS diagnosis moderated the association between specific personal social network features and physical function with respect to the direction (i.e., diversity of race, percentage of people who live >15 miles [<24 km] away, and percentage of people who drink) or strength (i.e., percentage of people with a perceived negative health influence) of the associations. Some of the differences in the results between the moderation analysis and the cross-sectional analysis for pwMS could be due to the inclusion of an MS diagnosis as a moderator variable (affecting the strength and direction of the association between network features and PROs) and inability to adjust for disease duration in this joint analysis because this covariate is not applicable for controls. Notably, a lower percentage of people who live >15 miles (>24 km) away and a lower percentage of people who drink within the personal networks of pwMS (but not controls) were each associated with worse physical function. pwMS with low physical function may need more support, which may explain their personal networks comprising a low percentage of people who live far away and a high proportion of kin who are more likely to live in the same household or in the vicinity. pwMS with low physical function may also be more inclined to seek out individuals whose healthy behaviors (e.g., minimal alcohol consumption) could have a positive health influence. The association between a higher percentage of people with a perceived negative health influence and worse physical function in pwMS (and controls) persisted. Interestingly, having higher perceived negative health influences had more of an impact on physical function in controls than in pwMS. We hypothesize that this effect is weaker in pwMS due to the likely resilience of pwMS, given their experience with a chronic neurological disease and having support from tight-knit circles, especially during the pandemic [60-63]. As pwMS have tight-knit networks with a high percentage of kin, it is conceivable that they have better support systems than controls. Interventions that empower individuals with the knowledge, skills, and support to effectively reduce the impact of perceived negative health influences (e.g., education on stress management and coping strategies) may be beneficial in the general population.

Strengths and Limitations
This study has several strengths. First, the longitudinal analysis of changes in personal social networks due to the pandemic in comparison with the prepandemic baseline had a within-subject design. Consequently, we postulated that changes in network features suggestive of personal network contraction in the same participants during the pandemic (compared to their prepandemic baseline) is likely attributable to the pandemic, possibly because of the necessary public health measures. Second, we used three independent but interrelated PROs as a pragmatic method to assess the real-world status of neurological disability and physical function during the early pandemic period when clinical research participation became severely restricted.
These validated PROs, two specifically for MS and one generalizable across health and disease, have shown strong correlations with physician-rated measures of neurological function (e.g., EDSS) in prior studies [43-47]. Third, we conducted both cross-sectional and longitudinal analyses in not only pwMS but also in controls, which enabled a comparison of the differential impact of the pandemic on their personal social networks. Fourth, the moderation analysis examining the role of having an MS diagnosis on the association between network features and physical function enabled an exploration of the strength and directionality of these complex relationships. Finally, we leveraged a large multicenter data set representative of the northeastern and mid-Atlantic regions of the United States with greater geographic diversity than prior studies, potentially increasing the generalizability of the study findings.

Our study also has limitations. First, the direction of the association between the percentage of people with a perceived negative health influence and physical function cannot be determined because we cannot infer causality. Therefore, it is possible that participants with greater disability or worse physical function could perceive that their personal network members exert a negative health influence. Conversely, participants with a higher percentage of network members with a perceived negative health influence could have greater disability or worse physical function. This limitation could be addressed by testing causality in future intervention studies. For example, in persons with a high percentage of people with a negative influence in the personal social network, we can compare the efficacy of interventions engaging the social network of the index person (e.g., providing closer monitoring of medication adherence and more encouragement to promote healthy behaviors) against the standard of care. Second, we did not assess the health status of network members as perceived by the index participant (e.g., if a network member was in bad or good health) or examine the association between the health status of the index participant and the perceived health status of their network members. This limitation could be addressed by incorporating relevant questions in future versions of the PERSNET survey. Third, the sample size of the longitudinal analysis was limited by the relatively modest number of participants with available quantified prepandemic personal networks as baseline. As such, the study did not have sufficient power for subgroup analysis stratified by demographic or clinical subtypes. Last, we sampled participants’ personal networks and neurological and functional status at one time point during the COVID-19 pandemic. The COVID-19 cases varied across areas and throughout the pandemic, whereas participants resided in urban, suburban, and rural environments that were differentially affected during the pandemic. Nevertheless, most of the study populations shared broader geographic regions (northeastern and mid-Atlantic United States) and completed the study response during the early period of the pandemic when federal- and state-level mandates were relatively uniform in terms of stay-at-home orders, physical distancing, and other mitigation guidelines.

Conclusions
In conclusion, this study highlights the impact of the COVID-19 pandemic on personal social networks in pwMS and controls. Our findings generate important hypotheses for testing future interventions that may modify personal social networks to improve health outcomes. Future longitudinal studies examining the long-term impact of the COVID-19 pandemic on the evolution of personal social networks and neurological outcomes in people with chronic neurological disorders such as MS are warranted.

Acknowledgments
This study was supported by the National Institute of Neurological Disorders and Stroke (R01NS124882).

Data Availability
The data sets generated for this study are not publicly available because protected health information is collected in the PERSNET survey. The deidentified data are available from the corresponding author on reasonable request and will be shared with approval from each participating site.

Conflicts of Interest
BWG has received grant or research support from Novartis, Genentech, EMD Serono, Celgene or Bristol Myers Squibb, Sanofi, Genzyme, Janssen, Horizon, Bayer, and Labcorp and has participated in speakers’ bureaus for Biogen. She serves on the editorial board of BMJ Neurology, Children, CNS Drugs, MS International, Journal of Neurology, and Frontiers in Epidemiology. E Longbrake has received honoraria for consulting for Bristol Myers Squibb, EMD Serono, Genentech, Genzyme, NGM Bio, TG Therapeutics, and Janssen. She has received research support from Genentech and Biogen. She is on the editorial staff of Annals of Neurology and serves on the editorial boards of Journal of Neuroimmunology and Neurology: Neuroimmunology & Neuroinflammation. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Personal network (PERSNET) survey.
[PDF File (Adobe PDF File), 149 KB - publichealth_v10i1e45429_app1.pdf]
References


Abbreviations

COVID-19: coronavirus disease 2019
CUIMC: Columbia University Irving Medical Center
EDSS: Expanded Disability Status Scale
GEMS: Genes and Environment in Multiple Sclerosis
MS: multiple sclerosis
MSReCOV: Multiple Sclerosis Resilience to COVID-19
MSRS-R: Multiple Sclerosis Rating Scale-Revised
**PDDS:** Patient Determined Disease Steps  
**PERSNET:** personal network  
**PRO:** patient-reported outcome  
**PROMIS:** Patient-Reported Outcomes Measurement Information System  
**pwMS:** people with multiple sclerosis  
**REDCap:** Research Electronic Data Capture  
**UPMC:** University of Pittsburgh Medical Center

---

*Edited by A Mavragani, T Sanchez; submitted 30.12.22; peer-reviewed by J Wu, A Gillespie; comments to author 12.04.23; revised version received 05.08.23; accepted 31.08.23; published 06.02.24.*

**Please cite as:**  
Impact of the COVID-19 Pandemic on the Personal Networks and Neurological Outcomes of People With Multiple Sclerosis: Cross-Sectional and Longitudinal Case-Control Study  
JMIR Public Health Surveill 2024;10:e45429  
URL: [https://publichealth.jmir.org/2024/1/e45429](https://publichealth.jmir.org/2024/1/e45429)  
doi:10.2196/45429  
PMID:38319703

©Claire Riley, Shruthi Venkatesh, Amar Dhand, Nandini Doshi, Katelyn Kavak, Elle Levit, Christopher Perrone, Bianca Weinstock-Guttman, Erin Longbrake, Philip De Jager, Zongqi Xia. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 06.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Association of Visceral Obesity Indices With Incident Diabetic Retinopathy in Patients With Diabetes: Prospective Cohort Study

Jiaheng Chen1*, MM; Yu Ting Li2,3*, MPH; Zimin Niu1, MM; Zhanpeng He4, MD; Yao Jie Xie5, PhD; Jose Hernandez6,7, DPhil; Wenyong Huang2,3, MD; Harry H X Wang1,8,9, PhD; Guangzhou Diabetic Eye Study Group2

1School of Public Health, Sun Yat-Sen University, Guangzhou, China
2State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, China
3Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, China
4Liwan Central Hospital of Guangzhou, Guangzhou, China
5School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, China (Hong Kong)
6Faculty of Medicine and Health, EDU, Digital Education Holdings Ltd, Kalkara, Malta
7Green Templeton College, University of Oxford, Oxford, United Kingdom
8JC School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, New Territories, China (Hong Kong)
9Usher Institute, Deanery of Molecular, Genetic & Population Health Sciences, The University of Edinburgh, Edinburgh, United Kingdom

*these authors contributed equally

Corresponding Author:
Harry H X Wang, PhD
School of Public Health
Sun Yat-Sen University
74 Zhongshan Road 2
Guangzhou, 510080
China
Phone: 86 20 87330672
Fax: 86 20 87330672
Email: haoxiangwang@163.com

Abstract

Background: Visceral adipose tissue plays an active role in the pathogenesis of type 2 diabetes and vascular dysfunction. The lipid accumulation product (LAP), visceral adiposity index (VAI), and Chinese VAI (CVAI) have been proposed as simple and validated surrogate indices for measuring visceral adipose tissue. However, the evidence from prospective studies on the associations between these novel indices of visceral obesity and diabetic retinopathy (DR) remains scant.

Objective: This study aimed to investigate the longitudinal associations of LAP, VAI, and CVAI with incident DR in Chinese patients with diabetes.

Methods: This was a prospective cohort study conducted in Guangzhou in southern China. We collected baseline data between November 2017 and July 2020, while on-site follow-up visits were conducted annually until January 2022. The study participants consisted of 1403 patients with a clinical diagnosis of diabetes, referred from primary care, who were free of DR at baseline. The LAP, VAI, and CVAI levels were calculated by sex-specific equations based on anthropometric and biochemical parameters. DR was assessed using 7-field color stereoscopic fundus photographs and graded according to the modified Airlie House Classification scheme. Time-dependent Cox proportional hazard models were constructed to estimate the hazard ratios with 95% CIs. Restricted cubic spline curves were fitted to examine the dose-response relationship between the 3 indices of visceral obesity and diabetic retinopathy (DR) remains scant.

Results: The mean age of study participants was 64.5 (SD 7.6) years, and over half (816/1403, 58.2%) were female. During a median follow-up of 2.13 years, 406 DR events were observed. A 1-SD increment in LAP, VAI, or CVAI was consistently associated with increased risk for new-onset DR, with a multivariable-adjusted hazard ratio of 1.24 (95% CI 1.09-1.41; \( P=0.001 \)), 1.22 (95% CI 1.09-1.36; \( P<0.001 \)), and 1.48 (95% CI 1.19-1.85; \( P=0.001 \)), respectively. Similar patterns were observed across tertiles in LAP (\( P \) for trend=.001), VAI (\( P \) for trend<.001), and CVAI (\( P \) for trend=.009). Patients in the highest tertile of LAP,
VAI, and CVAI had an 84%, 86%, and 82% higher hazard of DR, respectively, compared to those in the lowest tertile. A nonlinear dose-response relationship with incident DR was noted for LAP and VAI (both $P$ for nonlinearity<.05), but not for CVAI ($P$ for nonlinearity=.51). We did not detect the presence of effect modification by age, sex, duration of diabetes, BMI, or comorbidity (all $P$ for interaction>.10).

Conclusions: Visceral obesity, as measured by LAP, VAI, or CVAI, is independently associated with increased risk for new-onset DR in Chinese patients with diabetes. Our findings may suggest the necessity of incorporating regular monitoring of visceral obesity indices into routine clinical practice to enhance population-based prevention for DR.

(JMIR Public Health Surveill 2024;10:e48120) doi:10.2196/48120

KEYWORDS
Chinese visceral adiposity index; community-based cohort; diabetic retinopathy; lipid accumulation product; visceral adiposity index; visceral obesity indices

Introduction

Diabetes is an important global public health priority, with an estimated adult prevalence of 10.5% in 2021, rising to 12.2% in 2045 worldwide [1,2]. It poses an enormous threat to health and health care due to the associated mortality, disability, and costly long-term complications [3-5]. As one of the most common microvascular complications, diabetic retinopathy (DR) affects more than one-third of patients with diabetes and remains the leading cause of preventable visual impairment and blindness in working-age adults [6-9]. Given the growing prevalence of diabetes on a global scale [2,10], people at risk for DR are projected to increase rapidly over the coming decades [11]. Identifying modifiable risk factors for DR becomes critically imperative to inform clinical practice and public health recommendations in the context of addressing the ongoing epidemic of DR [12,13].

Observational epidemiologic studies have suggested hyperglycemia and hypertension as major risk factors for DR [6,14]; whereas clinical trial data demonstrated that stand-alone intensive control of blood glucose and blood pressure (BP) might not suffice to significantly reduce DR risk [15]. Meanwhile, genetic data show that the development of type 2 diabetes and obesity share environmental exposures and mechanisms [16]. Despite the established relationship between obesity and type 2 diabetes [17,18], meta-analyses of the association between overweight or obesity and DR have yielded mixed results [19-21]. Recent studies have reported that the distribution of adipose tissue rather than the total amount of fat is more important in the pathogenesis of insulin resistance, diabetes, and vascular dysfunction [22-25], thereby suggesting that visceral adipose tissue (VAT), in contrast to subcutaneous adipose tissue (SAT), tends to play a greater role in the development of DR. Nevertheless, previous studies mainly focused on generalized adiposity (as defined by the BMI) [26] or simple abdominal adiposity (as defined by waist-to-hip ratio [WHR]) [27,28], instead of more specific visceral adiposity.

At present, computed tomography (CT) and magnetic resonance imaging (MRI) techniques have been regarded as the “gold standard” for direct measurement of VAT [29]. Although the imaging technique has been used in large-scale studies in the West, such as the UK Biobank [30], it is less likely to be routinely adopted owing to its expensiveness and complex procedures [31]. Dual-energy x-ray absorptiometry (DEXA) and bioelectrical impedance analysis (BIA) have been used alternatively in epidemiological surveys [32,33]. However, access to these modalities to assess VAT, on top of the routine measures in clinical practice, may require trained technicians and dedicated facilities, with additional workload on health care in low-resource settings [29].

The lipid accumulation product (LAP), visceral adiposity index (VAI), and Chinese visceral adiposity index (CVAI), which could be easily implemented, have therefore been proposed as surrogate indices of VAT for wider use. The validation of these novel indices, when compared to traditional anthropometric adiposity measures, has demonstrated higher accuracy of visceral obesity discrimination and better prediction of type 2 diabetes [34-38]. However, findings from existing studies, mostly using a cross-sectional design, on associations between these indices of visceral obesity and DR are largely inconsistent [39-41]. More evidence from prospective studies is required to address this area of controversy. Therefore, we aimed to investigate the longitudinal associations of 3 validated visceral obesity indices, that is, LAP, VAI, and CVAI, with incident DR in patients with diabetes.

Methods

Study Design and Participants

This is an ongoing prospective cohort study among Chinese patients with diabetes conducted in Guangzhou in southern China. The study design was reported in detail elsewhere [28,42]. In brief, the participants consisted of primary care patients aged between 30 and 80 years with a clinical diagnosis of type 2 diabetes. All participants were referred through a generalist-specialist alliance consisting of 18 community health centers to a national-leading tertiary hospital specializing in ophthalmology (Zhongshan Ophthalmic Center), where a dilated, comprehensive eye examination was provided free of charge at baseline and at annual follow-up visits. The presence of type 2 diabetes was assessed by the attending primary care physician according to the Chinese Diabetes Society guideline and the World Health Organization recommendation when fasting plasma glucose ≥7.0 mmol/L; 2-hour plasma glucose ≥11.1 mmol/L during a 75-g oral glucose tolerance test; or hemoglobin A1c (HbA1c) ≥6.5% [43,44]. All participants had their HbA1c tested at the Zhongshan Ophthalmic Center to ensure that all enrolled patients who were not on
The indices were calculated based on demographics satisfactory correlations with VAT measured using CT scans LAP, VAI, and CVAI, all of which have demonstrated Visceral obesity was assessed using 3 validated indices, that is, Assessment of Visceral Obesity Indices provided in Multimedia Appendix 1.

Information on the definition and categorization of variables is Epidemiology Collaboration equation for Asians [52]. Detailed rate (eGFR) was calculated using the Chronic Kidney Disease automated, high-performance liquid chromatography system (Cobas 8000, Roche Diagnostics). HbA1c was measured using an automatic biochemistry modular analyzer (Cobas 8000, Roche Diagnostics). Waist circumference (WC) was taken at the midpoint between the lower margin of the last palpable rib and the top of the iliac crest, while hip circumference was measured at the largest circumference around the buttocks, both to the nearest 0.1 cm. Waist-to-height ratio (WHR) was calculated as WC (cm) divided by height (cm). The WHR was calculated as WC (cm) divided by hip circumference (cm). BP was measured in a seated position after a 10-minute rest by routinely validated automatic sphygmomanometers (HEM-907, Omron). The arm with the higher BP values was used. The average of 2 BP readings, 1-2 minutes apart, was recorded. Serum creatinine and lipid profiles, including plasma cholesterol and triglycerides, were directly measured using an automatic biochemistry chromatography system (G8, Sysmex Corporation). The estimated glomerular filtration rate (eGFR) was calculated using the Chronic Kidney Disease Epidemiology Collaboration equation for Asians [52]. Detailed information on the definition and categorization of variables is provided in Multimedia Appendix 1.

Visceral obesity was assessed using 3 validated indices, that is, LAP, VAI, and CVAI, all of which have demonstrated satisfactory correlations with VAT measured using CT scans [34,36]. The indices were calculated based on demographics (sex and age), anthropometrics (BMI and WC), and metabolic parameters (triglycerides and high-density lipoprotein cholesterol). The detailed formulas were described in Multimedia Appendix 1. Given the lack of consensus regarding the optimal cutoff point for identifying visceral obesity and the considerable disparities in the amount of VAT between males and females, the 3 indices were divided into sex-specific tertiles when treated as categorical variables.

Fundus Examination and Grading of DR
All study participants had color stereoscopic fundus photographs of seven standard fields taken in each eye after pupil dilation using a digital retinal camera (Canon CR-2). The photographs were graded at the Zhongshan Ophthalmic Center using the modified Airlie House Classification scheme as adapted for the Early Treatment Diabetic Retinopathy Study (ETDRS) [53]. A total of 2 trained ophthalmic specialists independently graded the fundus photographs, with disagreements (<8%) being resolved by the decision of a senior ophthalmologist. The grading of DR was determined based on the worst eye, ranging from the ETDRS classification levels from 10 to 85, that is, levels 10-20 (no apparent DR), level 35 (mild nonproliferative DR [NPDR]), levels 43-47 (moderate NPDR), level 53 (severe NPDR), and levels 61-85 (proliferative DR) [53]. Incident DR was defined as eyes with no apparent DR at baseline in which any DR was newly diagnosed at follow-up.

A more precise evaluation of retinal pathology in diabetic macular edema (DME) was performed with a swept-source optical coherence tomography device (DRI-OCT Triton, Topcon). The swept-source optical coherence tomography volume was captured in a 3D scan pattern over a 6x6 mm area for all eyes. The presence of DME, which was assessed separately from that of DR, was characterized by retinal thickening or hard exudates in the posterior pole that can develop at any stage of DR [54]. Given that the vision-threatening DR (VTDR) includes severe NPDR, proliferative DR, or DME [55,56], the presence of DME was taken into account in the sensitivity analysis.

Statistical Analysis
The Pearson correlation coefficient was examined for the 3 visceral obesity indices (ie, LAP, VAI, and CVAI) and the conventional obesity measures (ie, BMI, WC, WHtR, and WHR), respectively. Patients enrolled at baseline were followed until they were diagnosed with any DR or the recorded attendance at the most recent follow-up visits until January 2022, whichever came first. The incidence density was reported as the number of outcome events per 100 person-years. Time-dependent Cox proportional hazard models were constructed to explore the associations of LAP, VAI, and CVAI with incident DR at follow-up. The 3 visceral obesity indices and all other covariates, except for sex and education level, were considered time-varying variables in the main analysis. The model construction was described in detail in Multimedia Appendix 1.

Hazard ratios (HRs) with 95% CIs were estimated from the crude and adjusted models. Model 1 was adjusted for sex, age, and duration of diabetes. Model 2 was further adjusted for...
education level, current smoking, regular drinking, BMI, BP, HbA$_{1c}$, low-density lipoprotein cholesterol, eGFR, and use of insulin. Variable selection was determined from our previous knowledge of sociodemographic factors, health-related lifestyles, and anthropometric and biochemical parameters [6,9,14,55]. The proportional hazards assumption for model fit was tested using the scaled Schoenfeld residuals. The variance inflation factors were examined to ensure the absence of multicollinearity (ie, all variance inflation factors <5) in the regression model. We further modeled data as restricted cubic splines (RCSs) with 4 knots, located at the 5th, 35th, 65th, and 95th percentiles following the Akaike information criterion [57] of LAP, VAI, and CVAI, respectively, to assess the shape of the association between the 3 visceral obesity indices and risk of DR. We performed separate Cox regression analyses across patient subgroups according to sex (female vs male); age (<65 vs ≥65 years old); duration of diabetes (<10 vs ≥10 years); BMI (<24 vs ≥24 kg/m$^2$); and the presence of concurrent medical conditions including hypertension (yes vs no), dyslipidemia (yes vs no), and decreased renal function (yes vs no) to investigate the potential effect modifiers. The interaction between visceral obesity and the stratifying variable was explored by inserting a 2-factor interaction term into the regression model.

A series of sensitivity analyses were performed. First, we added the presence of DME to the outcome of interest, which was redefined as new-onset DR or DME at follow-up in patients free of both DR and DME at baseline. Second, we excluded patients who had incident DR within the first year of follow-up to account for the possible reverse causality bias. Third, we estimated HR from the Cox regression models without adjustment for BMI to avoid the plausible bias associated with over-adjustment. Fourth, we incorporated WC as a covariate in the regression model to detect whether the visceral obesity indices may have a role independent of abdominal obesity. We also estimated HR from time-fixed Cox regression models, in which baseline values of LAP, VAI, and CVAI were used to ascertain whether associations between visceral obesity indices and incident DR may vary from the time-dependent main analysis. Analyses were conducted using Stata (version 15.1; StataCorp LLC) and R (version 4.2.2; Core Team). A 2-tailed $P<.05$ was considered statistically significant.

**Ethical Considerations**

Data anonymization was performed by removing all patient identifiers from the data set before data analysis. Ethics approval was granted by the Zhongshan Ophthalmic Center Medical Ethics Committee (2017KYPJ094) at Sun Yat-Sen University as per the Declaration of Helsinki 2013. All participants provided written, informed consent.

**Results**

**Characteristics of Study Participants**

Of the 1403 patients included in the final analysis, slightly over half (816/1403, 58.2%) were female. The mean (SD) age of patients was 64.5 (7.6) years at baseline. The median (IQR) values of LAP, VAI, and CVAI were 48 (29.12-77.22), 2.69 (1.54-4.43), and 118.18 (94.77-141.17), respectively. During a median follow-up period of 2.13 years, we documented 406 new-onset DR events (ie, 289 cases of mild NPDR, 116 cases of moderate NPDR, and 1 case of severe NPDR), with an incidence density of 15.71 per 100 person-years. Patients who experienced new-onset DR at follow-up tended to be female, younger, with a longer duration of diabetes, on insulin, and had higher HbA$_{1c}$ and greater visceral obesity at baseline than their counterparts (Table 1). The correlation matrix showed statistically significant correlations among LAP, VAI, CVAI, BMI, WC, WHtR, and WHR (Table S1 in Multimedia Appendix 1). The sex distribution, duration of diabetes, lifestyle behaviors, comorbidity status, and levels of the majority of anthropometric and biochemical measurements were comparable between patients excluded during follow-up and those included in the final analysis, albeit slightly younger (64.45 vs 65.23 years) and with better glycemic control in the final sample (Table S2 in Multimedia Appendix 1).
Table 1. Baseline characteristics of study participants by incident diabetic retinopathy (DR) at follow-up. The 2-sample t test, Mann-Whitney U test, or chi-square test, where appropriate, was used for between-group comparison.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Absence of incident DR at follow-up (n=997)</th>
<th>Presence of incident DR at follow-up (n=406)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex, n (%)</td>
<td>560 (56.2)</td>
<td>256 (63.1)</td>
<td>.02</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, mean (SD)</td>
<td>64.72 (7.49)</td>
<td>63.77 (7.81)</td>
<td>.04</td>
</tr>
<tr>
<td>≥65, n (%)</td>
<td>515 (51.7)</td>
<td>186 (45.8)</td>
<td>.047</td>
</tr>
<tr>
<td>Duration of diabetes (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, median (IQR)</td>
<td>6.0 (3-11)</td>
<td>7.0 (3-13.5)</td>
<td>.002</td>
</tr>
<tr>
<td>≥10, n (%)</td>
<td>344 (34.5)</td>
<td>162 (40)</td>
<td>.053</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td>.56</td>
</tr>
<tr>
<td>Junior secondary school or below</td>
<td>316 (31.7)</td>
<td>120 (29.5)</td>
<td></td>
</tr>
<tr>
<td>Senior secondary school</td>
<td>399 (40)</td>
<td>175 (43.1)</td>
<td></td>
</tr>
<tr>
<td>College or above</td>
<td>282 (28.3)</td>
<td>111 (27.4)</td>
<td></td>
</tr>
<tr>
<td>Current smoking, n (%)</td>
<td>130 (13)</td>
<td>55 (13.5)</td>
<td>.82</td>
</tr>
<tr>
<td>Regular drinking, n (%)</td>
<td>87 (8.7)</td>
<td>45 (11.1)</td>
<td>.20</td>
</tr>
<tr>
<td>Use of insulin, n (%)</td>
<td>140 (14)</td>
<td>93 (22.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Presence of comorbidity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>548 (55)</td>
<td>246 (60.6)</td>
<td>.054</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>662 (66.4)</td>
<td>276 (68)</td>
<td>.57</td>
</tr>
<tr>
<td>Decreased renal function</td>
<td>414 (41.5)</td>
<td>163 (40.2)</td>
<td>.64</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, mean (SD)</td>
<td>24.59 (3.26)</td>
<td>24.73 (3.29)</td>
<td>.47</td>
</tr>
<tr>
<td>≥24, n (%)</td>
<td>552 (55.4)</td>
<td>229 (56.4)</td>
<td>.72</td>
</tr>
<tr>
<td>Waist circumference (cm), mean (SD)</td>
<td>85.87 (9.11)</td>
<td>86.17 (9.16)</td>
<td>.57</td>
</tr>
<tr>
<td>Blood pressure (mm Hg), mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>132.40 (17.98)</td>
<td>133.88 (18.45)</td>
<td>.17</td>
</tr>
<tr>
<td>Diastolic blood pressure</td>
<td>70.02 (10.14)</td>
<td>71 (10.44)</td>
<td>.10</td>
</tr>
<tr>
<td>HbA₁c, (%), mean (SD)</td>
<td>6.71 (1.16)</td>
<td>7.04 (1.34)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Cholesterol (mmol/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cholesterol, mean (SD)</td>
<td>4.83 (1.09)</td>
<td>4.81 (0.98)</td>
<td>.72</td>
</tr>
<tr>
<td>Triglycerides, median (IQR)</td>
<td>1.89 (1.31-2.85)</td>
<td>1.99 (1.39-2.93)</td>
<td>.21</td>
</tr>
<tr>
<td>LDL-Cb, mean (SD)</td>
<td>3.03 (0.99)</td>
<td>3.04 (0.85)</td>
<td>.89</td>
</tr>
<tr>
<td>HDL-C, mean (SD)</td>
<td>1.31 (0.42)</td>
<td>1.28 (0.36)</td>
<td>.14</td>
</tr>
<tr>
<td>Serum creatinine (μmol/L), mean (SD)</td>
<td>71.49 (19.68)</td>
<td>71.09 (19.66)</td>
<td>.73</td>
</tr>
<tr>
<td>eGFRc (mL/min per 1.73 m²), mean (SD)</td>
<td>90.32 (16.91)</td>
<td>90.25 (16.92)</td>
<td>.95</td>
</tr>
<tr>
<td>Indices, median (IQR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAPd</td>
<td>46.55 (27.64-75.44)</td>
<td>51.70 (31.80-80.08)</td>
<td>.03</td>
</tr>
<tr>
<td>VAIe</td>
<td>2.51 (1.47-4.37)</td>
<td>2.94 (1.79-4.54)</td>
<td>.003</td>
</tr>
<tr>
<td>CVAIf</td>
<td>117.47 (93.28-140.07)</td>
<td>119.58 (96.39-144.91)</td>
<td>.12</td>
</tr>
</tbody>
</table>

aHbA₁c: hemoglobin A1c.
bLDL-C: low-density lipoprotein cholesterol.
HDL-C: high-density lipoprotein cholesterol.
dGFR: estimated glomerular filtration rate.
LAP: lipid accumulation product.
VAI: visceral adiposity index.
CVAI: Chinese visceral adiposity index.

### Association of Visceral Obesity Indices With Incident DR

A 1-SD increment in LAP, VAI, or CVAI was consistently associated with increased risk for DR, with a multivariable adjusted HR (aHR) of 1.24 (95% CI 1.09-1.41; \( P = .001 \)), 1.22 (95% CI 1.09-1.36; \( P < .001 \)), and 1.48 (95% CI 1.19-1.85; \( P = .001 \)) respectively (Table 2). Similar patterns were observed across tertiles in LAP (\( P \) for trend<.001), VAI (\( P \) for trend<.001), and CVAI (\( P \) for trend=.009). Patients in the highest tertile of LAP, VAI, and CVAI had an 84% (aHR=1.84; 95% CI 1.30-2.62), 86% (aHR=1.86; 95% CI 1.35-2.57), and 82% (aHR=1.82; 95% CI 1.16-2.86) higher hazard of new-onset DR, respectively, compared to those in the lowest tertile (Table 2). A positive, nonlinear dose-response relationship with incident DR was noted for LAP and VAI (both \( P \) for nonlinear trend<.05), but not for CVAI (\( P \) for nonlinear trend=.51; Figure 1).

#### Table 2. Time-dependent Cox proportional hazard models with and without adjustment for covariates. The crude model referred to time-dependent Cox proportional hazard models with no adjustment. Model 1 was adjusted for sex, age, and duration of diabetes. Model 2 was then further adjusted for education level, current smoking, regular drinking, BMI, blood pressure, hemoglobin A1c, serum cholesterol level, estimated glomerular filtration rate, and use of insulin.

<table>
<thead>
<tr>
<th>Visceral obesity indices</th>
<th>Crude model</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cHR(^a) (95% CI)</td>
<td>( P ) value</td>
<td>aHR(^b) (95% CI)</td>
</tr>
<tr>
<td><strong>Lipid accumulation product(^c)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per SD increase</td>
<td>1.12 (1.01-1.25)</td>
<td>.03</td>
<td>1.14 (1.03-1.27)</td>
</tr>
<tr>
<td>Tertile 1</td>
<td>1.00 (Reference)</td>
<td>N/A(^d)</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Tertile 2</td>
<td>1.19 (0.89-1.60)</td>
<td>.25</td>
<td>1.27 (0.93-1.72)</td>
</tr>
<tr>
<td>Tertile 3</td>
<td>1.42 (1.07-1.89)</td>
<td>.02</td>
<td>1.53 (1.14-2.05)</td>
</tr>
<tr>
<td>( P ) for trend</td>
<td>N/A(^d)</td>
<td>.02</td>
<td>N/A(^d)</td>
</tr>
<tr>
<td><strong>Visceral adiposity index(^c)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per SD increase</td>
<td>1.18 (1.07-1.32)</td>
<td>.002</td>
<td>1.19 (1.07-1.32)</td>
</tr>
<tr>
<td>Tertile 1</td>
<td>1.00 (Reference)</td>
<td>N/A(^d)</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Tertile 2</td>
<td>1.36 (1.01-1.84)</td>
<td>.04</td>
<td>1.45 (1.07-1.97)</td>
</tr>
<tr>
<td>Tertile 3</td>
<td>1.60 (1.19-2.13)</td>
<td>.002</td>
<td>1.67 (1.24-2.24)</td>
</tr>
<tr>
<td>( P ) for trend</td>
<td>N/A(^d)</td>
<td>.003</td>
<td>N/A(^d)</td>
</tr>
<tr>
<td><strong>Chinese visceral adiposity index(^c)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per SD increase</td>
<td>1.16 (1.04-1.31)</td>
<td>.01</td>
<td>1.22 (1.07-1.37)</td>
</tr>
<tr>
<td>Tertile 1</td>
<td>1.00 (Reference)</td>
<td>N/A(^d)</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Tertile 2</td>
<td>1.25 (0.94-1.67)</td>
<td>.13</td>
<td>1.38 (1.02-1.86)</td>
</tr>
<tr>
<td>Tertile 3</td>
<td>1.39 (1.04-1.85)</td>
<td>.03</td>
<td>1.53 (1.13-2.07)</td>
</tr>
<tr>
<td>( P ) for trend</td>
<td>N/A(^d)</td>
<td>.03</td>
<td>N/A(^d)</td>
</tr>
</tbody>
</table>

\(^a\)cHR: crude hazard ratio.
\(^b\)aHR: adjusted hazard ratio.
\(^c\)SD of lipid accumulation product=38.90, SD of visceral adiposity index=2.73, and SD of Chinese visceral adiposity index=34.23.
\(^d\)N/A: not applicable.
Dose-response relationships were examined using RCS with 4 knots, located at the 5th, 35th, 65th, and 95th percentiles of LAP, VAI, and CVAI, respectively. The solid line represents the fitted curve, and the shaded areas represent the 95% CI bands. Time-dependent Cox proportional hazard models were adjusted for sex, age, duration of diabetes, education level, current smoking, regular drinking, BMI, BP, HbA1c, serum cholesterol level, eGFR, and use of insulin (test for overall trend: LAP P=.001, VAI P<.001, and CVAI P=.006; test for nonlinear trend: LAP P=.048, VAI P=.002, and CVAI P=.51).

Subgroup Analyses

When patients were classified by sex, age, duration of diabetes, BMI, and the presence of concurrent hypertension, dyslipidemia, and decreased renal function, the associations of a 1-SD increment in LAP, VAI, and CVAI with new-onset DR observed in the main analysis remained consistent across all patient subgroups. Multivariable-adjusted Cox models, in which interactions between visceral obesity and the stratifying variables were explored, showed no evidence of effect modification by age, sex, duration of diabetes, BMI, or comorbidity (all P for interaction>.10; Figure 2).

Sensitivity Analyses

Of the 1380 patients with diabetes free of DR and DME at baseline, 414 patients had incident DR or DME at follow-up (ie, 387 cases of DR alone, 16 cases of DME alone, and 11 cases of DR combined with DME). Separate analyses for incident DME (n=27) and VTDR (n=28; including 1 case of severe NPDR) as distinct outcomes were not performed due to the small number of events. The associations of LAP, VAI, and CVAI with incident DR remained unchanged when new-onset DME was added to the outcome of interest or when patients who had new-onset DR within the first year of follow-up were excluded. The strength of associations between visceral obesity indices and incident DR tended to be somewhat attenuated in Cox models without adjustment for BMI; however, the results remained significant (Figure 3).
Sensitivity analysis 1 was performed with time-dependent Cox regression models (n=1380), in which the outcome of interest was redefined as new-onset DR or DME, with models adjusted for sex, age, duration of diabetes, education level, current smoking, regular drinking, BMI, BP, HbA1c, serum cholesterol level, eGFR, and use of insulin. Sensitivity analysis 2 was performed with multivariable-adjusted, time-dependent Cox regression models (n=1320), in which patients who had incident DR within the first year of follow-up were excluded. Sensitivity analysis 3 was performed with multivariable-adjusted, time-dependent Cox regression models (N=1403), in which BMI was not adjusted for in the analysis.

The multivariable-adjusted hazard of incident DR at follow-up associated with LAP, VAI, and CVAI remained significant when WC was incorporated as a covariate in the time-dependent Cox model, as well as in the time-fixed Cox model in which baseline values of visceral obesity indices were used (Table S3 in Multimedia Appendix 1).

Figure 3. Associations of lipid accumulation product (LAP), visceral adiposity index (VAI), and Chinese VAI (CVAI) with incident diabetic retinopathy in the main analysis and sensitivity analyses. aHR: adjusted hazard ratio.

Discussion
Principal Findings
In this prospective cohort study conducted in southern China, we demonstrated longitudinal associations between 3 novel indices of visceral obesity and the incidence of DR in patients with diabetes. During a median follow-up of over 2 years, elevated levels of LAP, VAI, and CVAI were all independently associated with increased risk for new-onset DR, with positive dose-response trends. Our data suggested that visceral obesity was a significant risk factor for DR, with the associations independent of generalized and abdominal obesity. Multivariable-adjusted, time-dependent Cox regression models showed no evidence of effect modification by sex, age, duration of diabetes, BMI, and the presence of concurrent hypertension, dyslipidemia, and decreased renal function.

Comparison With Existing Literature
A recent systematic review of population-based cohort studies reported an annual incidence of DR ranging from 2.2% to 12.7% worldwide [58], while this study showed a slightly higher incidence rate. The differences might be explained by heterogeneity in the use of 7-field color stereoscopic fundus photographs for DR detection and in patient characteristics (eg, in terms of duration of diabetes) across studies. Over one-third of patients in this study had diabetes for more than 10 years and might be at a more advanced stage of diabetes with higher risks for DR.

The association between obesity and DR has been previously reported in a large number of studies, albeit with equivocal findings [19-21]. The lack of consensus may be due to the failure to address the mutually confounding effect resulting from the strong interrelation between generalized and abdominal obesity [27]. Obesity measures that account for the distribution of adipose tissue, and more specifically, visceral adiposity, have been considered a potentially stronger indicator for the risk of DR [25]. Positive associations between visceral obesity and DR have been reported in clinic-based studies among both Japanese and Singaporean adults with type 2 diabetes [59,60]. In contrast, the community-based Jogjakarta Eye Diabetic Study [61] reported opposite findings from Indonesian adults with type 2 diabetes, while a null association between visceral obesity and DR was observed in French patients with diabetes [62]. Plausible explanations for contradictory results include the use of different measurement methods (eg, CT [59], MRI [62], and BIA [60,61]), racial or ethnic disparities, and studies with relatively small sample sizes.

Simple and validated surrogate indices for measuring VAT, such as LAP [34,50], VAI [35], and CVAI [36], have been proposed given the resource availability and time constraints in routine clinical settings. A cross-sectional study in northern China reported positive associations of LAP and CVAI with DR among adults with type 2 diabetes [40], whereas another study with a similar design showed an inverse association between LAP and DR [39]. Cross-sectional community-based findings in eastern China, however, indicated nonsignificant associations of LAP, VAI, and CVAI with DR [41]. A lack of significant associations between visceral obesity indices and DR was also reported from a northern Chinese cohort of patients with type 2 diabetes [63], based on a time-fixed Cox model in which information on the duration of diabetes was not collected and the number of outcome events of DR as assessed by 4-field fundus photographs was much smaller (ie, 90 vs 406 cases) than that in this study, which took into account the time-varying...
effect in the Cox model with DR assessed by color stereoscopic fundus photographs of 7-standard fields.

To date, few studies have characterized the dose-response relationship between visceral obesity indices and DR using RCS functions. We found consistent associations of LAP, VAI, and CVAI with risk for DR, which appeared to indicate the presence of dose-response. Given that the shape of the RCS curve is largely influenced by the location and number of knots, the exploration of a specific threshold per se was not the focus of this study. We observed a nonlinear dose-response curve for LAP and VAI, but not for CVAI. This may be in part due to the inclusion of age as a component in the calculation of CVAI [36], yet further research is needed to better understand how different components used for the calculation of these surrogate indices are interacted with in basic molecular mechanisms and key pathogenic processes that drive abnormalities and lesions in the diabetic retina [64].

To our knowledge, this was one of the first studies using a prospective design to investigate the longitudinal associations of LAP, VAI, and CVAI with new-onset DR simultaneously while evaluating potential nonlinear associations. The anthropometric measurements, laboratory tests, and dilated-pupil retinal ophthalmoscopic examinations were performed annually by a regularly trained team of clinical staff who followed standard operating procedures with quality control. We adopted Early Treatment Diabetic Retinopathy Study 7-field color stereoscopic fundus photography, which has long been the gold standard for DR detection [12], to ensure the absence of DR in patients at baseline enrollment and the accurate capture of DR events during follow-up. The use of time-varying measures of visceral obesity indices and covariates in the Cox models took into account the plausible effects of time-varying exposure over the study period. The analyses were systematically performed using the 3 visceral obesity indices as continuous variables and in tertiles, with a consistent methodology adopted to deal with residual confounding and reverse causality. An extensive range of sensitivity analyses based on time-dependent and time-fixed multivariable-adjusted Cox models yielded little difference in estimated associations between visceral obesity indices and the incidence of DR, suggesting the robustness of our findings.

**Underlying Biological Mechanisms**

Although the biological mechanisms underlying the association between visceral obesity and DR are not yet fully understood, several hypotheses have been proposed. First, it was found that VAT adipocytes are more metabolically active and have greater lipolytic activity compared to SAT adipocytes [65]. Visceral fat accumulation is associated with a greater tendency to hyperglycemia, hyperinsulinemia, hypertriglyceridemia, and increased apolipoproteins B-rich lipoproteins, all of which could play a role in diabetes-related microvascular complications [65,66]. Second, evidence showed that VAT is more infiltrated with inflammatory cells (macrophages) than SAT [67] and is more capable of generating proinflammatory cytokines such as tumor necrosis factor-α, C-reactive protein, and interleukin-6 [65], which are involved in the pathological mechanisms leading to vascular dysfunction [24]. In addition, the plasminogen activator inhibitor-1, which is expressed more in VAT than SAT, has also been proven to be associated with increased susceptibility to DR [68,69]. Third, elevated vascular endothelial growth factor concentrations associated with visceral fat accumulation have been shown to be a potent mediator of angiogenesis and vascular permeability [11] and play a pivotal role in the retinal microvascular complications of diabetes [70,71]. Taken together, an excess of visceral fat could be linked to a state of chronic systemic inflammation and metabolic abnormalities, thereby predisposing patients to the complex progression of retinal vascular diseases [72].

**Implications for Research and Practice**

A recent review concluded that known risk factors for complications of diabetes appear to have limited implications as predictors of retinopathy development or progression [11]. This study provides novel, prospective evidence of the positive association between visceral obesity indices and new-onset DR. Given that stand-alone intensive control of blood glucose and BP might not suffice to significantly reduce DR risk [15], our findings suggest the necessity for regular monitoring of visceral adiposity apart from blood glucose and pressure. As visceral adiposity tends to be ignored by the general public, efforts should be made to increase awareness of the adverse effects of an excess of visceral fat on vascular permeability and growth. This may require tailored health education for high-risk patient groups to improve adherence to healthy lifestyles, thereby reducing excessive visceral fat. This is in line with the most recent American Diabetes Association guideline that advocates lifestyle improvement and obesity management in patients with diabetes [17,73].

From a public health perspective, routine monitoring of visceral obesity indices that are validated and easily implemented in low-resource settings may assist in identifying a broader range of patients who are at risk for DR. This would allow for early detection and timely intervention to avoid irreversible vision loss. As direct access to imaging modalities may not be widely feasible in daily practice [29], alternative approaches using surrogate indices that are broadly applicable to detect and measure intra-abdominal (visceral) fat distribution have gained increasing popularity. This study demonstrates the potential utility of visceral obesity indices, calculated using routinely available demographic, anthropometric, and biochemical measures, in predicting new-onset DR outcomes in remote or rural settings in low- and middle-income countries. Of note, our findings did not indicate which of the 3 indices was superior to one another but rather suggested opportunities for preventing DR through the application of these indices in risk assessment and management. Further interventional studies aimed at examining the effectiveness of continuous monitoring of visceral obesity indices for the prevention of DR are expected to offer important insights into the long-term management of diabetes and inform targeted public health intervention strategies and clinical guidelines.

**Limitations**

This study has some limitations that merit consideration. First, the VAT was not directly measured using CT or MRI or estimated using DEXA or BIA due to the resource constraints; instead, sex-specific equation-based indices that were previously

https://publichealth.jmir.org/2024/1/e48120
shown to be valid and reliable were used [34-36]. Second, the association of visceral obesity indices with incident DR may be underestimated, as the retinopathy outcome may not occur during the study period. Third, nearly one-third of patients enrolled at baseline were lost to follow-up or had missing outcome data. They appeared to be slightly older and had poorer glycemic control compared to their counterparts who attended follow-up, which might undermine the relationship between visceral obesity indices and incident DR in the final analysis despite largely comparable baseline characteristics concerning sex, duration of diabetes, lifestyles, comorbidity, and the majority of clinical measurements. Fourth, a separate analysis for incident DME or VTDR alone was not performed due to the small number of events. Fifth, we cannot rule out the possibility of residual bias because of confounding by unmeasured factors. Last but not least, the generalizability of our findings in Chinese diabetes to other patient populations should be interpreted with caution. Given that the present study was originally centered around evaluating the potential of visceral obesity indices in predicting new-onset DR rather than serving as a comprehensive validation with DEXA or BIA per se, further investigations on the basis of imaging techniques in a larger series of multiethnic patients are warranted to strengthen the evidence for the visceral adiposity-DR relationship.

In conclusion, this study provides prospective evidence that visceral obesity as measured by LAP, VAI, or CVAI is significantly associated with increased risk for new-onset DR, independent of generalized and abdominal obesity in Chinese patients with diabetes. Our findings may suggest the necessity of incorporating regular monitoring of visceral obesity indices in clinical practice to enhance population-based prevention for DR.

Acknowledgments
We wish to acknowledge the community liaison support from the local Health Commission (Bureau) in establishing the Generalist-Specialist Alliance. We also thank our research collaborators and frontline clinical staff involved in the study. Open Research Funds of the State Key Laboratory of Ophthalmology (303060202400377 and 303060202400362); National Natural Science Foundation of China (72061137002); and Traditional Chinese Medicine Research Program of Guangdong Province (20214159). The study sponsor or funder was not involved in the design of the study, the collection, analysis, and interpretation of data, or the writing of the report, and did not impose any restrictions regarding the publication of the report. We declare that generative AI was not used in any portion of the manuscript.

Data Availability
The data sets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Authors’ Contributions
JC and HHXW were responsible for conceptualization and writing of the original draft. Data curation and project administration was done by YTL, with support from WH who leads the Guangzhou Diabetic Eye Study Group. JC performed formal analysis. ZN and ZH contributed to validation. HHXW and WH supervised the work. YTL, YJX, and JH contributed to review and editing. All authors contributed to the interpretation of the data and read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental material.
[DOC File, 224 KB - publichealth_v10i1e48120_app1.doc ]

References


**Abbreviations**

- **aHR**: adjusted hazard ratio
- **BIA**: bioelectrical impedance analysis
- **BP**: blood pressure
- **CT**: computed tomography
- **CVAI**: Chinese visceral adiposity index
- **DEXA**: dual-energy x-ray absorptiometry
- **DME**: diabetic macular edema
- **DR**: diabetic retinopathy
- **eGFR**: estimated glomerular filtration rate
- **ETDRS**: Early Treatment Diabetic Retinopathy Study
- **HbA1c**: hemoglobin A1c
- **HR**: hazard ratio
- **LAP**: lipid accumulation product
- **MRI**: magnetic resonance imaging
- **NPDR**: nonproliferative diabetic retinopathy
- **RCS**: restricted cubic spline
- **SAT**: subcutaneous adipose tissue
- **VAI**: visceral adiposity index
- **VAT**: visceral adipose tissue
- **VTDR**: vision-threatening diabetic retinopathy
- **WC**: waist circumference
- **WHR**: waist-to-hip ratio
- **WHtR**: waist-to-height ratio

©Jiaheng Chen, Yu Ting Li, Zimin Niu, Zhanpeng He, Yao Jie Xie, Jose Hernandez, Wenyong Huang, Harry H X Wang, Guangzhou Diabetic Eye Study Group. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 06.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic
information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Age Differences in the Association of Sleep Duration Trajectory With Cancer Risk and Cancer-Specific Mortality: Prospective Cohort Study

Chenan Liu1,2,3,4*, MD, PhD; Qingsong Zhang5*, MD; Chenning Liu6*, MD, PhD; Tong Liu1,2,3,4, MD, PhD; Mengmeng Song7, MD, PhD; Qi Zhang8, MD, PhD; Hailun Xie1,2,3,4, MD, PhD; Shiqi Lin1,2,3,4, MD, PhD; Jiangshan Ren9, MD; Yue Chen1,2,3,4, MD, PhD; Xin Zheng1,2,3,4, MD, PhD; Jinyu Shi1,2,3,4, MD, PhD; Li Deng2,3,4, PhD; Hanping Shi1,2,3,4, MD, PhD; Shouling Wu10, MD, PhD

1Department of Gastrointestinal Surgery and Clinical Nutrition, Beijing Shijitan Hospital, Beijing, China
2National Clinical Research Center for Geriatric Diseases, Xuanwu Hospital, Capital Medical University, Beijing, China
3Key Laboratory of of Cancer Food for Special Medical Purposes for State Market Regulation, Beijing, China
4Laboratory for Clinical Medicine, Capital Medical University, Beijing, China
5Department of General Surgery, Kailuan General Hospital, Tangshan, China
6Department of Obstetrics and Gynecology, Dongguan Maternal and Child Health Care Hospital, Dongguan, China
7Cardiovascular Research Institute, University of California, San Francisco, CA, United States
8Department of Genetics, Yale University School of Medicine, New Haven, CT, United States
9Department of Oncology, First Hospital of Shanxi Medical University, Taiyuan, China
10Department of Cardiology, Kailuan General Hospital, Tangshan, China

*these authors contributed equally

Corresponding Author:
Shouling Wu, MD, PhD
Department of Cardiology
Kailuan General Hospital
57 Xinhua East Road, Lubei Street
Hebei
Tangshan, 063000
China
Phone: 86 15503631851
Email: drwusl@163.com

Abstract

Background: Baseline sleep duration is associated with cancer risk and cancer-specific mortality; however, the association between longitudinal patterns of sleep duration and these risks remains unknown.

Objective: This study aimed to elucidate the association between sleep duration trajectory and cancer risk and cancer-specific mortality.

Methods: The participants recruited in this study were from the Kailuan cohort, with all participants aged between 18 and 98 years and without cancer at baseline. The sleep duration of participants was continuously recorded in 2006, 2008, and 2010. Latent mixture modeling was used to identify shared sleep duration trajectories. Furthermore, the Cox proportional risk model was used to examine the association of sleep duration trajectory with cancer risk and cancer-specific mortality.

Results: A total of 53,273 participants were included in the present study, of whom 40,909 (76.79%) were men and 12,364 (23.21%) were women. The average age of the participants was 49.03 (SD 11.76) years. During a median follow-up of 10.99 (IQR 10.27-11.15) years, 2705 participants developed cancers. Three sleep duration trajectories were identified: normal-stable (44,844/53,273, 84.18%), median-stable (5877/53,273, 11.03%), and decreasing low-stable (2552/53,273, 4.79%). Compared with the normal-stable group, the decreasing low-stable group had increased cancer risk (hazard ratio [HR] 1.39, 95% CI 1.16-1.65) and cancer-specific mortality (HR 1.54, 95% CI 1.18-2.06). Dividing the participants by an age cutoff of 45 years revealed an increase in cancer risk (HR 1.88, 95% CI 1.30-2.71) and cancer-specific mortality (HR 2.52, 95% CI 1.22-5.19) only in participants younger than 45 years, rather than middle-aged or older participants. Joint analysis revealed that compared with participants who...
had a stable sleep duration within the normal range and did not snore, those with a shortened sleep duration and snoring had the highest cancer risk (HR 2.62, 95% CI 1.46-4.70).

Conclusions: Sleep duration trajectories and quality are closely associated with cancer risk and cancer-specific mortality. However, these associations differ with age and are more pronounced in individuals aged <45 years.

Trial Registration: Chinese Clinical Trial Registry ChiCTR–TNRC–11001489; http://tinyurl.com/2u89hrhx

(JMIR Public Health Surveill 2024;10:e50836) doi:10.2196/50836

KEYWORDS
sleep duration; aging; cancer risk; mortality; sleep; trajectory; adult

Introduction
Sleep, a body and brain repair agent, plays a vital role in development, growth, and metabolism [1]. Regular and high-quality sleep can promote brain waste clearance, memory recovery, immune health, and normal nutrient metabolism. Correspondingly, sleep disorders may seriously affect the health and quality of life of patients [2].

Unfortunately, sleep disorders are one of the most common issues in today’s society. The World Health Organization has reported that approximately 27% of the population experiences sleep disorders. Therefore, scientists have conducted extensive studies on sleep disorders, most of which have focused on rapid eye movement (REM) sleep behavior disorders [3] or sleep disorders caused by diseases such as obstructive sleep apnea (OSA) [4]. Recently, an authoritative study has identified sleep duration as 1 of the 6 key elements of sleep health [5]. Sleep duration is associated with the occurrence and prognosis of various diseases. Dong et al [6] reported that insufficient or excessive sleep duration can increase the risk of depression. In terms of metabolic health, decreased sleep duration is associated with a higher risk of and a higher severity score for metabolic syndrome [7]. Dynamic changes in sleep duration can also bring significant changes to the body. In experimental research, sleep-deprived mice experience disruptions in their blood sugar homeostasis, which is restored to normal when sleep is regained. Cohort studies have also confirmed this point [8]. A study on shift work showed a strong correlation between irregular sleep duration and type 2 diabetes, obesity, heart disease, and cancer [9]. Nevertheless, in most of these studies, the sleep duration of participants was assessed at a single time to determine their cancer risk or even mortality, which is biased. To the best of our knowledge, various factors affect sleep, including the external environment (sound, light, and air quality), depression, anxiety, and age [10]. When considering age, approximately 50% of the older population experiences sleep disorders. Changes in sleep patterns and circadian dysrhythmia are considered part of aging [11]. Sleep duration and quality significantly decrease with age. Surprisingly, compared with young people, lack of sleep in older people has less effect on performance [12]. Therefore, it is likely that a single sleep survey will estimate the risk of disease and death erroneously. For the first time, we measured the sleep duration trajectory by monitoring the sleep of participants from 2006 to 2010 and then elucidated the association of sleep duration trajectories with cancer risk and cancer-specific mortality.

Methods
Study Participants
All participants were from the Kailuan cohort, part of an ongoing prospective study in Tangshan, China (ChiCTR-TNRC-11001489, registered at the Chinese Clinical Trial Registry). As previously described [13], the Kailuan cohort was initiated in 2006, when Kailuan General Hospital collaborated with 11 other hospitals to conduct physical examinations on 101,510 participants; the examinations included clinical examinations, health questionnaires, imaging, and laboratory tests. Thereafter, participants were followed up and periodically resurveyed every 2 years, and relevant indicators were recorded [14]. In this study, to construct the sleep duration trajectory, we included individuals who consecutively participated in the physical examinations from 2006 to 2010 (n=57,927); 3492 participants with missing covariates such as snoring, sedentary lifestyle, marital status, education, waist circumference (WC), BMI, and other covariates were excluded. Furthermore, 254 participants with cancer or cancer at baseline and 908 participants with missing sleep duration information were excluded. Finally, 53,273 participants were included in this study (Figure 1).
Ethical Considerations
This study was conducted according to the principles of the Declaration of Helsinki and its revised version and was approved by the ethics committee of Kailuan General Hospital with the code sjtkyll-lx-2021(39). All participants agreed to participate in the study and provided written informed consent.

Exposure and Covariates
All participants were invited to the hospital to fill out the Pittsburgh Sleep Quality Index (PSQI) sleep questionnaire survey (Multimedia Appendix 1) [15]. They reported their sleep status over the past 6 months. Information about the participants’ average sleep duration (3-15 h, excluding naps) was extracted from the questionnaire as exposure. For participants who reported approximate sleep duration, the average value was calculated (for example, 6-7 h was considered 6.5 h). The covariates were age, sex, BMI, WC, marital status, education level, sedentary time, physical activity, smoking, drinking, hypertension, diabetes, snoring, salt intake, family history of tumors, high-sensitivity C-reactive protein (CRP), total cholesterol (TC), triglycerides (TG), and digestive system cancer–related factors (fatty liver, gallstones, cirrhosis, and hepatitis B virus infection). Multimedia Appendix 1, Table S1 presents the definitions of all covariates.

Outcome Assessment
The electronic medical records from provincial vital statistics offices, the Tangshan medical insurance system, and the Kailuan Social Security Information System were used to determine the time of tumor occurrence and death of participants as well as specific causes of death. Cancer was diagnosed via pathology or imaging. All diagnoses were recorded using the International Classification of Diseases, 10th revision (Multimedia Appendix 1). The follow-up time was from the date of completion of the investigation in 2010 until the occurrence of cancer, death, or the last follow-up date (December 31, 2021), whichever came first.

Statistical Analyses
Statistical analyses were conducted using SAS (version 9.4; SAS Institute) or R (version 4.2.0; R Foundation for Statistical Computing). All $P$ values were 2-sided, and a $P$ value of <.05 was considered statistically significant.

Self-reported sleep duration trajectories were constructed for participants from 2006 to 2010 using latent mixed modeling, which was performed using Proc Traj in SAS. The construction of the trajectories was gradual. First, we established 5 types of trajectories and then compared them with 4, 3, 2, and 1 trajectories. The Bayesian information criterion was used to evaluate the fit of each trajectory model. Furthermore, models with different functional forms were evaluated based on the significance levels of cubic, quadratic, and linear terms. Continuous variables with normal distributions were expressed as mean (SD) and compared between groups using a 1-way ANOVA. Data with skewed distributions were expressed as median (IQR) and compared using the Kruskal-Wallis test. Categorical variables were expressed as n (%); between-group comparisons were made using the chi-square test. At the same time, Bonferroni correction was used for multiple comparison corrections. After satisfying the proportional hazard ratio assumption, Cox proportional hazards regression models were used to describe the association of different trajectories with cancer risk and cancer-specific mortality. In the adjusted model, the covariates were adjusted. Furthermore, the effect of some potential factors on the risk of digestive system cancers was considered [14,16,17]. Therefore, we additionally adjusted for...
liver cirrhosis, hepatitis B, gallstones, and gallbladder polyps for digestive system cancers. Because significant differences were observed in sleep patterns and effects between young people and middle-aged and older people [18], we first divided the participants into young people (aged <45 years) and middle-aged and older people (aged ≥45 years) to conduct the study. After observing that sleep duration was more closely related in young people, we conducted more detailed analyses, including subgroup and sensitivity analyses, in participants aged <45 years. For subgroup analysis, previously reported potential modifiers, such as snoring, sex [19], BMI [20], and regular physical exercise [21], were selected. Simultaneously, to clarify the joint effect of sleep quality and sleep duration trajectory, snoring and sleep duration trajectory were combined and the participants were regrouped. In sensitivity analyses, participants who developed cancer within 1 year of follow-up, those with hepatitis B infection, those with regular physical activity, those with a family history of tumor, and those who used sleep medication at least once during the past month were excluded. Furthermore, the variables were adjusted for other components in the PSQI. To avoid bias owing to temporal changes, we also adjusted for time-varying covariates. Finally, to avoid overestimation of cancer risk by death as a competing event, we repeated the analysis using a competing risk model (Multimedia Appendix 1).

Results

Baseline Characteristics

A total of 53,273 participants were included in this study, of whom 40,909 (76.79%) were men and 12,364 (23.21%) were women. The average age of the participants was 49.03 (SD 11.76) years. During a median follow-up of 10.99 years, 2705 participants developed cancers (Multimedia Appendix 1, Table S2); 518 of them were younger than 45 years (Multimedia Appendix 1, Table S3) and 2187 were aged 45 years or older (Multimedia Appendix 1, Table S4). According to the sleep duration of participants from 2006 to 2010, we constructed 3 trajectory patterns. A total of 84.18% (44,844/53,273) of the participants were categorized into the normal-stable group (mean sleep duration range 7.32-7.46 h), 11.03% (5877/53,273) into the increasing median-stable group (mean sleep duration range 5.23-6.09 h), and 4.79% (2552/53,273) into the decreasing low-stable group (mean sleep duration range 5.01-7.48 h) (Figure 2). Among the participants overall (Table 1) and those younger than 45 years (Multimedia Appendix 1, Table S5), the increasing median-stable and decreasing low-stable groups were older, had higher WC and TC, and were more likely to drink, snore, and have fatty liver than the normal-stable group, but had similar BMI, CRP, and TG levels. However, among middle-aged and older participants (Multimedia Appendix 1, Table S6), those in the increasing median-stable and decreasing low-stable groups were older; had higher TC; were more likely to have a sedentary lifestyle, snore, have fatty liver, and consume a high-salt diet; and had lower TG, WC, and BMI compared with participants in the normal group. The Bonferroni correction indicated that when conducting pairwise comparisons among the 3 groups, there were differences in age, education, and alcohol use (Multimedia Appendix 1, Table S7).

Figure 2. Sleep duration trajectory patterns from 2006 to 2010.
Table 1. Basic characteristics of the participants by sleep duration trajectory.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Normal-stable (n=44,844)</th>
<th>Increasing median-stable (n=5877)</th>
<th>Decreasing low-stable (n=2552)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>48.51 (11.8)</td>
<td>52.04 (11.4)</td>
<td>51.32 (11.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45</td>
<td>17,674 (39.4)</td>
<td>1577 (26.8)</td>
<td>698 (27.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≥45</td>
<td>27,170 (60.6)</td>
<td>4300 (73.2)</td>
<td>1854 (72.6)</td>
<td></td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>34,323 (76.5)</td>
<td>4676 (79.6)</td>
<td>1910 (74.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BMI, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18.5 kg/m^2</td>
<td>17,555 (39.1)</td>
<td>2261 (38.5)</td>
<td>989 (38.8)</td>
<td>.13</td>
</tr>
<tr>
<td>18.5-23.9 kg/m^2</td>
<td>18,739 (41.8)</td>
<td>2552 (43.4)</td>
<td>1090 (42.7)</td>
<td></td>
</tr>
<tr>
<td>≥24 kg/m^2</td>
<td>8550 (19.1)</td>
<td>1064 (18.1)</td>
<td>473 (18.5)</td>
<td></td>
</tr>
<tr>
<td>Waist circumference (cm), median (IQR)</td>
<td>86.0 (80.0-93.0)</td>
<td>86.5 (80.0-93.0)</td>
<td>87.0 (81.0-93.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Married, n (%)</td>
<td>42,658 (95.1)</td>
<td>5479 (93.2)</td>
<td>2439 (95.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>College graduate or above, n (%)</td>
<td>10,175 (22.7)</td>
<td>1574 (26.8)</td>
<td>496 (19.4)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sedentary time ≥8 hours, n (%)</td>
<td>10,831 (24.2)</td>
<td>2205 (37.5)</td>
<td>606 (23.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Regular physical activity, n (%)</td>
<td>6027 (13.4)</td>
<td>1380 (23.5)</td>
<td>379 (14.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Smoker, n (%)</td>
<td>13,254 (29.6)</td>
<td>2527 (43)</td>
<td>734 (28.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Uses alcohol, n (%)</td>
<td>7479 (16.7)</td>
<td>1663 (28.3)</td>
<td>486 (19)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>17,615 (39.3)</td>
<td>2395 (40.8)</td>
<td>992 (38.9)</td>
<td>.08</td>
</tr>
<tr>
<td>Diabetes mellitus, n (%)</td>
<td>3570 (8)</td>
<td>502 (8.5)</td>
<td>226 (8.9)</td>
<td>.10</td>
</tr>
<tr>
<td>Snoring, n (%)</td>
<td>15,838 (35.3)</td>
<td>3412 (58.1)</td>
<td>950 (37.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Salt consumption, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 g/day</td>
<td>4047 (9)</td>
<td>760 (12.9)</td>
<td>197 (7.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>6-10 g/day</td>
<td>36,351 (81.1)</td>
<td>4050 (69)</td>
<td>2100 (82.3)</td>
<td></td>
</tr>
<tr>
<td>&gt;10 g/day</td>
<td>4433 (9.9)</td>
<td>1063 (18.1)</td>
<td>255 (10)</td>
<td></td>
</tr>
<tr>
<td>Family history of tumor, n (%)</td>
<td>1722 (3.8)</td>
<td>437 (7.4)</td>
<td>90 (3.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>C-reactive protein (mg/L), mean (SD)</td>
<td>2.16 (5.30)</td>
<td>2.17 (7.31)</td>
<td>2.32 (5.83)</td>
<td>.39</td>
</tr>
<tr>
<td>Total cholesterol (mmol/L), median (IQR)</td>
<td>4.9 (4.3-5.6)</td>
<td>5.0 (4.3-5.7)</td>
<td>5.0 (4.3-5.7)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Triglyceride (mmol/L), median (IQR)</td>
<td>1.3 (0.9-1.9)</td>
<td>1.28 (0.9-2.0)</td>
<td>1.3 (0.9-1.9)</td>
<td>.55</td>
</tr>
<tr>
<td>Fatty liver, n (%)</td>
<td>14,512 (32.4)</td>
<td>2197 (37.4)</td>
<td>907 (35.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gallstone disease, n (%)</td>
<td>978 (2.2)</td>
<td>146 (2.5)</td>
<td>59 (2.3)</td>
<td>.32</td>
</tr>
<tr>
<td>Cirrhosis, n (%)</td>
<td>337 (0.8)</td>
<td>47 (0.8)</td>
<td>24 (0.9)</td>
<td>.54</td>
</tr>
<tr>
<td>Hepatitis B virus infection, n (%)</td>
<td>1212 (2.7)</td>
<td>151 (2.6)</td>
<td>74 (2.9)</td>
<td>.68</td>
</tr>
</tbody>
</table>

Association Between Sleep Duration Trajectory and Cancer Risk

For all participants, no significant change in cancer risk was observed in the increasing median-stable group (HR 0.97, 95% CI 0.84-1.11); however, in the decreasing low-stable group, overall cancer risk for the participants increased significantly, by 39% (HR 1.39, 95% CI 1.16-1.65). Interestingly, after dividing the participants based on the age criterion (45 years), this trend was only present among young people (HR 1.88, 95% CI 1.30-2.71) rather than middle-aged or older participants (HR 1.17, 95% CI 0.96-1.42). Among participants younger than 45 years, those in the decreasing low-stable group were at an overall increased risk of cancer compared with those in the normal-stable group (HR 1.88, 95% CI 1.30-2.71) (Table 2). Among cancer types, participants in the decreasing low-stable group only showed an increased risk of lung cancer (HR 1.51, 95% CI 1.08-2.12), pancreatic cancer (HR, 3.08, 95% CI 1.05-9.08), and stomach cancer (HR 2.06, 95% CI 1.10-3.87) (Multimedia Appendix 1, Table S8). Similarly, the risk of lung cancer (HR 3.64, 95% CI 1.64-8.09), esophageal cancer (HR 4.92, 95% CI 1.08-12.16), pancreatic cancer (HR 7.16, 95% CI
1.83-31.62), liver cancer (HR 3.68, 95% CI 1.04-8.03), and digestive system cancer (HR 2.43, 95% CI 1.23-4.81) increased among participants younger than 45 years in the decreasing low-stable group and not among the middle-aged or older participants (Multimedia Appendix 1, Figure S1).

Table 2. Hazard ratios (HRs) for the association of sleep duration trajectory patterns with overall cancer risk.

<table>
<thead>
<tr>
<th>Sleep duration trajectory patterns</th>
<th>Cases/total</th>
<th>IR</th>
<th>Crude model HR (95% CI)</th>
<th>P value</th>
<th>Adjusted model HR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal-stable</td>
<td>2235/44,844</td>
<td>4.80</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Increasing median-stable</td>
<td>299/5877</td>
<td>4.93</td>
<td>1.03 (0.91-1.16)</td>
<td>.67</td>
<td>0.97 (0.84-1.11)</td>
<td>.62</td>
</tr>
<tr>
<td>Decreasing low-stable</td>
<td>171/2552</td>
<td>6.57</td>
<td>1.37 (1.17-1.60)</td>
<td>&lt;.001</td>
<td>1.39 (1.16-1.65)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age &lt;45 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal-stable</td>
<td>437/17,674</td>
<td>2.28</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Increasing median-stable</td>
<td>35/1577</td>
<td>2.03</td>
<td>0.89 (0.69-1.25)</td>
<td>.50</td>
<td>0.89 (0.59-1.36)</td>
<td>.75</td>
</tr>
<tr>
<td>Decreasing low-stable</td>
<td>31/698</td>
<td>4.12</td>
<td>1.81 (1.25-2.60)</td>
<td>.001</td>
<td>1.88 (1.30-2.71)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age ≥45 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal-stable</td>
<td>1798/27,170</td>
<td>6.56</td>
<td>Reference</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Increasing median-stable</td>
<td>264/340</td>
<td>6.08</td>
<td>0.93 (0.81-1.05)</td>
<td>.24</td>
<td>0.83 (0.71-1.02)</td>
<td>.10</td>
</tr>
<tr>
<td>Decreasing low-stable</td>
<td>140/1854</td>
<td>7.57</td>
<td>1.15 (0.97-1.37)</td>
<td>.11</td>
<td>1.17 (0.96-1.42)</td>
<td>.12</td>
</tr>
</tbody>
</table>

aIR: incidence rate per 1000 person-years.  
bThe model was adjusted for continuous (age, BMI, waist circumference, c-reactive protein, total cholesterol, triglycerides, and sleep duration in 2010) and categorical (sex, marital status, education level, sedentary time, physical activity, smoking, alcohol use, hypertension, diabetes, snoring, salt intake, and family history of tumor) variables.

Association Between Sleep Duration Trajectory and Cancer-Specific Mortality

Similar to the findings on sleep duration trajectories and cancer risk, for all participants, compared to the normal-stable group, individuals in the decreasing low-stable group had a significantly higher risk of cancer-specific mortality (HR 1.54, 95% CI 1.18-2.06). The same phenomenon was also observed among participants aged <45 years, where individuals in the decreasing low-stable group had a significantly higher risk of cancer-specific mortality (HR 2.52, 95% CI 1.22-5.19). However, no such association was found among participants aged ≥45 years (HR 1.28, 95% CI 0.94-1.75) (Table 3).

Table 3. Hazard ratios (HRs) for the association of sleep duration trajectory patterns with cancer-specific mortality.

<table>
<thead>
<tr>
<th>Sleep duration trajectory patterns</th>
<th>Cases/total</th>
<th>MR</th>
<th>Cancer-specific HR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal-stable</td>
<td>777/44,844</td>
<td>1.64</td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td>Increasing median-stable</td>
<td>115/5877</td>
<td>1.86</td>
<td>1.07 (0.86-1.33)</td>
<td>.71</td>
</tr>
<tr>
<td>Decreasing low-stable</td>
<td>65/2552</td>
<td>2.44</td>
<td>1.54 (1.18-2.06)</td>
<td>.01</td>
</tr>
<tr>
<td>Age &lt;45 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal-stable</td>
<td>108/17,674</td>
<td>0.56</td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td>Increasing median-stable</td>
<td>9/1577</td>
<td>0.52</td>
<td>0.80 (0.39-1.65)</td>
<td>.55</td>
</tr>
<tr>
<td>Decreasing low-stable</td>
<td>11/698</td>
<td>1.44</td>
<td>2.52 (1.22-5.19)</td>
<td>.01</td>
</tr>
<tr>
<td>Age ≥45 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal-stable</td>
<td>669/27,170</td>
<td>2.39</td>
<td>Ref.</td>
<td></td>
</tr>
<tr>
<td>Increasing median-stable</td>
<td>106/4300</td>
<td>2.39</td>
<td>1.01 (0.80-1.27)</td>
<td>.96</td>
</tr>
<tr>
<td>Decreasing low-stable</td>
<td>54/1854</td>
<td>2.84</td>
<td>1.28 (0.94-1.75)</td>
<td>.12</td>
</tr>
</tbody>
</table>

aMR: mortality rate per 1000 person-years.  
bThe model was adjusted for continuous (age, BMI, waist circumference, c-reactive protein, total cholesterol, triglycerides, and sleep duration in 2010) and categorical (sex, marital status, education level, sedentary time, physical activity, smoking, alcohol use, hypertension, diabetes, snoring, salt intake, and family history of tumor) variables.
Subgroup Analysis and Joint Analysis

We did not find any interactions between sleep duration trajectory and snoring (P for interaction .16), sex (P for interaction .80), overweight (P for interaction .50), and regular physical activity (P for interaction .90) (Multimedia Appendix 1, Table S9) in the subgroup analysis; however, in the joint analysis (Multimedia Appendix 1, Table S10), participants in the decreasing low-stable group were at the highest risk of cancer, especially those who snored (HR 2.62, 95% CI 1.46-4.70).

Additional Analyses

The competing risk model showed that after considering death as a competitive event, a closer association between sleep duration trajectory and cancer risk was observed. In both the cause-specific hazard functions (CSHF) model and the subdistribution hazard functions (SDHF) model, when compared to the normal stable group, participants in the decreasing low-stable group had an elevated risk of lung cancer (CSHF HR 2.42, 95% CI 1.07-5.53; SDHF HR 2.23, 95% CI 1.17-5.25), esophageal cancer (CSHF HR 4.07, 95% CI 1.12-14.89; SDHF HR 4.05, 95% CI 1.13-15.03), pancreatic cancer (CSHF HR 7, 95% CI 1.40-42.319; SDHF HR 6.88, 95% CI 1.38-27.6), liver cancer (CSHF HR 3.37, 95% CI 1.19-9.66; SDHF HR 3.08, 95% CI 1.08-10.40), digestive system cancers (CSHF HR 1.93, 95% CI 1.05-3.62; SDHF HR 1.99, 95% CI 1.23-5.53), and overall cancer risk (CSHF HR 2.01, 95% CI 1.39-3.17; SDHF HR 1.97, 95% CI 1.12-3.08) (Multimedia Appendix 1, Table S11). Finally, after performing multiple sensitivity analyses and adjusting for time-varying covariates, the results remained robust. Compared to the normal-stable group, participants in the decreasing low-stable group still had an increased risk of lung cancer, esophageal cancer, pancreatic cancer, liver cancer, digestive system cancers, and overall cancer risk (Multimedia Appendix 1, Table S12).

Discussion

Principal Findings

To the best of our knowledge, this is the first prospective study investigating age-related differences in the association of sleep duration trajectories with cancer risk and cancer-specific mortality. Compared with middle-aged and older individuals, young people in this study showed a closer association between sleep duration trajectories and the aforementioned risks. Despite some participants having baseline sleep durations within the normal range, once their sleep duration decreased, the aforementioned risks increased accordingly. Additionally, individuals with decreasing low-stable and normal-stable sleep duration trajectory and snoring had the highest risk of cancer.

Sleep duration is the most intuitive and concise indicator for evaluating sleep quality. A Mendelian randomization study from the UK Biobank showed that shorter sleep duration was associated with a higher risk of digestive tract cancers, such as stomach and pancreatic cancers [22], and most studies are in agreement with this study. In terms of cancer-related deaths, previous meta-analyses showed that compared with individuals who sleep for 7 to 8 hours, those who sleep for 4 to 5 hours were at an increased risk of cancer mortality [9]. Li et al. [23] found an association between short sleep time and a 24% increase in lung cancer mortality. However, there is no consensus due to the existence of contradictory results. A multicenter study from Japan showed that compared with 7 hours of sleep, less than 5 hours of sleep did not seem to affect the incidence rate or mortality of cancer [24]. Several meta-analyses also showed no relationship between sleep duration and breast cancer or overall cancer risk [25,26].

We speculate that the heterogeneity in these results may be because of population-specific factors. Svensson et al. [27] conducted one of the largest studies in Asia on the association between sleep duration and mortality, and the results indicated that age, an important modifier, greatly affected the association between sleep duration and prognosis. Another large prospective study was consistent with the abovementioned study, showing that shorter sleep duration increased the mortality rate in individuals younger than 65 years but not in those aged 65 years or older, and the effect of sleep duration on mortality was highest among younger individuals and negatively correlated with age [28]. These results are interesting and enlightening. We initially found an association between sleep duration trajectory and overall cancer risk in the entire population; however, the results differed significantly when the participants were stratified by age. Sleep duration changes only affected the cancer risk and mortality of participants younger than 45 years.

We hypothesize that this outcome is because of age-related differences in sleep patterns, because aging itself is probably accompanied by advanced sleep timing, shortened nocturnal sleep duration, and increased awakenings [29]. Therefore, changes in nocturnal sleep duration may have a lesser effect on older adults. Conversely, significant fluctuations in the sleep duration of young individuals may imply underlying issues in their physiological functioning. Mechanistically, regarding the central nervous system, as individuals age, neurotoxic substances accumulate, leading to cortical thinning and brain white-matter degeneration; additionally, the age-related dysregulation of neurotransmitters, including serotonin and adenosine, which are involved in sleep regulation, impairs the central sleep regulatory system [30]. Considering hormones and circadian rhythms, aging is closely related to a decrease in sex hormone levels. Studies have shown a significant correlation between decreased testosterone levels and shortened sleep duration and increased sleep fragmentation in older adults [31]. Cortisol, a hormone related to circadian rhythms, regulates sleep, and older adults often exhibit abnormalities in their cortisol rhythm, particularly an earlier peak in nocturnal cortisol levels, which leads to increased nocturnal awakenings and decreased slow-wave sleep [29]. The level of melatonin, an essential hormone that regulates circadian rhythms, differs depending on age, showing significantly reduced peak levels in older adults in comparison with younger individuals [32]. Therefore, fluctuations and changes in nocturnal sleep duration in older adults do not necessarily imply any disease but may simply be a normal aging-related progression. However, some studies have shown that changes in sleep patterns in older adults may increase the risk of lung cancer to some extent [33]. Nonetheless, these results do not indicate that sleep is unimportant in...
middle-aged and older adults. We can only cautiously conclude that compared with younger individuals, relying solely on sleep duration to assess the prognosis of middle-aged and older adults may lack sensitivity, and a comprehensive evaluation of their sleep patterns is crucial.

Compared with age, a factor that was associated with contradictory results, sleep quality is a critical factor. Zhang et al [19] reported similar results by conducting a 22-year prospective cohort study, finding that the effect of changes in sleep duration on colon cancer risk only existed in individuals who snored. Additionally, snoring can also increase the risk of physical weakness. A Chinese study showed that for participants with prefrailty, unhealthy sleep patterns such as snoring increased the risk of frailty by 42% [34]. In attempting to explore the underlying mechanisms, one study entered our field of view [35], which found that snoring or OSA can exacerbate intermittent hypoxia in mouse models and activate the key oxidative stress factor Bach1 in the body, thereby promoting the proliferation, invasion, and migration of lung cancer cells. Simultaneously, these factors promote the stemness of cancer cells and increase the degree of malignancy.

This study has the following limitations. First, the Kailuan cohort is a male-dominated cohort in northern China, which may not represent a general population because sleep differs across races and sexes [36]. Second, when evaluating sleep quality, we only included snoring, whereas REM sleep behavior disorder, restless leg syndrome, and OSA were not included in this study. Third, the information regarding sleep duration was obtained through questionnaires and participants’ recall; thus, the results may be affected by recall bias.

Conclusions

In conclusion, sleep duration trajectories are closely associated with the risk of cancers, as well as cancer-specific mortality. However, these associations vary with age and are more pronounced in individuals younger than 45 years. Furthermore, the joint analysis indicated that individuals in the decreasing low-stable sleep duration trajectory group who snored had the highest risk of cancer. The continuous monitoring of sleep duration may be of significant value in making diagnoses in young individuals.

Acknowledgments

This work was supported by the National Key Research and Development Program (2022YFC2009600; support was given to author HS). We would like to express our gratitude to all participants and staff involved in the Kailuan cohort.

SW, HS (shihp@ccmu.edu.cn), and LD (dengli070@foxmail.com) are cocorresponding authors of this manuscript.

Data Availability

The data sets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

Chenan Liu, TL, and Qingsong Z contributed to conceptualization; Chenning Liu and MS contributed to methodology; TL and MS contributed to software; Qi Zhang, Chenning Liu, and SL contributed to validation; Chenan Liu and TL contributed to formal analysis; HX and JS contributed to investigation; JR and XZ contributed to resources; Chenan Liu and SL contributed to data curation; Chen Liu, TL, and MS contributed to writing—original draft preparation; YC, and TL contributed to writing—review and editing; YC, LD, and XZ contributed to visualization; LD, SW, and HS contributed to supervision; SW contributed to project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Supplementary information on methods and supplementary tables and figures.

[DOCX File, 1026 KB - publichealth_v10i1e50836_app1.docx ]

References


Abbreviations

<table>
<thead>
<tr>
<th>CRP: c-reactive protein</th>
<th>CSHF: cause-specific hazard function</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR: hazard ratio</td>
<td>OSA: obstructive sleep apnea</td>
</tr>
<tr>
<td>PSQI: Pittsburgh Sleep Quality Index</td>
<td>REM: rapid eye movement</td>
</tr>
<tr>
<td>SDHF: subdistribution hazard function</td>
<td>TC: total cholesterol</td>
</tr>
<tr>
<td>TG: triglyceride</td>
<td>WC: waist circumference</td>
</tr>
</tbody>
</table>

Edited by A Mavragani; submitted 14.07.23; peer-reviewed by C Song, J Zhao, Y Han; comments to author 04.10.23; revised version received 21.10.23; accepted 13.12.23; published 07.02.24.

Please cite as:


Age Differences in the Association of Sleep Duration Trajectory With Cancer Risk and Cancer-Specific Mortality: Prospective Cohort Study

JMIR Public Health Surveill 2024;10:e50836
URL: https://publichealth.jmir.org/2024/1/e50836
doi:10.2196/50836
PMID:38324354

©Chenan Liu, Qingsong Zhang, Chenning Liu, Tong Liu, Mengmeng Song, Qi Zhang, Hailun Xie, Shiqi Lin, Jiangan Ren, Yue Chen, Xinzheng, JinYu Shi, Li Deng, Hangping Shi, Shouling Wu. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 07.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Arterial Stiffness and Obesity as Predictors of Diabetes: Longitudinal Cohort Study

Cancan Cui\textsuperscript{1*}, PhD; Zhiyuan Wu\textsuperscript{2*}, PhD; Jia Zheng\textsuperscript{1}, PhD; Jiaqi Chu\textsuperscript{3}, PhD; Jiajing Zhu\textsuperscript{1}, PhD; Yitian Qi\textsuperscript{1}, PhD; Zhenming Zhang\textsuperscript{1}, PhD; Zhijia Wang\textsuperscript{1}, PhD; Mengchao Zhang\textsuperscript{1}, PhD; Zhanhao Mo\textsuperscript{1}, PhD; Liangkai Cheng\textsuperscript{1}, PhD; Lei Ding\textsuperscript{1}, PhD; Xiaoping Kang\textsuperscript{4}, PhD; Lin Liu\textsuperscript{1}, PhD; Te Zhang\textsuperscript{5}, PhD

\textsuperscript{1}Department of Radiology, China-Japan Union Hospital of Jilin University, Jilin University, Changchun, China
\textsuperscript{2}Centre for Precision Health, School of Medical and Health Sciences, Edith Cowan University, WA, Australia
\textsuperscript{3}School of Medical Imaging, Dalian Medical University, Dalian, China
\textsuperscript{4}Center of Informatics, Beijing Xiaotangshan Hospital, Beijing, China
\textsuperscript{5}Department of Anesthesiology, China-Japan Union Hospital of Jilin University, Jilin University, Changchun, China

*these authors contributed equally

Corresponding Author:
Te Zhang, PhD
Department of Anesthesiology, China-Japan Union Hospital of Jilin University, Jilin University
No.126, Sendai Street, Erdao District
Changchun, 130022
China
Phone: 86 13894857383
Email: zhangte19@mails.jlu.edu.cn

Abstract

Background: Previous studies have confirmed the separate effect of arterial stiffness and obesity on type 2 diabetes; however, the joint effect of arterial stiffness and obesity on diabetes onset remains unclear.

Objective: This study aimed to propose the concept of arterial stiffness obesity phenotype and explore the risk stratification capacity for diabetes.

Methods: This longitudinal cohort study used baseline data of 12,298 participants from Beijing Xiaotangshan Examination Center between 2008 and 2013 and then annually followed them until incident diabetes or 2019. BMI (waist circumference) and brachial-ankle pulse wave velocity were measured to define arterial stiffness abdominal obesity phenotype. The Cox proportional hazard model was used to estimate the hazard ratio (HR) and 95% CI.

Results: Of the 12,298 participants, the mean baseline age was 51.2 (SD 13.6) years, and 8448 (68.7%) were male. After a median follow-up of 5.0 (IQR 2.0-8.0) years, 1240 (10.1%) participants developed diabetes. Compared with the ideal vascular function and nonobese group, the highest risk of diabetes was observed in the elevated arterial stiffness and obese group (HR 1.94, 95% CI 1.60-2.35). Those with exclusive arterial stiffness or obesity exhibited a similar risk of diabetes, and the adjusted HRs were 1.63 (95% CI 1.37-1.94) and 1.64 (95% CI 1.32-2.04), respectively. Consistent results were observed in multiple sensitivity analyses, among subgroups of age and fasting glucose level, and alternatively using arterial stiffness abdominal obesity phenotype.

Conclusions: This study proposed the concept of arterial stiffness abdominal obesity phenotype, which could improve the risk stratification and management of diabetes. The clinical significance of arterial stiffness abdominal obesity phenotype needs further validation for other cardiometabolic disorders.

(JMIR Public Health Surveill 2024;10:e46088) doi:10.2196/46088

KEYWORDS

arterial stiffness; baPWV; brachial-ankle pulse wave velocity; cohort analysis; obesity and abdominal obesity; type 2 diabetes
Introduction

The latest estimates showed that there were 537 million adults with diabetes worldwide in 2021, which was expected to reach 783 million by 2045 [1], leading to a heavy socioeconomic burden with huge costs for glucose therapy and complications treatment [2,3]. Therefore, early detection and prevention of diabetes are of great importance, which requires the precise identification of risk markers to promote the risk stratification and management of diabetes onset.

Multifactorial risk factor evaluation and management is a recommended strategy for preventing diabetes and its complications. Arterial stiffness is an age-related process that results from adverse changes in the structure and function of the elastic arterial vessel wall [4]. Previous studies have shown a close relationship between increased arterial stiffness and diabetes [5]. Arterial stiffness is an important risk factor for diabetes [6-8] and diabetes could accelerate the process of arterial stiffness, conversely [9]. At the same time, it is well known that obesity is commonly associated with a range of metabolic abnormalities, including insulin resistance, atherogenic dyslipidemia, and metabolic syndrome [10]. By far, obesity is the most important modifiable risk factor for diabetes onset and glucose intervention [11], formulating the term “diabesity” to describe the combined adverse health effects and close relationship between obesity and diabetes [12]. Obesity is defined as a common chronic disorder due to excess body fat and has become a global epidemic issue [13]. The diagnosis and classification of obesity are usually based on BMI and waist circumference for abdominal obesity. Of note, not all individuals with obesity are at the same risk of metabolic complications, leading to the concept of metabolic health obesity phenotype [14,15]. Although individuals with obesity in a metabolically healthy state have a relatively reduced risk of developing diabetes, they are still at a higher risk compared to the metabolically healthy normal-weight group [15]. Thus, the assessment of metabolic health obesity phenotype allows for better identification of people at different risk of diabetes onset. Of note, the metabolic health status assessment needs a complex physical examination and blood test. On the contrary, arterial stiffness measurement using pulse wave velocity is a simple, repeatable, and noninvasive process [16]. However, the combined effect of obesity and arterial stiffness status on diabetes onset has not been fully investigated to date.

Therefore, this study proposed the concept of arterial stiffness obesity phenotype and investigated the joint effect of increased arterial stiffness and abdominal obesity on incident diabetes. We hypothesized that arterial stiffness abdominal obesity phenotype could better stratify the risk of diabetes onset in the general population.

Methods

Study Population and Design

This study used data from the Beijing Xiaotangshan Examination Center [17], which is a large-scale longitudinal cohort study investigating the risk factors and biomarkers of cardiometabolic disorders. Beijing Xiaotangshan Examination Center included participants undergoing physical examinations from 2008, which is still ongoing and forms dynamic open cohort data with annual resurveys. In this study, a total of 12,505 participants aged 18 years or older who underwent the first comprehensive health examination without diabetes between 2008 and 2013 were primarily included as baseline. Then, we excluded 207 participants due to being underweight (n=126), being 18 years of age or younger (n=22), or lacking fasting glucose data (n=59) at baseline. Finally, 12,298 participants were selected and annually followed to incident diabetes or the end of 2019, including 8184 participants with available data on waist circumference as shown in Figure 1.
Figure 1. Flowchart of this study. This cohort study included 12,298 eligible participants from 2008 to 2019 in Beijing, China.

Data Collection and Definition

The data on demographic characteristics, lifestyle, disease history, and medication use were collected through a standard questionnaire and face-to-face interview. Educational level was categorized as illiteracy or primary school (primary), middle school or high school (secondary), and bachelor’s degree or above (third). Smoking and drinking status was divided into current or not. Physical activity was defined as having moderate or intense activity at work or during leisure time more than 4 times and 80 minutes each week. The disease history of dyslipidemia and hypertension was self-reported.

The physical examination and biochemical data were measured by the trained staff and acquired from the electronic record system of Beijing Xiaotangshan Hospital, including height, weight, waist circumference, and blood pressure. BMI was calculated as weight in kilograms divided by height in meters squared. Obesity was defined as BMI ≥28 kg/m², and abdominal obesity was defined as waist circumference ≥90 cm for male individuals and ≥85 cm for female individuals, using the standard for the Chinese population. The concentrations of fasting glucose and glycated hemoglobin A₁c (HbA₁c) were tested using the automatic biochemical analyzers Roche Cobas C 701 and SYSMEX HLC-723G8. Diabetes was defined as fasting glucose ≥7.0 mmol/L, HbA₁c ≥6.5%, or use of any glucose-lowering medication or self-reported diagnosis history of diabetes.

Arterial stiffness was measured by brachial-ankle pulse wave velocity (baPWV) using the Omron Colin BP-203RPE III device (Omron Health Care). Elevated arterial stiffness was defined as baPWV ≥1400 cm/s, as previously described. The ankle-brachial index (ABI) was calculated as the pressure ratio between posterior tibial artery and brachial artery.

Arterial Stiffness Abdominal Obesity Phenotype

Arterial stiffness obesity phenotype was divided into four groups: (1) normal arterial stiffness and no obesity (NASNO), (2) normal arterial stiffness and obesity (NASO), (3) elevated arterial stiffness and no obesity (ASNO), and (4) elevated arterial stiffness and obesity (ASO). On the other hand, waist circumference is another index to reflect the accumulation of abdominal fat, which is more closely associated with the risk of diabetes. Thus, arterial stiffness abdominal obesity phenotype was divided into four groups: (1) normal arterial stiffness and no abdominal obesity (NASNAO), (2) normal arterial stiffness and abdominal obesity (NASAO), (3) elevated arterial stiffness and no abdominal obesity (ASNAO), and (4) elevated arterial stiffness and abdominal obesity (ASAO).

Statistical Analysis

Baseline characteristics were described using mean (SD) and frequency (proportion) according to arterial stiffness abdominal obesity phenotype. The incidence rate and cumulative incidence of diabetes were calculated.

We used Kaplan-Meier curves to present the cumulative hazard of diabetes onset stratified by arterial stiffness obesity and arterial stiffness abdominal obesity phenotype. The effect of arterial stiffness (or obesity) on incident diabetes was evaluated stratified by obesity (arterial stiffness) status. Then, the Cox proportional hazard model was used to explore the longitudinal association of arterial stiffness abdominal obesity phenotype.
with incident diabetes. Hazard ratios (HRs) and 95% CIs were calculated in the following models: model 1 was primarily adjusted for age and sex, and model 2 was further adjusted for education level, physical activity, current smoking, current drinking, dyslipidemia or not, hypertension or not, mean arterial pressure (MAP), and fasting glucose concentration. We adjusted the level of MAP in the regression analysis as MAP was more dependent on baPWV level compared to systolic blood pressure and diastolic blood pressure. The interaction effect between arterial stiffness and abdominal obesity was tested as a multiplicative term. We performed multiple sensitivity analyses, including additionally adjusting HbA\textsubscript{1c} levels, excluding participants using antihypertensive medication, or repeating analyses among participants with an ABI >0.9. To address the dynamic changes of arterial stiffness abdominal obesity phenotype between baseline and follow-up, we repeated the analyses among participants of stable arterial stiffness abdominal obesity status. In addition, the effect of arterial stiffness abdominal obesity phenotype on diabetes onset was explored in subgroups of sex, age, and baseline fasting glucose level.

All statistical analyses were performed using R software (version 4.1.0; R Foundation for Statistical Computing), and a 2-sided \( P \) value <.05 was considered statistically significant.

**Ethical Considerations**

This study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Ethics Committee of the Beijing Xiaotangshan Center (XTS021431). The study data were anonymous. All participants provided written informed consent before taking part in this study.

**Results**

Of the 12,298 participants, the mean baseline age was 51.2 (13.6) years, and 8448 (68.7%) were male. Table 1 shows the detailed baseline characteristics according to arterial stiffness obesity phenotype, and characteristics according to arterial stiffness abdominal obesity phenotype are shown in Table S1 in Multimedia Appendix 1. During a median follow-up of 5.0 (IQR 2.0-8.0) years, diabetes occurred in 1240 participants. The incidence rates were 9.57, 27.83, 30.32, and 50.55 per 1000 person-years among NASNO, NASO, ASNO, and ASO groups (Table S2 in Multimedia Appendix 1) and 6.85, 20.04, 24.42, and 48.10 per 1000 person-years among NASNAO, NASAO, ASNAO, and ASAO groups, respectively (Table S3 in Multimedia Appendix 1).
Table 1. Baseline characteristics of 12,298 participants according to arterial stiffness obesity phenotype. This cohort study included eligible participants from 2008 to 2019 in Beijing, China. SI conversion factor: to convert fasting plasma glucose to mg/dL, multiply by 18.0.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>NASNO(^a) (n=5764)</th>
<th>NASO(^b) (n=1250)</th>
<th>ASNO(^c) (n=3952)</th>
<th>ASO(^d) (n=1332)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (N=12,298), n (%)</td>
<td>5764 (46.9)</td>
<td>1250 (10.2)</td>
<td>3952 (32.1)</td>
<td>1332 (10.8)</td>
</tr>
<tr>
<td>Sex (male), n (%)</td>
<td>3339 (57.9)</td>
<td>1014 (81.1)</td>
<td>3008 (76.1)</td>
<td>1087 (81.6)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>45.03 (9.56)</td>
<td>45.41 (9.77)</td>
<td>60.11 (13.98)</td>
<td>56.85 (13.45)</td>
</tr>
<tr>
<td>BMI(^e) (kg/m(^2)), mean (SD)</td>
<td>24.09 (2.29)</td>
<td>30.57 (9.62)</td>
<td>24.79 (2.08)</td>
<td>30.29 (9.73)</td>
</tr>
<tr>
<td>Waist(^f) (cm), mean (SD)</td>
<td>81.24 (8.56)</td>
<td>96.54 (6.97)</td>
<td>86.12 (7.21)</td>
<td>97.99 (7.38)</td>
</tr>
<tr>
<td>Abdominal obesity(^g), n (%)</td>
<td>928 (21.9)</td>
<td>875 (93.1)</td>
<td>866 (38.8)</td>
<td>740 (96.4)</td>
</tr>
<tr>
<td>Educational level, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>802 (13.9)</td>
<td>191 (15.3)</td>
<td>626 (15.8)</td>
<td>208 (15.6)</td>
</tr>
<tr>
<td>Secondary</td>
<td>3283 (57)</td>
<td>697 (55.8)</td>
<td>2293 (58)</td>
<td>771 (57.9)</td>
</tr>
<tr>
<td>Third</td>
<td>1679 (29.1)</td>
<td>362 (29)</td>
<td>1033 (26.1)</td>
<td>353 (26.5)</td>
</tr>
<tr>
<td>Physical activity, n (%)</td>
<td>2173 (37.7)</td>
<td>489 (39.1)</td>
<td>1550 (39.2)</td>
<td>477 (35.8)</td>
</tr>
<tr>
<td>Current smoking, n (%)</td>
<td>1703 (29.5)</td>
<td>426 (34.1)</td>
<td>1233 (31.2)</td>
<td>410 (30.8)</td>
</tr>
<tr>
<td>Current drinking, n (%)</td>
<td>3493 (60.6)</td>
<td>786 (62.9)</td>
<td>2357 (59.6)</td>
<td>832 (62.5)</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>74 (1.3)</td>
<td>21 (1.7)</td>
<td>216 (5.5)</td>
<td>65 (4.9)</td>
</tr>
<tr>
<td>MAP(^h),i (mmHg), mean (SD)</td>
<td>85.73 (9.32)</td>
<td>93.76 (9.43)</td>
<td>95.36 (10.25)</td>
<td>100.3 (10.76)</td>
</tr>
<tr>
<td>Fasting glucose (mmol/L), mean (SD)</td>
<td>5.14 (0.49)</td>
<td>5.41 (0.54)</td>
<td>5.38 (0.55)</td>
<td>5.61 (0.57)</td>
</tr>
<tr>
<td>HbA(_1)c(^j) (%), mean (SD)</td>
<td>5.42 (0.35)</td>
<td>5.53 (0.34)</td>
<td>5.56 (0.36)</td>
<td>5.64 (0.38)</td>
</tr>
</tbody>
</table>

\(^a\)NASNO: normal arterial stiffness and no obesity.
\(^b\)NASO: normal arterial stiffness and obesity.
\(^c\)ASNO: elevated arterial stiffness and no obesity.
\(^d\)ASO: elevated arterial stiffness and obesity.
\(^e\)Calculated as weight in kilograms divided by height in meters squared.
\(^f\)Waist circumference was measured only in 8184 participants.
\(^g\)Abdominal obesity was defined as a waist circumference of \(\geq 90\) cm for male individuals and \(\geq 85\) cm for female individuals. Due to data missing on waist circumference, the total numbers of NASNA, NASO, ASNO and ASO were 4237, 940, 2232 and 768, respectively.
\(^h\)MAP: mean arterial pressure.
\(^i\)MAP = \(1/3 \times \) systolic pressure + \(2/3 \times \) diastolic pressure.
\(^j\)HbA\(_1\)c: glycated hemoglobin A\(_1\)c.

Figure 2 presents the cumulative hazard of incident diabetes according to arterial stiffness obesity phenotype and arterial stiffness abdominal obesity phenotype. Of note, people with obesity had a significantly higher risk of diabetes onset regardless of arterial stiffness status (\(P<.001\); Figure S1A and B in Multimedia Appendix 1). Equally, those with elevated arterial stiffness had a higher risk of diabetes regardless of obesity status (Figure S1C and D in Multimedia Appendix 1). Figure S2A-D in Multimedia Appendix 1 present the mutual effect of arterial stiffness and abdominal obesity.
Figure 2. Kaplan-Meier curves of incident diabetes according to arterial stiffness obesity (left) phenotype and arterial stiffness abdominal obesity (right) phenotype. The left and right sides include 12,298 and 8184 eligible participants, respectively, from 2008 to 2019. ASAO: elevated arterial stiffness and abdominal obesity; ASNAO: elevated arterial stiffness and no abdominal obesity; ASNO: elevated arterial stiffness and no obesity; ASO: elevated arterial stiffness and obesity; NASAO: normal arterial stiffness and abdominal obesity; NASNAO: normal arterial stiffness and no abdominal obesity; NASNO: normal arterial stiffness and no obesity; NASO: normal arterial stiffness and obesity.

In the joint model, both continuous BMI (adjusted HR 1.06, 95% CI 1.04-1.07; \(P < .001\)) and baPWV (adjusted HR 1.18, 95% CI 1.11-1.25; \(P < .001\)) levels were significantly associated with diabetes onset. In the adjusted model, the highest risk of incident diabetes was observed in the ASO group (HR 1.94, 95% CI 1.60-2.35), followed by ASNO (HR 1.63, 95% CI 1.37-1.94) and NASO (HR 1.64, 95% CI 1.32-2.04) groups, compared with NASNO (Table 2). The interaction terms of general obesity and arterial stiffness were statistically significant \((P < .05)\). Similarly, ASAO had the highest risk (HR 2.30, 95% CI 1.71-3.11), followed by NASAO (HR 1.30, 95% CI 0.95-1.78) and ASNAO (HR 1.67, 95% CI 1.20-2.32) groups, compared with NASNAO.
Table 2. Association of arterial stiffness abdominal obesity phenotype with incident diabetes.

<table>
<thead>
<tr>
<th>Definition and variable</th>
<th>Model 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th></th>
<th>Model 2&lt;sup&gt;b&lt;/sup&gt;</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR&lt;sup&gt;c&lt;/sup&gt; (95% CI)</td>
<td>P value</td>
<td>HR (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td><strong>Definition 1&lt;sup&gt;d&lt;/sup&gt; (n=12,298)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous (per SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (SD 5.5 kg/m&lt;sup&gt;2&lt;/sup&gt;)</td>
<td>1.049 (1.034-1.065)</td>
<td>&lt;.001</td>
<td>1.055 (1.038-1.071)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>baPWV&lt;sup&gt;e&lt;/sup&gt; (SD 285 cm/s)</td>
<td>1.403 (1.347-1.461)</td>
<td>&lt;.001</td>
<td>1.176 (1.107-1.249)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Categorical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASNO&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>NASO&lt;sup&gt;f&lt;/sup&gt;</td>
<td>2.716 (2.191-3.366)</td>
<td>&lt;.001</td>
<td>1.642 (1.321-2.041)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ASNO&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2.248 (1.894-2.668)</td>
<td>&lt;.001</td>
<td>1.632 (1.373-1.940)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ASO&lt;sup&gt;i&lt;/sup&gt;</td>
<td>3.908 (3.246-4.705)</td>
<td>&lt;.001</td>
<td>1.935 (1.595-2.348)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>P value for interaction</td>
<td>N/A&lt;sup&gt;j&lt;/sup&gt;</td>
<td>&lt;.001</td>
<td>N/A</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Definition 2&lt;sup&gt;k&lt;/sup&gt; (n=8184)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous (per SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waist (SD 10.0 cm)</td>
<td>1.810 (1.657-1.977)</td>
<td>&lt;.001</td>
<td>1.817 (1.646-2.005)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>baPWV (SD 285 cm/s)</td>
<td>1.492 (1.400-1.591)</td>
<td>&lt;.001</td>
<td>1.263 (1.151-1.386)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Categorical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASNAO&lt;sup&gt;j&lt;/sup&gt;</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>NASAO&lt;sup&gt;m&lt;/sup&gt;</td>
<td>2.473 (1.826-3.349)</td>
<td>&lt;.001</td>
<td>1.299 (0.951-1.775)</td>
<td>.10</td>
</tr>
<tr>
<td>ASNAO&lt;sup&gt;n&lt;/sup&gt;</td>
<td>2.450 (1.783-3.366)</td>
<td>&lt;.001</td>
<td>1.671 (1.202-2.324)</td>
<td>.002</td>
</tr>
<tr>
<td>ASAO&lt;sup&gt;o&lt;/sup&gt;</td>
<td>4.707 (3.529-6.278)</td>
<td>&lt;.001</td>
<td>2.303 (1.705-3.113)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>P value for interaction</td>
<td>N/A</td>
<td>.04</td>
<td>N/A</td>
<td>.06</td>
</tr>
</tbody>
</table>

<sup>a</sup>Model 1: age and sex adjusted.

<sup>b</sup>Model 2: age, sex, education, physical activity, smoking, drinking, dyslipidemia, hypertension, mean arterial pressure, and fasting glucose level adjusted.

<sup>c</sup>HR: hazard ratio.

<sup>d</sup>Definition 1: arterial stiffness obesity phenotype is defined using brachial-ankle pulse wave velocity and BMI.

<sup>e</sup>baPWV: brachial-ankle pulse wave velocity.

<sup>f</sup>NASNO: normal arterial stiffness and no obesity.

<sup>g</sup>NASO: normal arterial stiffness and obesity.

<sup>h</sup>ASNO: elevated arterial stiffness and no obesity.

<sup>i</sup>ASO: elevated arterial stiffness and obesity.

<sup>j</sup>N/A: not applicable.

<sup>k</sup>Definition 2: arterial stiffness abdominal obesity phenotype is defined using brachial-ankle pulse wave velocity and waist circumference.

<sup>l</sup>NASNAO: normal arterial stiffness and no abdominal obesity.

<sup>m</sup>NASAO: normal arterial stiffness and abdominal obesity.

<sup>n</sup>ASNAO: elevated arterial stiffness and no abdominal obesity.

<sup>o</sup>ASAO: elevated arterial stiffness and abdominal obesity.

The results remained consistent even after adjusting for baseline HbA<sub>1c</sub> levels, excluding those using antihypertensive medication and participants with an ABI ≤0.9. Analyses among people with stable arterial stiffness abdominal obesity phenotype did not significantly change the results (Table 3).

Figure 3 presents the subgroup analyses of sex, age, and baseline fasting glucose level. Of note, only ASAO group had a significantly increased risk of diabetes among male and baseline glucose above 5.6 mmol/L compared with NASNAO.
Table 3. Adjusted regression results of sensitivity analysis. Analyses were adjusted for age, sex, education, physical activity, smoking, drinking, dyslipidemia, hypertension, mean arterial pressure, and fasting glucose level.

<table>
<thead>
<tr>
<th>Sensitivity and variable</th>
<th>Definition 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P value</th>
<th>Definition 2&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td></td>
<td>HR (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Sensitivity 1&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>NASNAO&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.953 (0.625-1.452)</td>
<td>.82</td>
<td>0.908 (0.574-1.436)</td>
<td>.68</td>
</tr>
<tr>
<td>NASAO&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.380 (0.993-1.919)</td>
<td>.06</td>
<td>1.488 (0.933-2.374)</td>
<td>.10</td>
</tr>
<tr>
<td>ASNAO&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.581 (1.103-2.266)</td>
<td>.01</td>
<td>1.763 (1.144-2.717)</td>
<td>.01</td>
</tr>
<tr>
<td>Sensitivity 2&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>NASAO</td>
<td>1.643 (1.306-2.066)</td>
<td>&lt;.001</td>
<td>1.305 (0.952-1.789)</td>
<td>.10</td>
</tr>
<tr>
<td>ASNAO</td>
<td>1.699 (1.416-2.039)</td>
<td>&lt;.001</td>
<td>1.669 (1.192-2.336)</td>
<td>.003</td>
</tr>
<tr>
<td>ASAO</td>
<td>2.019 (1.642-2.482)</td>
<td>&lt;.001</td>
<td>2.319 (1.709-3.149)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sensitivity 3&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>NASAO</td>
<td>1.790 (1.156-2.774)</td>
<td>.009</td>
<td>1.807 (0.974-3.353)</td>
<td>.06</td>
</tr>
<tr>
<td>ASNAO</td>
<td>1.494 (1.095-2.039)</td>
<td>.01</td>
<td>2.141 (1.143-4.013)</td>
<td>.02</td>
</tr>
<tr>
<td>ASAO</td>
<td>2.018 (1.442-2.824)</td>
<td>&lt;.001</td>
<td>3.052 (1.706-5.46)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sensitivity 4&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>NASAO</td>
<td>1.638 (1.317-2.038)</td>
<td>&lt;.001</td>
<td>1.316 (0.962-1.8)</td>
<td>.09</td>
</tr>
<tr>
<td>ASNAO</td>
<td>1.624 (1.364-1.933)</td>
<td>&lt;.001</td>
<td>1.698 (1.219-2.365)</td>
<td>.002</td>
</tr>
<tr>
<td>ASAO</td>
<td>1.952 (1.607-2.371)</td>
<td>&lt;.001</td>
<td>2.32 (1.713-3.141)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>Definition 1: arterial stiffness obesity phenotype is defined using brachial-ankle pulse wave velocity and BMI.
<sup>b</sup>Definition 2: arterial stiffness abdominal obesity phenotype is defined using brachial-ankle pulse wave velocity and waist circumference.
<sup>c</sup>HR: hazard ratio.
<sup>d</sup>Sensitivity 1: glycated hemoglobin A<sub>1c</sub> was additionally adjusted in the analysis; 3899 participants were enrolled in the current analysis under definition 1 and 3626 participants enrolled under definition 2.
<sup>e</sup>NASNAO: normal arterial stiffness and no abdominal obesity.
<sup>f</sup>NASAO: normal arterial stiffness and abdominal obesity.
<sup>g</sup>ASNAO: elevated arterial stiffness and no abdominal obesity.
<sup>h</sup>ASAO: elevated arterial stiffness and abdominal obesity.
<sup>i</sup>Sensitivity 2: participants using antihypertensive medication were excluded from the analysis; 11,325 participants were enrolled in the current analysis under definition 1 and 7836 participants enrolled under definition 2.
<sup>j</sup>Sensitivity 3: analyses performed among participants with stable arterial stiffness abdominal obesity phenotype at baseline and follow-up; 4818 participants were enrolled in the current analysis under definition 1 and 2905 participants enrolled under definition 2.
<sup>k</sup>Sensitivity 4: analyses performed among participants of ankle-brachial index >0.9; 12,161 participants were enrolled in the current analysis under definition 1 and 8122 participants enrolled under definition 2.
**Discussion**

**Overview**

In this study, we proposed the concept of arterial stiffness obesity phenotype using noninvasive and simple measurements. We found that arterial stiffness obesity phenotype could improve the risk stratification of diabetes onset. People with ASO had the highest risk of incident diabetes, independent of their baseline glucose level. The results remained consistent among multiple sensitivity analyses and subgroups of sex, age, and fasting glucose level.

Many studies have suggested a relationship between arterial stiffness and diabetes, indicating that arterial stiffness could predict the development of diabetes [5,6,8]. A Kailuan study of 14,159 participants found that arterial stiffness appeared to precede the increase in fasting glucose [7]. Notably, the independent association between arterial stiffness and diabetes was not consistent among all studies. A large prospective observational study of Japanese company employees found that atherosclerosis was observed only in those with diabetes combined with hypertension, and the increased arterial stiffness may not be associated with the onset of diabetes or prediabetes alone [20]. On the other hand, obesity is an important risk factor involved in the etiopathogenesis of diabetes and is the most important culprit of insulin resistance [21,22]. In a meta-analysis in the United States and Europe, men with obesity had a 7-fold higher risk of diabetes, and women with obesity had a 12-fold higher risk [23]. A longitudinal study from the United Kingdom investigated 369,362 participants aged between 2 and 15 years and found that most of the patients with diabetes were obese (47.1%) [24]. Similar results have been found in other countries [25]. Studies have shown that the distribution of adipose tissue is a key factor in the development of insulin resistance, independent of the stage of obesity [26]. A growing body of evidence suggests that obesity cannot be assessed by BMI. Individuals of normal weight but with excess visceral adipose tissue are at high risk of diabetes, while individuals with obesity who can expand their subcutaneous adipose tissue mass, especially in the hip and femoral regions, may be at much lower risk than expected [11]. Thus, both excess body weight (BMI) and ectopic fat (such as abdominal obesity) determine the risk of diabetes onset. This study supplemented the evidence about the combined effect of arterial stiffness and obesity or abdominal obesity on the incident diabetes using a larger cohort. Compared with the ideal vascular function and nonobese group, the highest risk of diabetes was observed in the elevated arterial stiffness with obesity or abdominal obesity group. Additionally, there exists a potential interaction effect between general obesity and arterial stiffness on diabetes onset.

Metabolic health obesity phenotype has been proposed as a health index [27]. In general, people with different metabolic health obesity phenotypes have a differentiated risk of type 2 diabetes, cardiovascular diseases, and all-cause mortality [28-32]. The number and severity of metabolic abnormalities could further identify the risk of adverse outcomes [33,34]. In terms of diabetes, data suggest that the risk of developing diabetes is 5-20 times higher in metabolically unhealthy people with obesity than in the metabolically healthy nonobese group, and the risk is 4 times higher for the metabolically healthy obese group [35]. However, the definition of metabolic health status needs physical examinations (blood pressure), blood test (glucose and lipid), and a questionnaire survey (disease history and medication use). In this study, we proposed the concept of arterial stiffness abdominal obesity phenotype to stratify people...
at different risks of diabetes onset. The measurements of arterial stiffness and obesity status are simple, fast, and noninvasive, which enhance the applicability of arterial stiffness abdominal obesity phenotype in the risk stratification and management of diabetes.

Potential mechanisms linking arterial stiffness and diabetes include endothelial dysfunction, chronic inflammation, oxidative stress, microvascular dysfunction, and shared genetic background [7,8,36]. First, endothelial dysfunction is associated with arterial stiffness [37]. Arterial stiffness may lead to increased arterial pulse pressure and pulsation shear, resulting in endothelial dysfunction and metabolic dysregulation [38]. It has been suggested that endothelial dysfunction can cause the development of diabetes, and there is a common pathway that may link arterial stiffness and endothelial dysfunction to the development of diabetes, or possibly that these 2 factors reinforce each other [8,39]. Second, arterial stiffness may lead to microvascular dysfunction, which in turn leads to damage to low-resistance organs (eg, the pancreas), resulting in reduced tissue perfusion, including insulin-mediated muscle perfusion. This will lead to impaired glucose metabolism, insulin resistance, and an elevated fasting glucose level [7,40]. In this process, endothelial dysfunction and impaired endothelium-dependent vasodilation may exacerbate insulin resistance by limiting glucose delivery to key target tissues [41]. Third, increased oxidative stress and chronic low-grade inflammation may be common risk factors for atherosclerosis and diabetes [42,43].

The mechanisms underpinning the relationship of obesity with diabetes are only partially understood. Most of the hypotheses about obesity causing diabetes in recent years have been based on the coexistence of insulin resistance. Randle et al [44], for the first time, explained the relationship between obesity and diabetes by the “glucose-fatty acid cycle,” proposing a theory that obesity inhibits the glycolytic enzymes pyruvate dehydrogenase, phosphofructokinase, and hexokinase, thus causing an imbalance in glucose metabolism [45]. Another hypothesis is that adipose tissue is a secretory organ that produces and releases a variety of factors that may lead to insulin resistance. Most of the data suggested that tumor necrosis factor alpha (TNF-α) plays a mediating role [46]. Upregulated TNF-α induces multiple adverse effects, such as impaired insulin signaling and inhibition of glucose transporter type 4 expression, which inhibits glucose uptake [47]. TNF-α could reduce the expression of lipocalin, a protein that is abundantly expressed in adipocytes and has direct anti diabetic and antiatherosclerotic effects [48]. In addition, there are several hypotheses about signaling pathways of adipose tissue inflammation [49], endoplasmic reticulum stress [50], oxidative stress [51], and accumulation of immune cells [52], which are related to insulin resistance and insulin secretion.

This cohort study proposed the concept of arterial stiffness abdominal obesity phenotype to stratify the risk of incident diabetes. However, the results should be interpreted in the context of limitations. First, baPWV measures the stiffness of both the elastic aorta and muscular artery, and other index of arterial stiffness such as carotid-femoral pulse wave velocity was not collected in this study. Second, this was an observational study design, and we were unable to claim the causal effect of arterial stiffness obesity groups on diabetes onset. Third, although we adjusted for the important confounding factors, including fasting glucose level, there was still a possibility of residual confounding bias, such as dietary factors. The observed results require further validation in other populations.

**Conclusions**

The findings indicated the combined effect of arterial stiffness and obesity status on diabetes onset, independent of fasting glucose level. This study proposed the concept of arterial stiffness abdominal obesity phenotype, providing a noninvasive and simple panel for the risk stratification and potential management of diabetes.

**Data Availability**

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

**Authors' Contributions**

TZ and LL are co-corresponding authors for this study.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Supplementary tables and figures.

[DOC File, 173 KB - publichealth_v10i1e46088_app1.doc]

**References**


**Abbreviations**

- **ABI:** ankle-brachial index
- **ASAO:** elevated arterial stiffness and abdominal obesity
- **ASNAO:** elevated arterial stiffness and no abdominal obesity
- **ASNO:** elevated arterial stiffness and no obesity
- **ASO:** elevated arterial stiffness and obesity
- **baPWV:** brachial-ankle pulse wave velocity
- **HbA1c:** glycated hemoglobin A1c
- **HR:** hazard ratio
- **MAP:** mean arterial pressure
- **NASAO:** normal arterial stiffness and abdominal obesity
- **NASNAO:** normal arterial stiffness and no abdominal obesity
- **NASNO:** normal arterial stiffness and no obesity
- **NASO:** normal arterial stiffness and obesity
- **TNF-α:** tumor necrosis factor alpha

©Cancan Cui, Zhiyuan Wu, Jia Zheng, Jiaqi Chu, Jiajing Zhu, Yitian Qi, Zhenming Zhang, Zhijia Wang, Mengchao Zhang, Zhanhao Mo, Liangkai Cheng, Lei Ding, Xiaoping Kang, Lin Liu, Te Zhang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 08.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
The Effect of the COVID-19 Pandemic on Non–COVID-19 Deaths: Population-Wide Retrospective Cohort Study

Abraham Ka-chung Wai1,2,3*, MBA, JD, MD; Tsz Fung Yip4*, BEng; Yui Hang Wong4, BSc; Chun Kit Chu1, MStat; Teddy Lee1, BPharm; Ken Hung On Yu1, PhD; Wang Leong So1, BPharm; Janet Y H Wong5, BNurs, PhD; Carlos King-ho Wong6, PhD; Joshua W Ho4*, PhD; Timothy Rainer1*, MD

1Department of Emergency Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
2Department of Accident & Emergency, The University of Hong Kong–Shenzhen Hospital, Shenzhen, China
3Department of Accident & Emergency, Queen Mary Hospital, Hong Kong, China (Hong Kong)
4School of Biomedical Sciences, The University of Hong Kong, Hong Kong, China (Hong Kong)
5School of Nursing & Health Studies, Hong Kong Metropolitan University, Kowloon, China (Hong Kong)
6Department of Family Medicine and Primary Care, The University of Hong Kong, Hong Kong, China (Hong Kong)

*these authors contributed equally

Corresponding Author:
Joshua W Ho, PhD
School of Biomedical Sciences
The University of Hong Kong
L4-44, Laboratory Block
21 Sassoon Road
Hong Kong
China (Hong Kong)
Phone: 852 6518 4979
Email: jwkho@hku.hk

Abstract

Background: Health care avoidance in the COVID-19 pandemic has been widely reported. Yet few studies have investigated the dynamics of hospital avoidance behavior during pandemic waves and inferred its impact on excess non–COVID-19 deaths.

Objective: This study aimed to measure the impact of hospital avoidance on excess non–COVID-19 deaths in public hospitals in Hong Kong.

Methods: This was a retrospective cohort study involving 11,966,786 patients examined between January 1, 2016, and December 31, 2021, in Hong Kong. All data were linked to service, treatment, and outcomes. To estimate excess mortality, the 2-stage least squares method was used with daily tallies of emergency department (ED) visits and 28-day mortality. Records for older people were categorized by long-term care (LTC) home status, and comorbidities were used to explain the demographic and clinical attributes of excess 28-day mortality. The primary outcome was actual excess death in 2020 and 2021. The 2-stage least squares method was used to estimate the daily excess 28-day mortality by daily reduced visits.

Results: Compared with the prepandemic (2016-2019) average, there was a reduction in total ED visits in 2020 of 25.4% (548,116/2,142,609). During the same period, the 28-day mortality of non–COVID-19 ED deaths increased by 7.82% (2689/34,370) compared with 2016-2019. The actual excess deaths in 2020 and 2021 were 3143 and 4013, respectively. The estimated total excess non–COVID-19 28-day deaths among older people in 2020 to 2021 were 1958 (95% CI 1100-2820; no time lag). Deaths on arrival (DOAs) or deaths before arrival (DBAs) increased by 33.6% (1457/4336) in 2020, while non–DOA/DBAs increased only by a moderate 4.97% (1202/24,204). In both types of deaths, the increases were higher during wave periods than in nonwave periods. Moreover, non-LTC patients saw a greater reduction in ED visits than LTC patients across all waves, by more than 10% (non-LTC: 93,896/363,879, 25.8%; LTC: 7,956/67,090, 11.9%). Most of the comorbidity subsets demonstrated an annualized reduction in visits in 2020. Renal diseases and severe liver diseases saw notable increases in deaths.

Conclusions: We demonstrated a statistical method to estimate hospital avoidance behavior during a pandemic and quantified the consequent excess 28-day mortality with a focus on older people, who had high frequencies of ED visits and deaths. This study serves as an informed alert and possible investigational guideline for health care professionals for hospital avoidance behavior and its consequences.
Introduction

The COVID-19 pandemic has posed challenges to population health beyond infection. On one hand, viral infection is a major health threat to vulnerable patients; on the other hand, among all patients, hospital non–COVID-19 mortality increased significantly. A retrospective, multicenter study in the United States reported that 30-day risk-adjusted non–COVID-19 mortality has increased by more than 20% since the pandemic [1].

Emergency departments (EDs) play an important role in the management of patients infected with COVID-19 and those with other medical emergencies. It is the first point of contact for many COVID-19 patients. EDs play an important role in diagnosis, treatment, and infection control (by isolation or hospitalization, if necessary). However, EDs manage general patients at the same time, meaning that they have to reserve capacity to manage patients with other medical emergencies.

A significant reduction in ED visits has been observed worldwide since the COVID-19 pandemic [1-4]. This reduction may be attributed to either active efforts to avoid contracting the virus in health care facilities or to passive avoidance caused by mandatory lockdown measures [3-5]. Most quantitative studies have reported excess mortality associated with reduced ED visits, although to different degrees [2], and only a few inferred the causative effect [6]. Disease-specific analysis with simpler pandemic dynamics of both excess and reduced mortality have also been reported [7,8], yet the relationships across gender, age groups, and comorbidities remain a significant knowledge gap.

Delayed medical visits and subsequent delayed diagnosis and extended symptomatic periods are most often associated with a time-dependent progression [9-11]. A meta-analysis of delayed cancer care during the pandemic involving 62 studies identified 38 different categories of delays and disruptions with impacts on treatment, diagnosis, or general health services. However, the most frequent determinants for disruptions were provider- or system-related and were caused by reductions in service availability [12-15] rather than patient-related hospital avoidance.

In our previous work, we identified a significant reduction in ED visits in the first 8 months of the COVID-19 pandemic that was associated with an increase in deaths certified in the ED and 28-day mortality in Hong Kong [16,17], despite the service in EDs never being disrupted or reduced. However, the causal effect of this observation has not been established. In this territory-wide retrospective cohort study, we measured the impact of hospital avoidance among ED patients on excess mortality using ED patient data from 2016 to 2021, a period during which Hong Kong experienced a unique COVID-19 pandemic with 4 distinct waves.

Methods

Participants and Source of Data

We performed a territory-wide retrospective cohort study using data from the Clinical Data Analysis and Reporting System (CDARS) [18], an administrative clinical database managed by the Hospital Authority of the Hong Kong Special Administrative Region, China. CDARS includes patients’ demographics, death data, diagnoses, procedures, drug prescriptions, dispensing history, and laboratory results from all public hospitals and clinics in Hong Kong. It contains inpatient and outpatient data from over 90% of the 7.47 million people served by the Hong Kong Hospital Authority [19]. The accuracy of diagnostic codes for case identification in CDARS has been validated in previous reports [20-22] with good reliability [23,24].

Full-scale emergency medicine services providing 24-hour, emergency, physician-led care are only available in 18 public hospitals under the hospital authority. In the nonpandemic years of 2016-2019, the total number of visits at these 18 EDs was 8.6 million, equivalent to an average of 287 visits per 1000 population yearly [25].

All patients attending the EDs of the 18 public hospitals in Hong Kong in the prepandemic period of January 1, 2016, to December 31, 2019, and the pandemic period of January 1, 2020, to December 31, 2021, were included. Data extraction took place on February 1, 2022, so all patients were followed up for at least 28 days.

To avoid confounding related to visits by COVID-19–positive patients, all COVID-19–positive visits and inpatient episodes were excluded from analysis. Daily COVID-19 case numbers and confirmed COVID-19 death numbers were obtained from an actively maintained online repository. Data included sex, age, long-term care home (LTC) resident status, ED visit service data, International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes, and treatments. Among all records, there were 3716 ED visits recorded with positive polymerase chain reaction (PCR) testing results for COVID-19, and these records were excluded from the analysis.

Ethical Considerations

All patient record entries are anonymized. The institutional review board of the University of Hong Kong/Hospital Authority West Cluster approved the study (UW 20-112) and granted a waiver of participant consent.

Visit Reduction and 28-Day Mortality

Deaths within 28 days of the last ED visit are defined as 28-day mortality. This was imputed by death registration minus visit
date, which must be less than or equal to 28 days. Daily 28-day mortality counts were tallied by the deceased patients’ last visit dates. As visit reduction was hypothesized to take place prior to subsequent excess mortality, counting deaths by visit dates (instead of simply deaths by dates) is preferable. The percentage change in ED visits was analyzed for fair comparison across age and sex strata in each COVID-19 wave and after-wave period.

**Excess 28-Day Mortality Estimation by Causal Inference**

In this study, we used the 2-stage least squares (2SLS) method [26] to estimate the daily excess 28-day mortality from daily reduced visits. The 2SLS is a common tool for causal inference. Others have characterized the extent of excess deaths by regression [27], difference-in-difference [28], and propensity matching [29]. Yet these methods cannot infer causation between attendance and death in a population. The 2SLS model is visualized as follows (Multimedia Appendix 1, Figure S1 contains further details):

Five binary instrumental variables (IV) or Z were introduced to model the effect of COVID-19 waves on visit reduction: after-wave and wave 1 to wave 4. After-wave IV is 1 for dates after each wave and excludes all COVID-19 waves; otherwise it is 0. IV for waves 1 to 4 are 1 for dates within a COVID-19 wave; otherwise they are 0. Waves 1, 2, 3, and 4 were designated with reference to government actions (Multimedia Appendix 1, Table S1), which were usually implemented at the onset and recession of COVID-19 waves; hence, these act as pivotal timestamps of pandemic severity.

On validity of the wave periods as IVs, calendar time has been reported to be a valid IV [6,7] when dramatic changes in practice occur in a relatively short period of time. As calendar time has been used in COVID-19 literature as an IV [8-10], we believe that government announcements, policy changes, and surges in COVID-19 case numbers from nonwave to wave periods in Hong Kong can also be regarded as drastic changes and are not seasonal (ie, have no L/U-Z association), making the wave periods a valid IV. We also believe that by excluding COVID-19–positive patients from the study population, the association between COVID-19 and 28-day death numbers (ie, Z-Y) has been minimized.

Age and sex strata with mean daily 28-day mortality from 2016 to 2021 less than or equal to 1 were excluded from the causal inference analysis due to model instability. The strata were the age groups 0-17, 18-34, and 35-44 years and both sexes (male and female).

We used the ivreg package [11,30] of R (R Foundation for Statistical Computing) to implement 2SLS. First, the exposure (daily number of ED visits) was regressed on the 5 binary IVs: \( \beta_1 \) to \( \beta_5 \). Then, we regressed ead (death, i), that is, ED 28-day mortality, with \( i \) lagged days on the predicted exposure, (ie, attendance), which yielded the local average treatment effect (LATE) among the compliers of each age and gender stratum. In both stages, we added 4 terms for seasonality adjustment: year (2016 was 0), \( t \) (number of days since the first day of a year), squared \( t \), and cubic \( t \).

Time lag \( i \) was also introduced to investigate the lagging effect of visit reduction on future 28-day deaths. For example, the excess death coefficient, \( \theta_1 \), can be interpreted as follows: if \( \theta_1 \) is \(-0.02\) for a population at a time lag of 5 days, there is a reduction of 100 ED visits on a certain day that caused 2 excess deaths 5 days away. Multimedia Appendix 1, Figures S2A and 2B show the dynamics of visits and deaths estimated by 2SLS models in tandem with COVID-19 waves.

To establish the validity of using COVID-19 waves as instrumental variables, a partial \( F \) statistic test was conducted between instruments and exposure [31] (visit with no time lag).

**Individual-Level Longitudinal Analysis of the Incidence Rate of Deaths on Arrival or Deaths Before Arrival Among Older People**

Deaths on arrival (DOAs) or deaths before arrival (DBAs) among 2 groups of people (COVID-19 and pre–COVID-19) aged 65 years or older were identified by patient IDs and filtered by those who visited at least twice in their respective control periods. The COVID-19 DOA/DBA group of older people visited EDs from between July 1, 2020, and July 1, 2021. Changes in their individual incidence rates (daily ED visits) between January 1, 2019, and January 24, 2020, (the control period) and between January 25, 2020, (the pandemic start), and June 30, 2020 (the treatment period), were calculated. The pre–COVID-19 DOA/DBA group of older people visited EDs between July 1, 2018, and July 1, 2019. Changes in their individual incidence rates between January 1, 2017, and June 30, 2017 (the control period), and January 1, 2018, and June 30, 2018 (the treatment period), were calculated.

**DOA/DBAs and Comorbidities Among Older People**

To further investigate clinical attributes of the excess 28-day mortality among older patients, we divided patient mortality in EDs into DOA/DBAs and non-DOA/DBAs.

Among DOA/DBAs, most had no ICD-9-CM diagnosis code or were simply diagnosed as, for example, cardiac arrest, so their comorbidities cannot be inferred meaningfully. For non-DOA/DBAs, their diagnostic codes were mapped to Deyo comorbidities using the R package icd [32]. Episode ED admissions and 28-day mortality were counted for each selected comorbidity.

**Results**

We examined a total of 11,966,786 ED visit records from January 1, 2016, to December 31, 2021, among which there were 212,288 (1.8%) patient deaths within 28 days of their last visit to the ED. We found that 28-day ED mortality was 43,335 in 2020 and 44,205 in 2021, compared to the 2016-2019 average of 40,192 (Table 1).
Table 1. Yearly changes in absolute numbers of and ratios between total 28-day emergency department mortalities and registered all-cause deaths in Hong Kong government censuses; the values include the 2016-2019 average and values for 2020 and 2021.

<table>
<thead>
<tr>
<th>Time period</th>
<th>28-day ED\textsuperscript{a} mortalities, n</th>
<th>Increase in 28-day ED mortalities vs 2016-2019 average, n</th>
<th>Registered all-cause deaths, n</th>
<th>Increase in registered all-cause deaths vs 2016-2019 average, n</th>
<th>28-day ED mortalities/registered all-cause deaths, %</th>
<th>Increase in 28-day ED mortalities/registered all-cause death vs 2016-2019 average, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-2019 average</td>
<td>34,370</td>
<td>_b</td>
<td>47,523</td>
<td>—</td>
<td>72.3</td>
<td>—</td>
</tr>
<tr>
<td>2020</td>
<td>37,059</td>
<td>2689</td>
<td>50,666</td>
<td>3143</td>
<td>73.1</td>
<td>0.8</td>
</tr>
<tr>
<td>2021</td>
<td>37,662</td>
<td>3292</td>
<td>51,536</td>
<td>4134</td>
<td>73.1</td>
<td>0.8</td>
</tr>
</tbody>
</table>

\textsuperscript{a}ED: emergency department.

\textsuperscript{b}Not applicable.

Visit Reductions and 28-Day Mortality

Compared with the average in 2016-2019, there was a reduction in total ED visits in 2020 of 25.4% (548,116/2,142,609). Total ED visits were 1,594,493 in 2020 and 1,812,703 in 2021, compared to the 2016-2019 average of 2,142,609 (Multimedia Appendix 1, Table S2). During the same period, the 28-day mortality of non–COVID-19 ED deaths increased by 7.82% (3143/40,192) compared with 2016-2019 (Table 1). The COVID-19 pandemic progressed in Hong Kong through 4 distinct waves, with the fourth wave lasting until February 18, 2021, resulting in 149 COVID-19 deaths in 2020 and 64 in 2021. ED visits in 2016-2019 was roughly stable throughout the year, while in 2020 and 2021 they fluctuated and mostly showed noticeable decreases at the start of each wave and subsequent resurgences (Figures 1A and 1B). These decreases in ED visits in 2020 and 2021 generally correspond to an increase in excess ED deaths, particularly during the first and fourth waves (Figures 1E and 1F).

The number of ED visits and the 28-day mortality of ED patients in 2016-2019 and in 2020 and 2021 are shown in Figures 1C to 1F. Compared with the 2016-2019 average value in Figure 1C, the number of visits in 2020 and 2021 demonstrated an observable reduction, to between 3000 and 4000 daily, during each wave of COVID-19, followed by a gradual rebound to a higher visit number until the next wave. With relatively improved infection control in 2021, the number of visits reached
close to prepandemic levels in late 2021 (Figure 1E). The 28-day mortality in 2020 roughly followed the pattern of the 2016-2019 average until the fourth wave, during which the mortality reached its peak throughout the study period (Figure 1D and 1F). The median ED waiting time from entry to reaching a cubicle in 2020 displayed similar dynamics as the ED visit curve (Multimedia Appendix 1, Figure S3).

**Excess 28-day Mortality Estimation by Causal Inference**

Next, the reduced ED visits and 28-day excess mortality across age and sex groups were characterized. Visits in each wave period were adjusted to same-period 2016-2019 data, while after-wave period data across 2020 and 2021 were summed and adjusted. In all adult groups, female patients saw a larger reduction in annualized visits than male patients of the same age in every wave (Figure 2A). However, it is also notable that in the 55-to-64 and ≥65-year age groups, the patients were less affected by further waves with time. ED visits and excess 28-day deaths across age strata with percentage changes from the 2016-2019 average and 2020 and 2021 are shown in Figure 2B. The reduction of ED visits was attenuated in 2021 when compared with 2020, but it was not restored to the 2016-2019 average.

**Figure 2.** Seasonally adjusted changes in Hong Kong public hospital emergency department visits and 28-day mortality, segmented by age and sex, from 2016 to 2019 and in 2020 and 2021, per wave period and yearly. (A) Line plots of seasonally adjusted visit change percentage among age and sex groups across waves, with after-wave periods shown by dashed lines (see Methods). The seasonally adjusted visit change percentage is calculated as the same-period change percentage against the 2016 to 2019 average. (B) Emergency department visits and excess 28-day deaths across age strata with percentage changes between 2020 and the 2016 to 2019 average, as well as 2021 vs the 2016 to 2019 average. Numerical values are percentage changes from the 2016 to 2019 average to the annual sum of each year (2020 and 2021). (C) Line plots of excess 28-day mortality per 100 reduced emergency department visits among 45-54, 55-64, and ≥65 year age groups and sex groups (95% CIs are shown by the shaded areas) across 0-to-14-day delays. This was estimated by instrumental variable analysis, with wave and after-wave periods as instrument variables, emergency department attendance as exposure, and emergency department 28-day deaths as outcome.

Although the ≥65-year age group had the highest absolute numbers of ED visits in the study period (3,849,923 episodes; Multimedia Appendix 1, Table S1), the 0-to-17-year age group had by far the largest reduction in ED visits in 2020 and 2021 from the 2016-2019 average (range −49.4% to −59.3%, n=−65,232/132,330; −102,411/172,781; Figure 2B and Multimedia Appendix 1, Table S3).

Apart from being the largest contributor to the total number of ED visits, the older group also saw the largest excess 28-day mortality in 2020 and 2021 (Figure 2C). Older male patients had a similar excess 28-day mortality rate (with a 95% CI greater than 0) to female patients, with an immediate (no time lag) excess 28-day death rate of 0.953 (95% CI 0.382-1.52) per 100 reduced ED visits among male patients and 1.15 (95% CI 0.382-1.52) per 100 reduced ED visits among female patients and 1.15 (95% CI 0.382-1.52).
0.763-1.54) among female patients (Multimedia Appendix 1, Table S4). All partial $F$ statistics had a $P$ value <.05.

Multimedia Appendix 1, Table S5 provides estimates of older male and female total excess deaths by visit reduction in 2020 and 2021. The estimated total excess non–COVID-19 28-day deaths due to reduced ED visits among older people throughout 2020 and 2021 was 1958 (95% CI 1100-2820; no time lag). The actual excess deaths in 2020 and 2021 were 3143 and 4013, respectively, with the 2016-2019 average in the census [33,34] as the benchmark.

**Individual-Level, Longitudinal Analysis of Incidence Rate of DOA/DBAs Among Older People**

The ED excess 28-day mortality of older patients can be categorized as DOA/DBAs and non-DOA/DBAs. DOA/DBAs increased by 1457 or 35.1% in 2020, while non-DOA/DBA mortalities increased by only a moderate 1202 or 4.65% (Multimedia Appendix 1, Table S6). Table 2 provides a breakdown of the changes in both DOA/DBAs and non-DOA/DBAs; the increases were higher during wave periods than in nonwave periods. The corresponding results in 2021 generally follow the same trends as the ones in 2020.

Results in Table 3 show that there were more DOA/DBAs among older people without ED visits or who reduced their visits in the COVID-19 group when compared to their number of ED visits at least twice in prior months. Moreover, non-LTC patients saw a greater reduction in ED visits than LTC patients across all waves by more than 10% (non-LTC: 93,896/363,879, 25.8%; LTC: 7956/67,090, 11.9%) (Figure 3). We further broke down DOA/DBA and non-DOA/DBA mortalities among older people by their LTC residence status (Table 4). In 2020, DOA/DBAs among non-LTC residents increased by 42.4% (1284/3026), which is a more than 7-fold greater increase compared to that among LTC residents (173/1310, +13.2%). In 2021, the trends were similar, apart from a smaller LTC death increase. Taken together, these results show that non-LTC patients had a much higher rate of DOA/DBA excess ED deaths during the COVID-19 pandemic (when normalized against 2016-2019).

**Table 2.** Excess emergency department 28-day deaths in the ≥65-year age group, broken down by deaths on arrival or deaths before arrival (DOA/DBA) status; wave and nonwave periods in 2020 and 2021, as defined in Multimedia Appendix 1, Table S1; and same-period 2016-2019 average deaths.

<table>
<thead>
<tr>
<th>Year</th>
<th>2016-2019 average deaths, n</th>
<th>Change, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2020</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-DOA/DBA, nonwave (179 days)</td>
<td>11,486</td>
<td>565 (4.92)</td>
</tr>
<tr>
<td>Non-DOA/DBA, wave (186 days)</td>
<td>12,717</td>
<td>561 (4.41)</td>
</tr>
<tr>
<td>DOA/DBA, nonwave</td>
<td>1973</td>
<td>571 (28.9)</td>
</tr>
<tr>
<td>DOA/DBA, wave</td>
<td>2316</td>
<td>933 (40.3)</td>
</tr>
<tr>
<td><strong>2021</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-DOA/DBA nonwave</td>
<td>20,278</td>
<td>1815 (8.95)</td>
</tr>
<tr>
<td>Non-DOA/DBA wave</td>
<td>3852</td>
<td>345 (8.96)</td>
</tr>
<tr>
<td>DOA/DBA nonwave</td>
<td>3508</td>
<td>895 (25.5)</td>
</tr>
<tr>
<td>DOA/DBA wave</td>
<td>765</td>
<td>301 (39.3)</td>
</tr>
</tbody>
</table>

**Table 3.** Same-period comparison of the incidence rate of deaths on arrival/deaths before arrival (DOA/DBAs) among people aged ≥65 years during COVID and before COVID. All DOA/DBAs were among people with at least 1 emergency department visit during the control period. Increases and reductions in the incidence rate describe the number of DOA/DBAs among people aged ≥65 years who had increased or reduced individual incidence rates in the treatment period compared to the control period. “No visits in treatment period” describes the number of DOA/DBAs among people aged ≥65 years without emergency department visits during the treatment period. $P$ values were derived from a 2-sample $z$ test.

<table>
<thead>
<tr>
<th></th>
<th>DOA/DBAs among people aged ≥65 years during COVID-19$^a$</th>
<th>DOA/DBAs among people aged ≥65 years before COVID-19$^b$</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase in incidence rate (n=416)</td>
<td>194</td>
<td>222</td>
<td>.053</td>
</tr>
<tr>
<td>No visits in treatment period (n=628)</td>
<td>427</td>
<td>201</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Reduction in incidence rate (n=542)</td>
<td>307</td>
<td>235</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

$^a$Between July 1, 2020, and July 1, 2021; control period: January 1, 2019, to December 31, 2019; treatment period: January 25, 2020, to June 30, 2020.

$^b$Between July 1, 2018, and July 1, 2019; control period: January 1, 2019, to December 31, 2019; treatment period: January 1, 2018, to June 30, 2018.
Figure 3. Seasonally adjusted change (vs 2016-2019 average) in public hospital emergency department visits across waves, with the after-wave period in dashed lines.

Table 4. Seasonally adjusted change in 28-day emergency department mortality among people aged ≥65 years by long-term care (LTC) and death on arrival/death before arrival (DOA/DBA) status.

<table>
<thead>
<tr>
<th>Year</th>
<th>Seasonally adjusted change, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LTC, non-DOA/DBA (n=8760)</td>
</tr>
<tr>
<td></td>
<td>LTC, DOA/DBA (n=1310)</td>
</tr>
<tr>
<td></td>
<td>Non-LTC, non-DOA/DBA (n=15,444)</td>
</tr>
<tr>
<td></td>
<td>Non-LTC, DOA/DBA (n=3026)</td>
</tr>
<tr>
<td>2021</td>
<td>LTC, non-DOA/DBA (n=8760)</td>
</tr>
<tr>
<td></td>
<td>LTC, DOA/DBA (n=1310)</td>
</tr>
<tr>
<td></td>
<td>Non-LTC, non-DOA/DBA (n=15,444)</td>
</tr>
<tr>
<td></td>
<td>Non-LTC, DOA/DBA (n=3026)</td>
</tr>
</tbody>
</table>
DOA/DBAs and Comorbidities Among Older People
Most of the comorbidity subsets demonstrated an annualized reduction in visits in 2020, as shown in Figure 4. Patients diagnosed with renal diseases and severe liver diseases saw a notable increase in deaths.

Figure 4. Seasonally adjusted change in episode admission among people aged ≥65 years by comorbidity and 28-day all-cause mortality, as well as absolute yearly change in 28-day mortality, vs the 2016-2019 average in 2020 wave and nonwave periods. The blue line is the 1:1 visit and death change line, on which high mortality-rate situations like cardiac arrest and sudden death normally fall. Deaths on arrival/deaths before arrival were excluded. Comorbidities were defined by the Charlson comorbidities. CHF: congestive heart failure; DM: diabetes mellitus; MI: myocardial infarction; PUD: peptic ulcer disease; PVD: peripheral vascular disease.

Discussion

Principal Findings
This study used 2SLS modeling to quantify hospital avoidance behavior and resulting excess mortality in the COVID-19 pandemic across sex, age groups, and comorbidities. Through causal inference, we discovered a higher prevalence of hospital avoidance behavior in female patients, children, and adolescents during the study period. Although surges of COVID-19 patients may overwhelm health systems and increase excess 28-day deaths [35], the waiting time in EDs in 2020 did not vastly differ from the 2016-2019 average in Hong Kong (Multimedia Appendix 1, Figure S3) [36]. Thus, the capacity of ED service was not severely impacted in 2016-2021 and we can exclude the effect of ED capacity on excess death.

The association between COVID-19 waves and hospital avoidance was observed among all age and sex groups. However, the extent of reduction varied in each age group. Lange et al [37] reported a similarly greater age-related reduction in ED use under the National Syndromic Surveillance Program for patients with selected comorbidities. On the other hand, Hung et al [38] did not discover any statistically significant differences in hospital avoidance among different age groups in their survey.

Stratified by sex, we found that women avoided hospitals more than men. This concurs with the literature on hospital avoidance behavior during the pandemic [36-40].

The largest relative reduction in ED visits in this study was found in children and adolescents (Figure 2A, Multimedia Appendix 1, Table S2), which is consistent with previous reports. Parents or caretakers may weight infection risk more heavily than the risk of delayed illness management. They might resort to visiting clinics rather than the ED or even adopt home care, given that children and adolescents are the healthiest of all age groups.

At a population level, there was a statistically significant excess death rate among older people associated with reduced ED visits. The results in Table 3 show that at the individual level, there was a significant difference in hospital avoidance behavior among older people during the pandemic compared to the prepandemic period. These observations imply that in Hong Kong, excess deaths among older people confirmed by the ED were more often due to the individual’s decision to avoid hospitals and stay home despite being severely ill rather than because of worsened hospital treatment received during the pandemic. Further breakdown of the results (Multimedia Appendix 1, Tables S6 and S7) by LTC status shows that there was a significant decrease in non-LTC ED visits during the pandemic, and there were nearly half the number of non-LTC ED DOA/DBAs among older people (n=280) compared to their
Conclusion

We demonstrated a statistical method to estimate hospital avoidance behavior during a pandemic and quantified the consequent excess 28-day mortality with a focus on older people, who have a high frequency of ED visits and deaths. This study serves as an informed alert and possible investigational guideline for health care professionals on hospital avoidance behavior and its toll.

Acknowledgments

This work was supported in part by AIR@InnoHK, administered by the Innovation and Technology Commission. ChatGPT was used to aid journal article reference searching.
Data Availability
A deidentified subject-level data set will be made available on a case-by-case basis on reasonable request to the corresponding author for research purposes 6 months after publication.

Authors' Contributions
All authors reviewed the literature, contributed to data interpretation, and read and approved the final article. CKW, AKW, and TR conceptualized the study. CKW, TFY, and YHW designed the study. AKW, TL and WLS collected the data. TFY and YHW designed the algorithms. TFY, YHW, CKC, CKW, and AKW analyzed data. TFY, YHW, CKW, KHOY, and AKW drafted the article. All authors critically revised the article. CKC and AKW confirm they had access to the data from the Clinical Data Analysis and Reporting System. TFY and YHW confirm they had access to the COVID-19 data from an online repository. TFY, YHW, and CKC accessed and verified the data sets during the study.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary Tables S1-S9 and Figures S1-S4.

References


Abbreviations

2SLS: 2-stage least squares
CDARS: Clinical Data Analysis and Reporting System
CHF: congestive heart failure
DM: diabetes mellitus
DOA/DBA: death-on-arrival/death-before-arrival
ED: emergency department
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
IV: instrumental variable
LATE: local average treatment effect
LTC: long-term care
MI: myocardial infarction
PCR: polymerase chain reaction
PUD: peptic ulcer disease
PVD: peripheral vascular disease

©Abraham Ka-chung Wai, Tsz Fung Yip, Yui Hang Wong, Chun Kit Chu, Teddy Lee, Ken Hung On Yu, Wang Leong So, Janet Y H Wong, Carlos King-ho Wong, Joshua W Ho, Timothy Rainer. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 13.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete

https://publichealth.jmir.org/2024/1/e41792

 JMIR Public Health Surveill 2024 | vol. 10 | e41792 | p.1459 (page number not for citation purposes)
bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Exploration of the Healthy Donor Effect Among 0.6 Million Blood Donors in China: Longitudinal Study

Shu Su\textsuperscript{1,2*}, PhD; Yang Sun\textsuperscript{1,3*}, PhD; Xiaoyun Gu\textsuperscript{4}, BSc; Wenjie Wu\textsuperscript{4}, BSc; Xiaodong Su\textsuperscript{4}, MSc; Ting Ma\textsuperscript{1}, PhD; Aowei Song\textsuperscript{1}, MSc; Xinxin Xie\textsuperscript{1}, MSc; Liqin Wang\textsuperscript{4}, MSc; Qianke Cheng\textsuperscript{4}, BSc; Lingxia Guo\textsuperscript{6}, BSc; Lei Zhang\textsuperscript{1,7,8,9*}, PhD; Jiangcun Yang\textsuperscript{1*}, PhD

\textsuperscript{1}Department of Transfusion Medicine, Shaanxi Provincial People’s Hospital, Xi’an, China
\textsuperscript{2}Department of Epidemiology and Biostatistics, The Second Affiliated Hospital of Chongqing Medical University, Chong Qing, China
\textsuperscript{3}Data Center, Shaanxi Provincial People’s Hospital, Xi’an, China
\textsuperscript{4}Department of Information Technology, Shaanxi Health Information Center, Xi’an, China
\textsuperscript{5}Department of Medical Record Management, Shaanxi Provincial People’s Hospital, Xi’an, China
\textsuperscript{6}Planning Development and Information Office, Health Commission of Shaanxi Province, Xi’an, China
\textsuperscript{7}China-Australia Joint Research Center for Infectious Diseases, School of Public Health, Xi’an Jiaotong University Health Science Center, Xi’an, China
\textsuperscript{8}Artificial Intelligence and Modelling in Epidemiology Program, Melbourne Sexual Health Centre, Alfred Health, Melbourne, Australia
\textsuperscript{9}Central Clinical School, Faculty of Medicine, Monash University, Melbourne, Australia
* these authors contributed equally

Corresponding Author:
Jiangcun Yang, PhD
Department of Transfusion Medicine
Shaanxi Provincial People’s Hospital
256 Youyi West Road
Xi’an, 710068
China
Phone: 86 02985251331 ext 3073
Fax: 86 029 85236987
Email: yjc65@sina.com

Abstract

Background: The World Health Organization emphasizes the importance of completely voluntary blood donation to maintain safe and sustainable blood supplies. However, the benefits of blood donation for donors, such as reducing the risk of disease, remain a topic of debate due to the existence of the healthy donor effect (HDE). This effect arises because of inherent health differences between blood donors and the general population, and it is also considered a methodological issue.

Objective: This study aims to generate a more detailed health profile of blood donors from a donor cohort study to mitigate and quantify the HDE and properly interpret the association between blood donation and disease outcomes among blood donors.

Methods: A retrospective cohort study was conducted between January 2012 and December 2018 among donors before their first donation. One-to-one propensity score matching was conducted through a random selection of individuals without any history of blood donation, as reported from their electronic health records. We conducted a Poisson regression between blood donors and non–blood donors before the first donation to estimate the adjusted incidence rate ratio (AIRR) of selected blood donation–related diseases, as defined by 13 categories of International Classification of Diseases, Tenth Revision (ICD-10) codes.

Results: Of the 0.6 million blood donors, 15,115 had an inpatient record before their first donation, whereas 17,356 non–blood donors had an inpatient record. For the comparison between blood donors and the matched non–blood donors, the HDE (the disease incidence rate ratio between non–blood donors and blood donors) was an AIRR of 1.152 (95% CI 1.127-1.178; \( P<.001 \)). Among disease categories not recommended for blood donation in China, the strongest HDE was observed in the ICD-10 D50-D89 codes, which pertain to diseases of the blood and blood-forming organs as well as certain disorders involving the immune mechanism (AIRR 3.225, 95% CI 2.402-4.330; \( P<.001 \)). After age stratification, we found that people who had their first blood donation between 46-55 years old had the strongest HDE (AIRR 1.816, 95% CI 1.707-1.932; \( P<.001 \)). Both male and female donors had significant HDE (AIRR 1.082, 95% CI 1.05-1.116; \( P=.003 \); and AIRR 1.236, 95% CI 1.196-1.277; \( P<.001 \), respectively) compared with matched non–blood donors.
**Conclusions:** Our research findings suggest that the HDE is present among blood donors, particularly among female donors and those who first donated blood between the ages of 46 and 55 years.

**Trial Registration:** Chinese Clinical Trial Registry ChiCTR2200055983; https://www.chictr.org.cn/showproj.html?proj=51760

**KEYWORDS**
healthy donor effect; blood donors; predonation; health profile; longitudinal study; propensity score matching

**Introduction**

The World Health Organization (WHO) promotes voluntary blood donation as the safest and most effective means to maintain an adequate supply of safe blood and blood products for transfusion. To ensure the safety and sustainability of blood supplies, the WHO advocates for completely voluntary blood donation without compensation. Despite the collection of approximately 118.54 million blood donations worldwide [1], the benefits of blood donation for donors are still a matter of debate, as some studies suggest the existence of the healthy donor effect (HDE).

The HDE refers to the observation that blood donors exhibit lower disease morbidity and mortality compared with the general population. This phenomenon can be attributed to the fact that blood donors are typically selected from a healthier subset of the population, due to both donor selection procedures and self-selection factors [2,3]. The HDE has been examined in numerous past studies by comparing the health status of blood donors to that of non–blood donors [3-8]. Although some studies have failed to reveal any significant effects on health status resulting from blood donation [4,6,7], others found substantial differences such as decreased cancer incidence [5] and better cardiovascular status in blood donors compared with the general population [9]. The findings regarding the HDE have been largely inconclusive and sometimes contradictory, mainly because researchers were not able to deal with inevitable selection bias, that is, the fact that donors have better healthier status due to the donation policy requirements. In practice, it is nearly impossible to determine whether a beneficial health status observed in blood donors reflects the inherent healthier status of donors or the actual health gains due to blood donation [8]. Thus, it is important to mitigate and quantify the HDE to properly interpret the association between blood donation and disease outcomes.

Multiple methods have been used to address the HDE in various studies. The main strategy is to apply an exposure window or qualification period to potentially mitigate the HDE [10]. It mainly involves using 1 qualification period (ie, all donors must donate for at least 5 or 10 years to be eligible for the study) to adjust for the HDE [11-13]. However, this effort to reduce the HDE is conducted after blood donation. In addition, some studies reduce selection bias by analyzing only healthier groups; comparing differences within blood donors; and dealing with mixed bias, including the HDE, through ANOVA or some form of regression analysis [14-17]. However, adjusting age, sex, and demographic factors to reduce the HDE through some form of statistical strategy will affect the probability of subsequent exposure, and adjusting for confounding factors is not enough to address the HDE sufficiently [15,16]. To avoid the possible bias, another method is to compare or restrict the study within the active donor population, which still cannot exclude the existence of the residual HDE from most observational studies [14,17,18]. Currently, there is no evidence to certify the existence of the HDE prior to donation. The major weakness is that the potential HDE is not quantified and is usually estimated by researchers [10].

The first step to reduce the HDE is to accurately estimate it. The precise assessment of the HDE can be retrospectively obtained during the predonation period. Leveraging our large blood donor cohort, which is among the largest in the world, we are able to match non–blood donors and thus overcome the major limitations of previous studies on the HDE. This would allow us to obtain unprecedented insights into the health status of blood donors and non–blood donors and to quantify the HDE in a more accurate manner. Our study was conducted as a retrospective cohort study, wherein the health status of blood donors and non–blood donors were compared by one-to-one propensity score matching (PSM). Our main objective is to confirm the existence of the HDE in the population while also assessing the impacts of age and sex on the HDE. Through this comprehensive analysis, we aim to generate a more detailed health profile of blood donors to certify the HDE and better understand the underlying factors that contribute to it.

**Methods**

**Data Source**

We extracted the disease information from blood donors in Shaanxi Province, which is located in Northwest China. The Shaanxi Blood Donation Database is a computerized, combined donation and transfusion register from 3.4 million individuals who donated blood voluntarily between 1998-2018, and their relevant blood information was collected in the Shaanxi Blood Donor Database by linking with province-wide Shaanxi electronic health records (EHRs) and centralized hospital medical records. From this linkage, all the inpatient records of the donors can be searched. This database also tracks the disease information of participants without any history of blood donation. We have provided detail information for this cohort in our previous publication [19]. The disease burden of blood donors and non–blood donors was demonstrated. This natural and representative population cohort could be used to describe the health profiles of blood donors and investigate the potential HDE for selected clinically diagnosed diseases (as listed below) for donors, compared with non–blood donors, before the first donation date.
We initially included 3,389,981 donors in the cohort, and 1,907,146 had health records from January 1, 2012, to December 31, 2018. We excluded 1,063,418 participants with disease records before 2012, resulting in 641,523 qualified EHR records of eligible donors. Based on this information, we established a control group with the same criteria by using one-to-one matching to the donor group. Our study recognized the substantial impact of factors such as population aging, sex disparities, environmental issues, and rural-urban differences on individual health [20-23]. Thus, to provide a more accurate assessment of the HDE, we used a PSM analysis. This approach allowed us to match non–blood donors to blood donors with identical demographic characteristics (sex, age, and residence) from the EHRs and investigate the occurrence of a wide range of inpatient diseases defined by International Classification of Diseases, Tenth Revision (ICD-10) codes prior to the first donation date. The number of disease records was calculated during 2012-2018 and was extracted from the inpatient record, which is automatically uploaded from the hospital system. The matching was conducted by random sampling without the replacement of individuals without any history of blood donation from the EHR. We only extracted the disease information related to the blood donation criteria, resulting in the blood donor data set and non–blood donor data set having 58,235 and 57,002 inpatient records, respectively. Then, we further restricted the disease records to before the donation period, resulting in 15,115 donor records and 17,356 nondonor records, respectively (Figure 1).

Figure 1. Study flowchart of the longitudinal study.

Selection of Diseases
According to the national guidelines for blood donation requirements [24], only specific diseases that are related to blood donation were selected for study. Thus, diseases corresponding to injury, poisoning, pregnancy, childbirth and the puerperium, and others were excluded. We only collected disease information for 13 categories—certain infectious and parasitic diseases (A00-B99); neoplasms (C00-D48); diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89); endocrine, nutritional, and metabolic diseases (E00-E90); mental and behavioral disorders (F00-F99); diseases of the nervous system (G00-G99); diseases of the circulatory system (I00-I99); diseases of the respiratory system (J00-J99); diseases of the skin and subcutaneous tissue (L00-L99); diseases of the genitourinary system (N00-N99); external causes of morbidity and mortality (V01-Y98); and factors influencing health status and contact with health services (Z00-Z99)—to conduct the comparison between blood donors and non–blood donors before donation.

Inclusion and Exclusion Criteria
In this cohort, we only included (1) blood donors who gave their first donation between 2012 and 2018 and (2) blood donors who have EHR records between 2012 and 2018 in Shaanxi Province. We excluded (1) blood donors without a blood donation date; (2) blood donors with a disease record before 2012 or who left Shaanxi Province; and (3) blood donors with incomplete demographic information. For the control group of
non–blood donors, identical inclusion and exclusion criteria were implemented, with the exception of content specific to blood donation. This standardized approach ensures comparability between the blood donor and non–blood donor cohorts, establishing a robust foundation for our comprehensive investigation into the HDE.

Assessing the HDE

We then specifically collected the disease information of these blood donors between baseline and the date of their first blood donation. Similarly, we also extracted the disease information of the matching non–blood donors during the same time and compared the disease rates to quantify the HDE. For example, we collected the number of certain infectious and parasitic diseases (A00-B99) diagnosed among donors and nondonors during the study period, which was used as the numerator, whereas the total observed number of donors and nondonors (it is the same number in this study; n=641,523) was used as the denominator (Figure 2). Then, we further estimated the incidence rate ratio between these 2 groups.

According to the Chinese blood donation guidelines, we set the age criteria for entering the cohort as 18-55 years, and we stratified age into 4 groups (18-25, 26-35, 36-45, and 46-55 years). The age when the participants entered the cohort was used. The model compared all the disease rates among male and female donors and nondonors among the 4 age groups. Poisson regression models were built to identify donors’ and nondonors’ health status in the respective exposure windows. The observed period was defined as from 2012 to the first donation date, and the non–blood donors used the same cutoff point accordingly (Figure 2).

Figure 2. Study schematic of the longitudinal study.

Statistical Analysis

Poisson regression is often used for count data where the outcome variable represents the number of events that occur in a fixed interval of time or space. It is also appropriate for rare events [25]. This methodology has been widely used in HDE studies [3,26]. We used Poisson regression to compare the likelihood of becoming hospitalized between blood donors and non–blood donors prior to the first blood donation. The adjusted incidence rate ratio (AIRR) was used as an estimate of the ratio of the incident risks for an event for the non–blood donors to the risks for the blood donors during the study period; that is, the reference group is blood donors. The model was adjusted by occupation, education, and marital status. The waiting period between hospital discharge and first blood donation was calculated for each age group in each disease category. Descriptive and inferential statistical analyses were conducted as part of our methodology. Mean and SD were used to summarize age and donation counts, whereas median and IQR were used to summarize the duration in days between discharge and the first donation. For categorical variables, we described them using frequencies and percentages. In the context of inferential statistical tests, we used both multivariate regression analyses and sensitive analysis to evaluate model assumptions and validate the suitability of our chosen analytical techniques. All statistical analyses and data handling were conducted with computer software (SAS version 9.4; SAS Institute).

Ethical Considerations

Ethical approval for the study was obtained from the institutional review board of the People’s Hospital of Shaanxi Province (No: 2020-R002). The study was preregistered at China National Medical Research Register (MR-61-21-011750) and Chinese Clinical Trial Registry (ChiCTR2200055983). Since our study is a secondary analysis using existing data with primary consent, where the original consent or institutional review board approval covers secondary analysis, no additional consent was needed for this study. The data have been deidentified and can only be accessed by the researcher group for this study.

Results

Baseline Characteristics

This study included 641,523 blood donors and 641,523 non–blood donors, which were matched by age, sex, and region. There were 396,681 male donors and 244,842 female donors in the study, and a total of 3,396,992 person-years (PYs) were observed. The mean age was 30.4 (SD 9.3) years, with most of them (253,045/641,523, 39.4% for both groups) being aged 18-25 years. The majority (429,421/641,523, 66.9%) of the blood donors came from Central Shaanxi, 26.6% (170,274/641,523) came from South Shaanxi, and 6.5% (41,828/641,523) came from North Shaanxi. Age, sex, and region distributions of the non–blood donors were the same as
those of blood donors. Regarding education, 59.2% (379,592/641,523) of the blood donors and 56.9% (365,294/641,523) of non–blood donors had received an education of junior high school or below. Regarding occupation, 65.3% (418,757/641,523) of the blood donors and 61.2% (392,540/641,523) of the non–blood donors were farmers. Regarding marriage status, 71.4% (458,070/641,523) of the blood donors and 70.5% (452,156/641,523) of the non–blood donors were married. Baseline characteristics of the participants are displayed in Table 1.

### Table 1. Basic demographic characteristics of participants stratified by blood donors and non-blood donors.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Blood donors (n=641,523)</th>
<th>Non–blood donors (n=641,523)</th>
<th>(P) value(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Male</td>
<td>396,681 (61.8)</td>
<td>396,681 (61.8)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>244,842 (38.2)</td>
<td>244,842 (38.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td></td>
<td></td>
<td>&gt;.99</td>
</tr>
<tr>
<td>18-25, n (%)</td>
<td>253,045 (39.4)</td>
<td>253,045 (39.4)</td>
<td></td>
</tr>
<tr>
<td>26-35, n (%)</td>
<td>180,185 (28.1)</td>
<td>180,185 (28.1)</td>
<td></td>
</tr>
<tr>
<td>36-45, n (%)</td>
<td>164,645 (25.7)</td>
<td>164,645 (25.7)</td>
<td></td>
</tr>
<tr>
<td>46-55, n (%)</td>
<td>43,648 (6.8)</td>
<td>43,648 (6.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of donation, mean (SD)</strong></td>
<td>1.45 (1.07)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Junior high and below</td>
<td>379,592 (59.2)</td>
<td>365,294 (56.9)</td>
<td></td>
</tr>
<tr>
<td>Senior high</td>
<td>119,669 (18.7)</td>
<td>114,611 (17.9)</td>
<td></td>
</tr>
<tr>
<td>University and above</td>
<td>64,081 (10)</td>
<td>81,930 (12.8)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>78,181 (12.2)</td>
<td>79,688 (12.4)</td>
<td></td>
</tr>
<tr>
<td><strong>Marriage status, n (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Single</td>
<td>132,613 (20.7)</td>
<td>139,177 (21.7)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>458,070 (71.4)</td>
<td>452,156 (70.5)</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td>8007 (1.2)</td>
<td>7729 (1.2)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>42,833 (6.7)</td>
<td>42,461 (6.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Occupation, n (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Worker</td>
<td>74,266 (11.6)</td>
<td>73,061 (11.4)</td>
<td></td>
</tr>
<tr>
<td>Farmer</td>
<td>418,757 (65.3)</td>
<td>392,540 (61.2)</td>
<td></td>
</tr>
<tr>
<td>Self-employed</td>
<td>121,609 (19)</td>
<td>140,808 (21.9)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>26,891 (4.2)</td>
<td>35,114 (5.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Region, n (%)</strong></td>
<td></td>
<td></td>
<td>&gt;.99</td>
</tr>
<tr>
<td>North Shaanxi</td>
<td>41,828 (6.5)</td>
<td>41,828 (6.5)</td>
<td></td>
</tr>
<tr>
<td>Central Shaanxi</td>
<td>429,421 (66.9)</td>
<td>429,421 (66.9)</td>
<td></td>
</tr>
<tr>
<td>South Shaanxi</td>
<td>170,274 (26.5)</td>
<td>170,274 (26.5)</td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\)Chi-square test.

### Disease Burden Among Donors

Among the blood donors, 15,115 participants had inpatient records during the observed period, with a median follow-up duration of 1061 (IQR 510-1803) days. The disease incidence consistently showed an upward trend with age: it increased from 26.69 per 10,000 PYs for the 18-25 years age group to 86.98 per 10,000 PYs for the 46-55 years age group. Diseases of the digestive system (K00-K93) were most reported diseases among blood donors with 3240 records, which was followed by diseases of the circulatory system (I00-I99; n=2856) and diseases of the genitourinary system (N00-N99; n=2519). Diseases of the circulatory system (I00-I99) were the most reported diseases with an incidence rate of 30.30 per 10,000 PYs in the 46-55 years age group.

### Disease Burden Among Non–Blood Donors

Among non–blood donors, 17,356 participants had inpatient records during the observed period. The disease incidence increased with age: it ranged from 34.63 per 10,000 PYs for the 18-25 years age group to 149.51 per 10,000 PYs for the 46-55 years age group. There were 3462 patients who had...
diseases of the digestive system (K00-K93), which was the most reported disease, followed by diseases of the circulatory system (I00-I99; n=3042) and diseases of the genitourinary system (N00-N99; n=2533). The highest incidence of 44.36 per 10,000 PYs was found among diseases of the circulatory system (I00-I99) in the 46-55 years age group.

**HDE on the Overall Population**

Overall, the risk of reporting ICD-10 diseases in nondonors was generally higher than that from donors before the first blood donation (AIRR 1.152, 95% CI 1.127-1.178; \(P<.001\)), which was defined as the existence of HDE. The top 3 diseases with the strongest HDE were observed in diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89; AIRR 3.225, 95% CI 2.402-4.330; \(P<.001\)); certain infectious and parasitic diseases (A00-B99; AIRR 1.995, 95% CI 1.812-2.196; \(P<.001\)); and factors influencing health status and contact with health services (Z00-Z99; AIRR 1.980, 95% CI 1.796-2.184; \(P<.001\)). When considering the specific diseases, non–blood donors had a lower risk in relation to diseases of the nervous system (G00-G99; AIRR 0.939, 95% CI 0.841-1.049; \(P=.13\)); diseases of the respiratory system (J00-J99; AIRR 0.933, 95% CI 0.882-1.135; \(P=.29\)); diseases of the skin and subcutaneous tissue (L00-L99; AIRR 1.101, 95% CI 0.946-1.281; \(P=.21\)); and diseases of the genitourinary system (N00-N99; AIRR 1.006, 95% CI 0.952-1.064; \(P=.32\); Figure 3 and Table S1 in Multimedia Appendix 1).

**Figure 3.** Adjusted incidence rate ratio (AIRR) in comparing blood donation–related diseases between donors and nondonors: the healthy donor effect. ICD-10: International Classification of Diseases, Tenth Revision.

**HDE by Sex Differences**

Our study found that male and female individuals both have significant HDE, but it was stronger in female individuals (AIRR 1.236, 95% CI 1.196-1.277; \(P<.001\)) than male individuals (AIRR 1.082, 95% CI 1.050-1.116; \(P=.003\)). The most substantial difference was found in the risk of having factors influencing health status and contact with health services (Z00-Z99), as female non–blood donors in all age groups had a higher risk than blood donors, but only male non–blood donors aged 46-55 years had a higher risk (AIRR 2.273, 95% CI 1.205-4.289; \(P<.001\)). For diseases, the most significant difference was observed in the risk of diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89); the HDE of female donors was stronger than that of male donors (female: AIRR 4.294, 95% CI 2.887-6.388; \(P<.001\) vs male: AIRR 2.042, 95% CI 1.299-3.210; \(P<.001\)). After age stratification, male donors who had their first donation between 46-55 years old had the strongest HDE among all age groups (AIRR 2.193, 95% CI 2.011-2.392; \(P<.001\)), whereas female donors who first donated between 18-25 years old had the strongest HDE (AIRR 1.571, 95% CI 1.465-1.685; \(P<.001\); Figure 4 and Tables S2-3 in Multimedia Appendix 1).
HDE by Age Groups

In terms of age groups, among non–blood donors, the 46-55 years age group showed the highest risk (AIRR 1.816, 95% CI 1.707-1.932; P < .001), followed by the 18-25 years age group (AIRR 1.302, 95% CI 1.245-1.362; P < .001). Conversely, the 26-35 and 36-45 years age groups did not show significant differences between donors and nondonors (AIRR 0.979, 95% CI 0.936-1.023; P = .21 and AIRR 0.988, 95% CI 0.951-1.025; P = .32, respectively). Notably, donors who initiated donation after the age of 46 years had a consistently better health status compared with non–blood donors. It was shown that non–blood donors had a 5-times higher risk than donors in diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89; AIRR 5.181, 95% CI 1.968-13.636; P < .001) and a 3-times higher risk in endocrine, nutritional, and metabolic diseases (E00-E90; AIRR 3.384, 95% CI 2.612-4.383; P < .001) and certain infectious and parasitic diseases (A00-B99; AIRR 3.181, 95% CI 2.183-4.635; P < .001). Non–blood donors in all age classes were associated with higher risks for certain infectious and parasitic diseases (A00-B99); diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89); endocrine, nutritional, and metabolic diseases (E00-E90); mental and behavioral disorders (F00-F99); and factors influencing health status and contact with health services (Z00-Z99) compared with donors (Figure 3 and Table S1 in Multimedia Appendix 1).

Sensitive Analysis

Upon conducting a comparative analysis, we made necessary adjustments to the overall findings by excluding the disease categories of external causes of morbidity and mortality (V01-Y98) and factors influencing health status and contact with health services (Z00-Z99). Subsequently, our results demonstrated a consistent increase in the risk of HDE among both blood donors and non–blood donors, with an AIRR of 1.117 (95% CI 1.092-1.143).

Gaps Between Hospital Discharge and First Blood Donation

The overall median waiting period between inpatients being discharged from the hospital and the first blood donation was 534 (IQR 241-960) days. This duration showed a declining trend with respect to age groups. The longest waiting period was found in the 18-25 years age group (median 592, IQR 267-1027 days). The shortest waiting period was found in the 46-55 years age group (median 439, IQR 204-813 days). In the 26-35 years age group, the gap was a median of 531 (IQR 240-970) days, whereas in the 36-45 years age group, the gap was a median of 521 (IQR 239-942) days. The longest gap was observed among those who had the disease category of neoplasms (C00-D48; median 673, IQR 358-1137 days).
followed by factors influencing health status and contact with health services (Z00-Z99; median 660, IQR 373-1109 days), and certain infectious and parasitic diseases (A00-B99; median 654, IQR 298-1154 days). The shortest gap was found in those with the disease category of endocrine, nutritional, and metabolic diseases (E00-E90; median 456, IQR 152-877 days).

Regarding the overall trend, the waiting period showed a descending trend, with the gap changing from a median of 592 (IQR 267-1027) days in 18-25 years age group to a median of 439 (IQR 204-813) days in the 46-55 years age group (Table 2).

Table 2. Gaps between hospital discharge and the first blood donation among blood donors.

<table>
<thead>
<tr>
<th>Disease categories</th>
<th>Days between discharge and the first donation, median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18-25 years age group</td>
</tr>
<tr>
<td>Certain infectious and parasitic diseases (A00-B99)</td>
<td>770 (335-1192)</td>
</tr>
<tr>
<td>Neoplasms (C00-D48)</td>
<td>595 (286-978)</td>
</tr>
<tr>
<td>Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89)</td>
<td>496 (262-916)</td>
</tr>
<tr>
<td>Endocrine, nutritional, and metabolic diseases (E00-E90)</td>
<td>451 (216-828)</td>
</tr>
<tr>
<td>Mental and behavioral disorders (F00-F99)</td>
<td>533 (301-940)</td>
</tr>
<tr>
<td>Diseases of the nervous system (G00-G99)</td>
<td>472 (164-953)</td>
</tr>
<tr>
<td>Diseases of the circulatory system (I00-I99)</td>
<td>608 (256-1054)</td>
</tr>
<tr>
<td>Diseases of the respiratory system (J00-J99)</td>
<td>649 (291-1124)</td>
</tr>
<tr>
<td>Diseases of the digestive system (K00-K93)</td>
<td>604 (263-1023)</td>
</tr>
<tr>
<td>Diseases of the skin and subcutaneous tissue (L00-L99)</td>
<td>432 (178-999)</td>
</tr>
<tr>
<td>Diseases of the genitourinary system (N00-N99)</td>
<td>611 (259-1052)</td>
</tr>
<tr>
<td>External causes of morbidity and mortality (V01-Y98)</td>
<td>N/A</td>
</tr>
<tr>
<td>Factors influencing morbidity and mortality (V01-Y98)</td>
<td>661 (378-1054)</td>
</tr>
<tr>
<td>Overall</td>
<td>592 (267-1027)</td>
</tr>
</tbody>
</table>

*N/A: not applicable.

**Discussion**

The findings of our study provide robust evidence of the existence of the HDE among blood donors in China. In line with the blood donation guidelines of the country, the overall health status of blood donors was found to be superior to that of non–blood donors, as evidenced by a reduced risk of developing selected diseases. In particular, individuals who had their first blood donation at or after the age of 46 years exhibited a significantly lower risk of developing diseases across all categories compared with non–blood donors. The results of our study highlight a marked disparity in the HDE by sex, with concerns that donating blood may negatively impact the health of participants with these conditions. Although previous research has aimed to disprove the HDE, our study is the first to both confirm its presence and quantify its magnitude.

Although blood donors generally exhibit a better health status than non–blood donors during the predonation period, our findings indicate that non–blood donors are superior in certain disease age groups. The disparity in health status between blood donors and non–blood donors could be attributed to several factors. First, our data revealed that blood donors with a higher disease rate than non–blood donors were all younger than the age of 45 years. This observation suggests that older donors may be more conscientious of their health and well-being [27,28]. Second, certain diseases, although they may require hospitalization, do not necessarily prohibit patients from donating blood. For instance, our data revealed that the majority of blood donors with inpatient records for respiratory diseases had contracted influenza or pneumonia (J09-J18). As the risk of transmitting these diseases directly through blood or blood products is extremely low, individuals who have recovered from these conditions and are free from clinical symptoms may still be eligible to donate blood [29]. Third, previous studies have shown that appropriate blood donation can help reduce the risk of certain diseases. Blood donation has been found to increase the concentrations of high-density lipoprotein and apolipoprotein A while also lowering the potential for oxidation of low-density lipoprotein (LDL) particles. Lower levels of LDL peroxidation may have a protective effect against certain diseases [30]. Additionally, blood donation has been associated with a reduced risk of circulatory diseases and tumors [31,32], possibly due to the reduction of iron concentrations in the body through blood loss therapy. However, some studies suggest that iron loss through blood donation may also be associated with certain diseases [6,33,34].
This study demonstrated that female individuals exhibited a stronger HDE than male individuals. This difference can be attributed to social and physiological factors, as well as the different incidence of common diseases in male and female individuals [35]. Furthermore, female individuals demonstrated a strong HDE in relation to oncology and digestive diseases, which may be related to life circumstances, work stress, and lifestyle factors such as smoking and alcohol consumption [36,37]. Interestingly, the results of age stratification indicated that the strength of the HDE differs between age categories. Specifically, female individuals between the ages of 18-45 years exhibited a stronger HDE than their male counterparts, whereas male individuals aged 46-55 years exhibited a stronger HDE compared to female individuals in the same age group. These results may be explained by the overall physiological differences that exist between male and female individuals in their reproductive stages, such as the menstrual cycle and pregnancy, which could lead to female donors paying more attention to their health and neonatal health and following the donation guidelines more closely, resulting in an increased HDE compared to male individuals [38]. Importantly, it is worth noting that the strength of HDE was no longer accentuated in female individuals after the age of 46 years, as they have entered the nonreproductive phase and the aforementioned reproductive factors have disappeared, resulting in a stronger HDE in male individuals compared to female individuals. These results highlight the importance of considering sex effects in the expression of the HDE.

We also found that the average time interval between participants donating blood and being discharged from hospital was almost 2 years. Given the shortage of blood storage, it is recommended that restrictions on donation requirements be discussed, similar to other countries [39], such as allowing inpatients with cured diseases to donate blood after a certain period. This is especially important during the COVID-19 outbreak, where blood supply shortages have been reported [40], and China has updated guidelines stating that patients with COVID-19 can donate blood after 7 days of seroconversion [41]. We are the first to demonstrate that older people have a shorter interval between discharge and blood donation than younger people, which may be due to 2 reasons. First, as the age limit for blood donation is ≤55 years, donors must donate blood before this age limit. Many countries have a higher age limit than China, and considering that life expectancy has increased in China, the age limit for blood donation may also need to be increased. Furthermore, China encourages donors to donate blood by allowing the donor or a family member to receive an equal volume of the donated blood with priority.

Older adults may have been more reluctant to donate blood previously but have become more accepting after this promotion policy was enacted. Consistent with this policy, donors aged 46-55 years have increased from 11.7% in 2012 to 19.6% in 2018 in Shaanxi Province, whereas donors aged 18-25 years have decreased from 36.7% in 2012 to 30.3% in 2018 [42]. However, it is crucial to ensure the safety of blood transfusions, particularly for donors with hospitalization records. To address this, we recommend that blood centers link donors’ identification with their EHRs, allowing health care staff to access their previous disease history and determine their eligibility for blood donation. In China, the National Health Commission has been promoting the establishment of EHRs with various policies and financial support since 2010 [43]. The health system can be effectively used to inform donors about their medical history, and only healthy individuals with eligible medical histories could be qualified to donate blood or blood components.

Some limitations should be noted in this study. First, as a retrospective study, the records were extracted from the EHR, and the number of people who left the province was not recorded, which may have influenced the hospitalization rate. However, based on the current data, the proportion of immigrants from Shaanxi Province is approximately 10%, and this is unlikely to significantly affect the results. Second, environmental factors and family medical history, both of which may have an impact on respiratory diseases and tumors, were not taken into account in the study. Third, blood routine and liver function data were not readily available for non–blood donors, so we cannot include these factors in the PSM process at baseline. Given that blood routine and liver function tests involve multiple indicators, participants with abnormal data in these tests may potentially have asymptomatic diseases that have not yet been diagnosed. This limitation introduces a potential source of confounding bias in our investigation of related diseases. Finally, the duration of hospitalization may have been influenced by the economic and health care quality factors of the admitting hospitals, which could not be considered in the analysis. Despite these limitations, this study is the first and largest population-based study to provide comprehensive evidence of the existence of the HDE, especially for the female donors and donors who donate over the age of 46 years. We highlighted the substantial differences in blood-related diseases between blood donors and non–blood donors prior to donation. Consequently, this has the potential to introduce bias in comparing donor effects between the 2 groups. Future studies should aim to develop more accurate methods to calculate donor effects and eliminate this bias.
JCY is supported by the 2021 Science and Technology Talent Support Program of Shaanxi Provincial People’s Hospital (grant 2021LJ-14). SS is supported by the National Natural Science Foundation of China (grant 82304246); the Natural Science Foundation of Chongqing (grant CSTB2023NSCQ-MSX0198); the joint project of the Chongqing Health Commission and Science and Technology Bureau (grant 2024QNXM056); and the Second Affiliated Hospital of Chongqing Medical University Research (grant 2023JITXG26). LZ is supported by the National Natural Science Foundation of China (grant 81950410639); Outstanding Young Scholars Funding (grant 3111500001); Xi’an Jiaotong University Basic Research and Profession Grant (grants xtr022019003 and xzy032020032); Xi’an Jiaotong University Young Talent Support Grant (grant YX6f004); and part of National Key R&D Program of China (grant 2022YFC2505100).

Data Availability
The data sets used in this study are not publicly available because they contain information that could compromise research participant privacy, but they are available from the corresponding author upon reasonable request.

Authors' Contributions
SS, YS, and JCY had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. SS, LZ, and JCY contributed to the concept and design. SS, YS, XG, WW, XS, TM, AS, XX, LW, QC, LG, and JCY contributed to the acquisition, analysis, and interpretation of data. SS, YS, JCY, and LZ drafted the manuscript. YS, AS, WW, and XG critically revised the manuscript for important intellectual content. SS and YS conducted the statistical analysis. LZ and JCY contributed equally to supervision and are cocorresponding authors.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Healthy donor effect in the general population and by gender differences.

References


### Abbreviations

AIRR: adjusted incidence rate ratio  
EHR: electronic health record  
HDE: healthy donor effect  
ICD-10: International Classification of Diseases, Tenth Revision  
LDL: low-density lipoprotein  
PSM: propensity score matching  
PY: person-year  
WHO: World Health Organization
The Use of ICD-9-CM Coding to Identify COVID-19 Diagnoses and Determine Risk Factors for 30-Day Death Rate in Hospitalized Patients in Italy: Retrospective Study

Barbara Giordani¹, MSc; Alessandra Burgio², MSc; Francesco Grippo², MSc; Alessandra Barone¹, MSc; Erica Eugeni¹, MA, PhD; Giovanni Baglio¹, MD, MSc

¹Research, National Outcomes Evaluation Programme (PNE) and International Relations Unit, Italian National Agency for Regional Healthcare Services, Rome, Italy
²Italian National Institute of Statistics, Rome, Italy

Corresponding Author:
Barbara Giordani, MSc
Research, National Outcomes Evaluation Programme (PNE) and International Relations Unit
Italian National Agency for Regional Healthcare Services
Via Piemonte 60
Rome, 00187
Italy
Phone: 39 06 42749713
Email: giordani@agenas.it

Abstract

Background: In Italy, it has been difficult to accurately quantify hospital admissions of patients with a COVID-19 diagnosis using the Hospital Information System (HIS), mainly due to the heterogeneity of codes used in the hospital discharge records during different waves of the COVID-19 pandemic.

Objective: The objective of this study was to define a specific combination of codes to identify the COVID-19 hospitalizations within the HIS and to investigate the risk factors associated with mortality due to COVID-19 among patients admitted to Italian hospitals in 2020.

Methods: A retrospective study was conducted using the hospital discharge records, provided by more than 1300 public and private Italian hospitals. Inpatient hospitalizations were detected by implementing an algorithm based on specific International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code combinations. Hospitalizations were analyzed by different clinical presentations associated with COVID-19 diagnoses. In addition, 2 multivariable Cox regression models were performed among patients hospitalized “due to COVID-19” from January 1 to December 31, 2020, to investigate potential risk factors associated with 30-day death and the temporal changes over the course of the pandemic; in particular, the 30-day death rates during the first and the second waves were analyzed across 3 main geographical areas (North, Center, and South and Islands) and by discharge wards (ordinary and intensive care).

Results: We identified a total of 325,810 hospitalizations with COVID-19–related diagnosis codes. Among these, 73.4% (n=239,114) were classified as “due to COVID-19,” 14.5% (n=47,416) as “SARS-CoV-2 positive, but not due to COVID-19,” and 12.1% (n=39,280) as “suspected COVID-19” hospitalizations. The cohort of patients hospitalized “due to COVID-19” included 205,048 patients, with a median age of 72 years and a higher prevalence of male patients (n=124,181, 60.6%). The overall 30-day death rate among hospitalized patients due to COVID-19 was 9.9 per 1000 person-days. Mortality was lower for women (hazard ratio [HR]=0.83; P<.001) and for patients coming from high migration pressure countries, especially Northern Africans (HR=0.65; P<.001) and Central and Eastern Europeans (HR=0.66; P<.001), compared to patients coming from Italy and high-income countries. In the southern regions and the Islands, mortality was higher compared to the northern regions (HR=1.17; P<.001), especially during the second wave of COVID-19 among patients with a transfer to intensive care units (HR=2.52; P<.001).

Conclusions: To our knowledge, the algorithm is the first attempt to define, at a national level, selection criteria for identifying COVID-19 hospitalizations within the HIS. The implemented algorithm will be used to monitor the pandemic over time, and the patients selected in 2020 will be followed up in the next years to assess the long-term effects of COVID-19.
Annex 1 of the Ministerial Decree and the guidelines issued on specific only with the Ministerial Decree of October 28, 2020 [14], by adapting unspecific codes already existing in the homogeneity in the criteria and methods of HDR coding (MoH) published the first guidelines to ensure the needed therefore, on March 20, 2020, the Italian Ministry of Health currently used in Italy for coding in the HDRs.

Indeed, with the beginning of the COVID-19 pandemic, the importance of human genetics in the onset and course of severity of the disease [6-8], evidence is arising about the association with a greater risk of death from COVID-19 [5]. Although environmental, clinical, and social characteristics have a role as risk factors for SARS-CoV-2 infection and the severity of the disease [6-8], evidence is arising about the importance of human genetics in the onset and course of COVID-19 [9,10].

In Italy, despite early data provided by the COVID-19 Integrated Surveillance System [11], it has been difficult to accurately quantify hospital admissions of patients with a COVID-19 diagnosis using the Hospital Information System (HIS), mainly due to the heterogeneity of codes used in the hospital discharge records (HDRs) during different waves of the COVID-19 pandemic.

Indeed, with the beginning of the COVID-19 pandemic, the World Health Organization provided, in response to member state requests, codes and instructions for COVID-19 coding in the International Classification of Diseases, 10th Revision (ICD-10) and International Classification of Diseases, 11th Revision (ICD-11) [12], but not in the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), which is the international classification of diseases currently used in Italy for coding in the HDRs.

Therefore, on March 20, 2020, the Italian Ministry of Health (MoH) published the first guidelines to ensure the needed homogeneity in the criteria and methods of HDR coding throughout the country [13]. Such guidelines provide instructions to identify COVID-19 and related clinical conditions by adapting unspecific codes already existing in the ICD-9-CM. Only with the Ministerial Decree of October 28, 2020 [14], specific ICD-9-CM codes were introduced for COVID-19.

February 19, 2021, provided new instructions for both the principal and secondary diagnoses, aiming to overcome differences in coding among different Italian regions. However, the harmonization of COVID-19 information on HDRs has been gradual in different Italian regions, and sometimes, it was done retrospectively by updating and changing codes in HDRs.

The algorithm presented in this paper, based on the COVID-19 codes included in different MoH guidelines, the ICD-9-CM SARS codes, and the codes that reported a dramatic increase in 2020, has made it possible to quantify patients with COVID-19–related diagnoses admitted to Italian hospitals. Moreover, the algorithm made it possible to classify hospitalizations as “due to COVID-19,” “SARS-CoV-2 positive, but not due to COVID-19,” and “suspected COVID-19,” and investigate the risk factors associated with mortality “due to COVID-19” among patients admitted to Italian hospitals in 2020.

Moreover, the algorithm is important because other researchers could use it for potential investigations on specific issues related to the pandemic.

Methods

Source

A nationwide retrospective study was conducted using the HDRs on health care services in 2020, provided by more than 1300 public and private Italian hospitals. Hospital discharge data are routinely collected by the Italian MoH and contain patient demographic information (eg, gender and age), admission and discharge dates, up to 6 discharge diagnoses (ICD-9-CM), up to 11 medical procedures or surgical interventions, and status at discharge (eg, alive, deceased, or transferred to another hospital). In addition, the National Tax Registry was used to determine vital status or death following hospitalization. HDRs were linked with National Tax Registry records using deterministic record linkage.

Ethical Considerations

This study is conducted in accordance with the Ministerial Decree, “Evaluation in terms of quality, safety and appropriateness of activities provided for accreditation and contractual agreements with health facilities,” released on December 19, 2022, Article 4 [15]. This decree grants a mandate, inter alia, to the Italian National Agency for Regional Healthcare Services to monitor hospital activity and publish reports based on anonymized Italian hospitalizations. We did not include any identifiable data of patients’ personal information, including name, identity information, address, and telephone number; therefore, ethics approval was not required.
Algorithm to Identify COVID-19 Hospitalizations

Inpatient hospitalizations were detected by implementing an algorithm based on specific ICD-9-CM code combinations. This was carried out through the integration of 2 distinct approaches. One approach, deductive (ie, from the general rule to practical use), was based on the ICD-9-CM codes reported in the official documents of the MoH for COVID-19 case identifications. In addition, we used codes referring to SARS (though not specifically due to SARS-CoV-2), assuming that these codes, mainly at the beginning of the pandemic and before the MoH guidelines publication, were used to indicate COVID-19 hospitalizations. These codes are presented in Table 1.

<table>
<thead>
<tr>
<th>Codes sources</th>
<th>Codes</th>
</tr>
</thead>
</table>
| **First guidelines by MoH**<sup>a</sup> (March 20, 2020) | • 078.89 (other specified diseases due to viruses)  
• V01.79 (contact with or exposure to other viral diseases)  
• V07.0 (need for isolation)  
• V71.83 (observation and evaluation for suspected exposure to other biological agents) |
| **MoH Decree (October 28, 2020)** | • 043 (COVID-19 disease)  
• 480.4 (COVID-19 pneumonia)  
• 518.9 (COVID-19 acute respiratory distress syndrome)  
• 519.7 (COVID-19 other respiratory infections)  
• V01.85 (exposure to SARS-CoV-2)  
• V01.82 (exposure to SARS-associated coronavirus)  
• V07.00 (need for isolation after contact with SARS-COV-2)  
• V12.04 (personal history of COVID-19)  
• V71.84 (observation and evaluation for suspected exposure to SARS-COV-2) |
| **Use of ICD-9-CM<sup>b</sup> codes of SARS** | • 079.82 (SARS-associated coronavirus)  
• 480.3 (pneumonia due to SARS-associated coronavirus)  
• V01.82 (exposure to SARS-associated coronavirus) |

<sup>a</sup>MoH: Ministry of Health.  
<sup>b</sup>ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification.

The other approach, inductive (ie, from an empirical observation to a general rule) was used to confirm and integrate the algorithm. In detail, comparing the number of hospitalizations between 2019 and 2020, a dramatic increase was found in the principal diagnosis with the ICD-9-CM code 484.8 for “pneumonia in other infectious diseases classified elsewhere.” This code was presumed to be associated with COVID-19 and was integrated into the algorithm.

The hospitalizations were then grouped into clinical profiles according to a hierarchical approach based on 3 main criteria, as follows:

- The main reason for hospitalization: admissions categorized as “due to COVID-19” (codes identifying COVID-19 in the principal diagnosis) or “SARS-CoV-2 positive, but not due to COVID-19” (codes identifying COVID-19 only in secondary diagnosis).
- Degree of certainty: diagnosis confirmed or categorized as “suspected COVID-19”; the latter included hospitalizations with nonspecific diagnoses or V codes (exposure to SARS-CoV-2, observation, or isolation), for which it is difficult to ascertain if they refer to COVID-19 cases.
- Clinical presentation: pneumonia or acute respiratory distress syndrome, other acute respiratory infections, full-blown COVID-19 without respiratory symptoms, and asymptomatic or pauci/asymptomatic COVID-19.

All diagnosis codes and their combined use (algorithm) in the principal and secondary diagnoses used to select and classify COVID-19 hospitalizations are shown in Table S1 in Multimedia Appendix 1.

Statistical Analysis

Hospitalizations were analyzed with respect to the distribution of different clinical presentations associated with diagnoses of COVID-19.

A further analysis was performed only among patients hospitalized “due to COVID-19” from January 1 to December 31, 2020, to investigate potential risk factors associated with 30-day mortality. The cohort included all patients with at least 1 “due to COVID-19” hospitalization, excluding repeated admissions due to transfers to other hospitals or readmissions. Time-to-event techniques were used to analyze survival. The follow-up of patients began from the hospital admission date until either the ascertainment of death (outcome) or up to 30 days after admission (censoring).

The 30-day death rate was calculated as the total number of all-cause deaths up to 30 days after admission divided by the total follow-up days of hospitalized patients in the cohort and was reported per 1000 patient-days. To investigate the association of one or more risk factors with the 30-day death rate among patients hospitalized “due to COVID-19,” a multivariable Cox proportional-hazards regression model (Cox, 1972) was used, after having checked the proportional hazard assumption.

https://publichealth.jmir.org/2024/1/e44062
Sociodemographic data (eg, age, gender, area of residence, or nationality) were considered in the analysis. Patients were classified according to the following geographic areas of nationality: Italy or other high-income countries; migrants from high migration pressure countries (HMPC), which include Central and Eastern European countries (including those belonging to the European Union) and Malta; Northern African countries; other African countries; Central and South American countries, Asian countries (excluding Japan, Israel, and South Korea); and Oceania (excluding Australia and New Zealand) [16].

The transfer to ICU has been used as a proxy for the clinical severity of the disease.

The Charlson Comorbidity Index (CCI) [17,18] was used to account for current comorbidities or comorbidities over the previous 4 years in patients that could affect the death rate. To calculate the CCI, 17 medical conditions were considered [19,20]. The CCI was categorized into the following scores: CCI=0, CCI=1 or 2, and CCI³3.

To detect differences in the 30-day death rate during the year 2020, three pandemic periods were considered, as follows:

- First wave, including the Italian lockdown (January 1 to May 5, 2020)
- Postlockdown period (May 6 to October 7, 2020)
- Second wave (October 8 to December 31, 2020)

Two further multivariable Cox regression models were performed to assess temporal changes over the course of the pandemic; in particular, the 30-day death rates during the first and second waves were analyzed across 3 main geographical areas (ie, North, Center, and South and Islands) and by discharge wards (ie, ordinary and intensive care).

The selection of patients and hospital admissions was performed using SAS Studio (version 3.81; SAS Institute; Enterprise edition). The Cox proportional-hazards regression was performed using STATA (version 14.0; StataCorp). Tests were 2-sided, and statistical significance was set at \( P<.05 \).

Results

COVID-19 Hospitalizations Analysis

We identified a total of 325,810 hospitalizations with COVID-19-related diagnosis codes within the public and private accredited Italian hospitals throughout the year 2020 (Table 2).

A total of 239,114 (73.4%) out of the initial HDR selection were classified as “due to COVID-19” hospitalizations (in most cases, with a diagnosis of pneumonia and acute respiratory distress syndrome), and 47,416 (14.5%) were classified as “SARS-CoV-2 positive, but not due to COVID-19” hospitalizations, mainly in association with diseases of the circulatory system, injuries, chronic diseases of respiratory system, and complication of pregnancy, childbirth, and the puerperium.

The remaining 12.1% (n=39,280) were “suspected COVID-19 hospitalizations.” A high proportion of such hospitalizations (n=29,327, 9%) were also found in association with other pathological conditions reported in principal diagnosis.

There was a significant increase in the length of ICU stay out of the overall length of hospital stay (calculated in days), compared with the prepandemic period (1,031,037/54,170,360, 1.9% in 2019 vs 1,204,216/46,113,945, 2.6% in 2020; \( P<.001 \)).

In addition, this proportion was 2.6% (33,509/1,295,932) in 2020 for the “SARS-CoV-2 positive, but not due to COVID-19” and “suspected COVID-19” hospitalizations, while it increased to 8.4% (281,783/3,355,386) among the “due to COVID-19” hospitalizations (data are not shown in the Tables).
Table 2. The number of admissions categorized by reason for hospitalization and clinical presentation of COVID-19 in Italy, 2020.

<table>
<thead>
<tr>
<th>Reason for hospitalization and clinical presentation</th>
<th>Values (N=325,810), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Due to COVID-19, n=239,114 (73.4%)</strong></td>
<td></td>
</tr>
<tr>
<td>Pneumonia or acute respiratory distress syndrome</td>
<td>222,655 (93.1)</td>
</tr>
<tr>
<td>Other acute respiratory infections</td>
<td>2275 (1)</td>
</tr>
<tr>
<td>Full-blown COVID-19 (without respiratory symptoms)</td>
<td>5718 (2.4)</td>
</tr>
<tr>
<td>Pauciasymptomatic COVID-19</td>
<td>8466 (3.5)</td>
</tr>
<tr>
<td>Total</td>
<td>239,114 (100)</td>
</tr>
<tr>
<td><strong>SARS-CoV-2–positive but not due to COVID-19 (with secondary diagnoses of COVID-19), n= 47,416 (14.5%)</strong></td>
<td></td>
</tr>
<tr>
<td>Diseases of the circulatory system</td>
<td>10,318 (21.8)</td>
</tr>
<tr>
<td>Injuries</td>
<td>6038 (12.7)</td>
</tr>
<tr>
<td>Chronic diseases of the respiratory system</td>
<td>4424 (9.3)</td>
</tr>
<tr>
<td>Complication of pregnancy, childbirth, and the puerperium</td>
<td>4089 (8.6)</td>
</tr>
<tr>
<td>Diseases of the digestive system</td>
<td>3208 (6.8)</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>2993 (6.3)</td>
</tr>
<tr>
<td>Factors influencing health status</td>
<td>2844 (6)</td>
</tr>
<tr>
<td>Infectious and parasitic diseases</td>
<td>2441 (5.1)</td>
</tr>
<tr>
<td>Symptoms, signs, and ill-defined conditions</td>
<td>2328 (4.9)</td>
</tr>
<tr>
<td>Diseases of the genitourinary system</td>
<td>2030 (4.3)</td>
</tr>
<tr>
<td>Diseases of the nervous system and sense organs</td>
<td>1985 (4.2)</td>
</tr>
<tr>
<td>Diseases of the musculoskeletal system and connective tissue</td>
<td>1851 (3.9)</td>
</tr>
<tr>
<td>Mental disorders</td>
<td>1075 (2.3)</td>
</tr>
<tr>
<td>Other</td>
<td>1792 (3.8)</td>
</tr>
<tr>
<td>Total</td>
<td>47,416 (100)</td>
</tr>
<tr>
<td><strong>Due to suspected COVID-19, n=9953 (3.1%)</strong></td>
<td></td>
</tr>
<tr>
<td>Pneumonia or acute respiratory distress syndrome</td>
<td>8277 (83.2)</td>
</tr>
<tr>
<td>Other acute respiratory infections in suspected COVID-19</td>
<td>678 (6.8)</td>
</tr>
<tr>
<td>Need for isolation</td>
<td>998 (10)</td>
</tr>
<tr>
<td>Total</td>
<td>9953 (100)</td>
</tr>
<tr>
<td><strong>Associated with suspected COVID-19 (non–COVID-19 disease), n=29,237 (9%)</strong></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>325,810 (100)</td>
</tr>
</tbody>
</table>

“Due to COVID-19” Patients Analysis

Table 3 shows the baseline demographic and clinical characteristics of patients hospitalized “due to COVID-19.”

The cohort included 205,048 patients, with a median age of 72 years. Of these patients, 147,266 (71.8%) were older than 60 years, and there was a higher prevalence of male patients (n=124,181, 60.6%). Additionally, 6.1% (n=12,364) were HMPC citizens coming from Central and Eastern Europe (n=3721, 1.8%), Asia (n=2842, 1.4%), Central and South America (n=2637, 1.3%), Northern Africa (n=1786, 0.9%), and other African countries (n=1378, 0.7%). More than two-thirds of patients resided in the northern Italian regions, while 31.5% (n=64,491) were in Central and Southern Italy. Over the 3 periods of analysis, 39.4% (n=80,843) of patients were hospitalized in the first wave, 53.9% (n=110,446) were hospitalized in the second wave, and only 6.7% (n=13,759) were hospitalized in the postlockdown period. A total of 199,283 (97.2%) patients reported a Charlson Comorbidity Index score lower than or equal to 2.

The overall 30-day death rate in patients hospitalized “due to COVID-19” was 9.9 per 1000 person-days. After adjustment, mortality was lower for women (HR=0.83; P<.001) and for patients coming from HMPC. Among the latter, North African patients (HR=0.65; P<.001) and patients from Central and Eastern Europe (HR=0.66; P<.001) were around 35% less likely to die within 30 days, compared to patients coming from Italy and other high-income countries; Asian patients as well as Central and Southern American patients were 17% less likely to die within 30 days.
Table 3. The baseline demographic and clinical characteristics of patients hospitalized “due to COVID-19.”

<table>
<thead>
<tr>
<th>Characteristics and categories</th>
<th>Patients (N=205,048), n (%)</th>
<th>The 30-day death rate per 1000 person-days (overall=9.9 per 1000)</th>
<th>Adjusted hazard ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>124,181 (60.6)</td>
<td>10.2</td>
<td>1</td>
<td>N/Aa</td>
</tr>
<tr>
<td>Female</td>
<td>80,867 (39.4)</td>
<td>9.4</td>
<td>0.83</td>
<td>0.82-0.85</td>
</tr>
<tr>
<td><strong>Age (years)b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-40</td>
<td>10,164 (5)</td>
<td>0.5</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>41-50</td>
<td>15,421 (7.5)</td>
<td>1.1</td>
<td>2.18</td>
<td>1.80-2.63</td>
</tr>
<tr>
<td>51-60</td>
<td>32,197 (15.7)</td>
<td>2.2</td>
<td>3.77</td>
<td>3.17-4.49</td>
</tr>
<tr>
<td>61-70</td>
<td>39,080 (19.1)</td>
<td>5.8</td>
<td>8.99</td>
<td>7.590-10.66</td>
</tr>
<tr>
<td>&gt;70</td>
<td>108,186 (52.7)</td>
<td>17.5</td>
<td>28.98</td>
<td>24.49-34.30</td>
</tr>
<tr>
<td><strong>Nationality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy or other high-income</td>
<td>192,684 (93.9)</td>
<td>10.5</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>countries</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Africa</td>
<td>1786 (0.9)</td>
<td>2.3</td>
<td>0.65</td>
<td>0.54-0.78</td>
</tr>
<tr>
<td>Other African Countries</td>
<td>1378 (0.7)</td>
<td>1.4</td>
<td>0.77</td>
<td>0.59-1.00</td>
</tr>
<tr>
<td>Central and South American</td>
<td>2637 (1.3)</td>
<td>2.3</td>
<td>0.83</td>
<td>0.71-0.96</td>
</tr>
<tr>
<td>countries</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>2842 (1.4)</td>
<td>1.6</td>
<td>0.83</td>
<td>0.70-0.99</td>
</tr>
<tr>
<td>Central and Eastern Europe</td>
<td>3721 (1.8)</td>
<td>2.5</td>
<td>0.66</td>
<td>0.59-0.75</td>
</tr>
<tr>
<td><strong>Italian geographic areac</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>140,557 (68.5)</td>
<td>10.2</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>Center</td>
<td>34,195 (16.7)</td>
<td>8.4</td>
<td>0.90</td>
<td>0.88-0.93</td>
</tr>
<tr>
<td>South and Islands</td>
<td>30,296 (14.8)</td>
<td>10.3</td>
<td>1.17</td>
<td>1.14-1.20</td>
</tr>
<tr>
<td><strong>Periods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>January 1 to May 5, 2020</td>
<td>80,843 (39.4)</td>
<td>11</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>May 6 to October 7, 2020</td>
<td>13,759 (6.7)</td>
<td>4.4</td>
<td>0.45</td>
<td>0.43-0.47</td>
</tr>
<tr>
<td>October 8 to Dec 31, 2020</td>
<td>110,446 (53.9)</td>
<td>9.9</td>
<td>0.85</td>
<td>0.83-0.86</td>
</tr>
<tr>
<td><strong>Charlson Comorbidity Index</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>96,225 (46.9)</td>
<td>6.7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1 or 2</td>
<td>103,058 (50.3)</td>
<td>12.7</td>
<td>1.37</td>
<td>1.35-1.40</td>
</tr>
<tr>
<td>≥3</td>
<td>5765 (2.8)</td>
<td>19.4</td>
<td>1.68</td>
<td>1.61-1.76</td>
</tr>
<tr>
<td><strong>Ward of discharge</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordinary care</td>
<td>176,801 (86.2)</td>
<td>8.5</td>
<td>1</td>
<td>N/A</td>
</tr>
<tr>
<td>Intensive care</td>
<td>28,247 (13.8)</td>
<td>19.5</td>
<td>2.55</td>
<td>2.50-2.60</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
bMedian age was 72 years.
cThe North includes Valle d’Aosta, Piemonte, Liguria, Lombardia, Trentino-Alto Adige, Friuli-Venezia Giulia, Veneto, and Emilia-Romagna regions. The Center includes Toscana, Umbria, Marche, and Lazio regions. The South and Islands include Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, and Sardegna regions.

In addition, we found a higher 30-day survival in the central regions compared with the northern regions, which were the most affected by COVID-19 (HR=0.90; P<.001), while the southern regions and the Islands showed a higher risk of 30-day mortality (HR=1.17; P<.001).

Transfers to ICUs involved 13.8% (28,247/205,048) of hospitalized patients.

Patients referred to ICUs were two and a half times more likely to die within 30 days, compared to patients hospitalized in ordinary wards (HR=2.55; P<.001).
Table 4 shows the survival analysis by geographic area and wards during 2 pandemic waves.

Over the first wave, among patients in ordinary wards, we found a higher 30-day death rate in the northern regions (10.9 per 1000) compared to the central (7.7 per 1000) and southern islands regions (5.6 per 1000). On the contrary, among patients with a transfer to ICUs, the southern regions and the Islands showed a higher 30-day death rate (28.4 per 1000) compared to the northern regions (15.2 per 1000).

Table 4. Survival analysis of patients in hospitals due to COVID-19 during the first and the second waves of the pandemic by geographic area and wards of discharge in Italy in 2020.

<table>
<thead>
<tr>
<th>Wards and geographic area</th>
<th>First wave (N=80,843)</th>
<th>Second wave (N=110,446)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients, n</td>
<td>30-day death rate (per 1000 person-days)</td>
</tr>
<tr>
<td>Ordinary care (first wave: n=68,880; second wave: n=95,915)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>56,849</td>
<td>10.9</td>
</tr>
<tr>
<td>Center</td>
<td>7591</td>
<td>7.7</td>
</tr>
<tr>
<td>South and Islands</td>
<td>4440</td>
<td>5.6</td>
</tr>
<tr>
<td>Intensive care unit (first wave: n=11,963; second wave: n=14,531)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>9144</td>
<td>15.2</td>
</tr>
<tr>
<td>Center</td>
<td>1726</td>
<td>15.5</td>
</tr>
<tr>
<td>South and Islands</td>
<td>1093</td>
<td>28.4</td>
</tr>
</tbody>
</table>

aAdjusted for gender, age class, nationality, and Charlson Comorbidity Index.
bN/A: not applicable.

Discussion

Principal Findings

In Italian hospitals, due to the lack of COVID-19 codes in the ICD-9-CM system, multiple codes were used for COVID-19 patient admissions throughout the pandemic outbreak; this made it difficult to identify COVID-19 hospitalizations. This study enabled the definition of a specific combination of codes to identify COVID-19 hospitalizations at the national level within the HIS. Additionally, it facilitated the investigation of risk factors associated with mortality due to COVID-19 among patients admitted to Italian hospitals in 2020. The selection algorithm made it possible to identify 239,114 hospitalizations “due to COVID-19” from HDR. This number is very similar to the data on hospital admissions gathered from the COVID-19 Integrated Surveillance System, which amounted to 240,542 [11]. The monthly trend (data are not shown in Tables) is also very similar in the 2 data sources. Only in a few regions, the number of cases derived from HDRs is lower than the number obtained from the Surveillance System, especially in the first wave of the pandemic, likely due to a miscoding of data at the beginning of the epidemiological crisis. Nevertheless, the use of HDR data for COVID-19 impact analyses remains preferable to Surveillance System data for several reasons. The Surveillance System is based on data provided by the regions to monitor the evolution of the pandemic with timely data. Nevertheless, HDR data, although less timely, have higher quality, as they are collected following criteria and rules established by national legislation and guidelines, which guarantee greater homogeneity and comparability across different periods and territories. Moreover, HDR contains more information than the Surveillance System, especially on comorbidities, and allows comparisons with non–COVID-19 hospitalizations as well.

Throughout the second wave, among patients in ordinary wards, there was a significant decrease in 30-day death rates in the northern regions (8.7 per 1000), whereas such rates increased up to 7.4 per 1000 person-days in the southern regions and the Islands. On the other hand, among patients with a transfer to ICUs, the 30-day death rate increased in all geographic areas, particularly in the southern regions and the Islands, where the number of ICU patients more than tripled from the first to the second wave; this resulted in a significantly higher mortality risk in the southern regions and the Islands compared to the northern ones (HR=2.52; P<0.001).

Overall, we observed a decrease in the 30-day death rate among patients hospitalized “due to COVID-19” between the 2 waves of the pandemic. This cannot be explained by the introduction of new pharmacological therapies [21]; however, it could be due to a better use of existing therapies associated with gained experience in managing the disease by health care providers and a more effective preparedness of the health care system, including reorganization of hospitals and increasing number of ICU beds for patients with COVID-19 [22].

In particular, in the second wave, mortality was reduced in the northern regions among patients treated in ordinary wards, with a lower proportion of cases referred to ICUs (from 9144/65,993, 13.9% to 7990/67,406, 11.9%); this may have led to a higher concentration of patients with severe symptoms in ICUs, consistent with a small increase in 30-day death rates (from 15.2 to 17.0 per 1000 person-days). On the contrary, in the southern regions and the Islands, mortality increased between the 2 waves, both in the ordinary wards and in the ICUs, and
this might be due to a higher burden on hospital care compared to the first wave, which had affected Southern Italy to a lesser extent. Furthermore, the South reported a decrease in non–COVID-19 hospitalizations, which was greater than the national mean [23], a sign indicating that the hospital system was struggling.

Our study also highlighted the lower 30-day death rate for women compared to men, regardless of age. The observed disparity may be attributed to sex-based differences in immunological responses, comorbidities prevalence, or differences in behaviours [24,25]. Consistent with several studies, this analysis showed that critically ill older patients were at high risk of disease severity and mortality [26,27]. Some studies also highlighted important post–COVID-19 sequelae in patients with cancer, which adversely affect survival and oncological outcomes after recovery [28].

Finally, our findings, consistent with the findings of a study carried out in the city of Milan [29], confirmed that mortality among migrants in hospitals due to COVID-19 was lower than that of Italians and people coming from high-income countries. The lower 30-day death rate observed among HMPC migrants can be explained as the effect of specific selection processes (especially in the early and late stages of the migration path), which tend to maintain the overall health status of the foreign population high. One of them is the so-called “healthy migrant effect,” a dynamic for which only people in good health tend to emigrate [30]. Another selection dynamic, known as the “salmon effect,” is due to the habit of older migrants, especially if ill, to go back to their country of origin [31].

Moreover, an Italian study on data retrieved from the COVID-19 Integrated Surveillance System found an interaction between nationality and the epidemic phases; the analysis stratified by year showed that compared to Italians, mortality and other severe clinical outcomes in non-Italian nationals slightly increased from the last months of 2020 through 2021. The authors hypothesize that reduced access to vaccination by the immigrated population could partly explain such an increase [32], as a high risk of missed vaccination was observed among non–Italian-born people living in Italy, both overall and for individuals ≥50 years of age [33].

Limitations
This study has several limitations. First, results may be affected by the information quality of the HDR or by missing or incorrect coding of some diagnoses, leading to possible underestimation of comorbidities. Therefore, to minimize the effect of underreporting bias, comorbidities over the previous 4 years were also retrieved.

Second, the study is based on the HDR, which collects administrative data and does not contain clinical information of patients, which is useful to better characterize COVID-19 severity. For this reason, the transfer to the ICU has been used in this study as a proxy for the clinical severity of the disease.

Third, due to the lack of specified COVID-19 codes from the onset of the pandemic, the overall hospitalizations detected through the algorithm might be underestimated.

Finally, the issue of patients hospitalized due to “suspected COVID-19,” which cannot be defined as COVID-19 cases, remains unsolved.

Conclusions
To our knowledge, the algorithm represents the first attempt to define, at the national level, selection criteria for identifying COVID-19 hospitalizations within the HIS. While awaiting the adoption of more updated International Classification of Diseases coding systems for hospital diagnoses and procedures in Italy, consistent with the World Health Organization recommendations, the implemented algorithm will be used to monitor the pandemic over time, and the patients selected in 2020 will be followed up in the subsequent years to assess the long-term effects of COVID-19.

Further analyses will be useful to assess the impact of the anti–COVID-19 vaccination campaign on the severity and mortality of the disease and to investigate possible inequalities between population subgroups.

Acknowledgments
We would like to thank Roberto Blaco for his technical support about the COVID-19 coding, Stella Lanzi for the Italian-to-English translation, Simona Gorietti for reference editing, and Elisa Guglielmi for organizational and administrative support. These individuals were not compensated for their role in the study.

This research received no external funding.

Data availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request, by contacting direzione.ricerca@agenas.it.

Authors' Contributions
All authors discussed the results and contributed to the final manuscript. All authors read and approved the final manuscript.

Conflicts of Interest
None declared.
References


11. COVID-19 integrated surveillance data in Italy. URL: [https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-integrated-surveillance-data] [accessed 2022-09-09]


Abbreviations

CCI: Charlson Comorbidity Index
HDR: hospital discharge record
HIS: Hospital Information System
HMPC: high migration pressure countries
HR: hazard ratio
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10: International Classification of Diseases, 10th Revision
ICD-11: International Classification of Diseases, 11th Revision
ICU: intensive care unit
MoH: Italian Ministry of Health
Effects of Face Mask Mandates on COVID-19 Transmission in 51 Countries: Retrospective Event Study

Anatol-Fiete Näher1,2,3, MA, MD, Dr med; Matthias Schulte-Althoff2,4, MSc; Marvin Kopka2,5, MSc; Felix Balzer2, MSc, MD, PhD, Dr med; Francisco Pozo-Martín6, MSc, PhD

1Digital Global Public Health, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
2Institute for Medical Informatics, Charité - Universitätsmedizin, Berlin, Germany
3Method Development, Research Infrastructure, and Information Technology, Robert Koch Institute, Berlin, Germany
4Department of Information Systems, School of Business and Economics, Freie Universität, Berlin, Germany
5Department of Psychology and Ergonomics, Technische Universität Berlin, Berlin, Germany
6Evidence-based Public Health Unit, Center for International Health Protection, Robert Koch Institute, Berlin, Germany

Corresponding Author:
Anatol-Fiete Näher, MA, MD, Dr med
Digital Global Public Health
Hasso Plattner Institute
University of Potsdam
Rudolf-Breitscheid-Strasse 187
Campus III, Bldg G2
Potsdam, 14482
Germany
Phone: 49 331 5509176
Email: anatolfiete.naeher@hpi.de

Abstract

Background: The question of the utility of face masks in preventing acute respiratory infections has received renewed attention during the COVID-19 pandemic. However, given the inconclusive evidence from existing randomized controlled trials, evidence based on real-world data with high external validity is missing.

Objective: To add real-world evidence, this study aims to examine whether mask mandates in 51 countries and mask recommendations in 10 countries increased self-reported face mask use and reduced SARS-CoV-2 reproduction numbers and COVID-19 case growth rates.

Methods: We applied an event study approach to data pooled from four sources: (1) country-level information on self-reported mask use was obtained from the COVID-19 Trends and Impact Survey, (2) data from the Oxford COVID-19 Government Response Tracker provided information on face mask mandates and recommendations and any other nonpharmacological interventions implemented, (3) mobility indicators from Google’s Community Mobility Reports were also included, and (4) SARS-CoV-2 reproduction numbers and COVID-19 case growth rates were retrieved from the Our World in Data—COVID-19 data set.

Results: Mandates increased mask use by 8.81 percentage points ($P=.006$) on average, and SARS-CoV-2 reproduction numbers declined on average by $-0.31$ units ($P=.008$). Although no significant average effect of mask mandates was observed for growth rates of COVID-19 cases ($-0.98$ percentage points; $P=.56$), the results indicate incremental effects on days 26 ($-1.76$ percentage points; $P=.04$), 27 ($-1.89$ percentage points; $P=.05$), 29 ($-1.78$ percentage points; $P=.04$), and 30 ($-2.14$ percentage points; $P=.02$) after mandate implementation. For self-reported face mask use and reproduction numbers, incremental effects are seen 6 and 13 days after mandate implementation. Both incremental effects persist for $>30$ days. Furthermore, mask recommendations increased self-reported mask use on average ($5.84$ percentage points; $P<.001$). However, there were no effects of recommendations on SARS-CoV-2 reproduction numbers or COVID-19 case growth rates ($-0.06$ units; $P=.70$ and $-2.45$ percentage points; $P=.59$). Single incremental effects on self-reported mask use were observed on days 11 (3.96 percentage points; $P=.04$) and 25 to 27 (4.20 percentage points; $P=.048$ and 5.91 percentage points; $P=.01$) after recommendation. Recommendations also affected reproduction numbers on days 0 ($-0.07$ units; $P=.03$) and 1 ($-0.07$ units; $P=.03$) and between days 21 ($-0.09$ units; $P=.04$) and 28 ($-0.11$ units; $P=.05$) and case growth rates between days 1 and 4 ($-1.60$ percentage points; $P=.03$ and $-2.19$ percentage points; $P=.03$) and on day 23 ($-2.83$ percentage points; $P=.05$) after publication.
Conclusions: Contrary to recommendations, mask mandates can be used as an effective measure to reduce SARS-CoV-2 reproduction numbers. However, mandates alone are not sufficient to reduce growth rates of COVID-19 cases. Our study adds external validity to the existing randomized controlled trials on the effectiveness of face masks to reduce the spread of SARS-CoV-2.

(JMIR Public Health Surveill 2024;10:e49307) doi:10.2196/49307

KEYWORDS
nonpharmacological interventions; face masks; infectious diseases; acute respiratory infections; COVID-19; real-world evidence

Introduction

Background

Viral and bacterial acute respiratory infections (ARIs) are among the leading causes of death worldwide. With a total of 3.8%, ARIs contributed the fourth largest share to the global burden of disease in 2019 [1]. Thus, ARIs pose a significant threat to global health. A measure to prevent ARI, which has also been discussed extensively in the context of the COVID-19 pandemic, is the wearing of face masks. At the heart of this discussion are the following 2 considerations: first, a large proportion of COVID-19 infections are transmitted by asymptomatic individuals [2,3], and second, masks worn by infected individuals are thought to reduce transmission risks. Therefore, in the face of a potentially large number of undetected cases, it has been argued that wearing masks could contribute to a sustained decrease in the spread of COVID-19 [4]. In contrast, a recently updated Cochrane review did not reach a clear conclusion regarding the preventive efficacy of face masks for viral respiratory infections [5]. On the basis of 78 randomized controlled trials (RCTs) conducted during influenza seasons as well as during the H1N1 influenza and COVID-19 pandemics, the authors conclude that masks alone are not sufficient to reduce the spread of respiratory viruses. However, given the results, it should be noted that interpretation is complicated by the fact that adherence was low, and outcome measures varied across a large proportion of RCTs. Furthermore, the observations of Jefferson et al [5] are at odds with clinical trials demonstrating the efficacy of masks in protecting uninfected individuals from COVID-19 [6,7]. Clinical trials have also shown that masks reduce the risk of transmission of influenza or SARS-CoV-2 from previously infected individuals [8].

A key advantage of studies that are based on real-world data is their high external validity. As such, they can significantly complement the evidence from RCTs on the efficacy of face masks in protecting against ARIs. ARIs can be most effectively prevented at the population level when masks are worn by both susceptible and infectious individuals [8,9]. Following this line of reasoning, the World Health Organization issued a recommendation to publicly wear face masks to contain the spread of SARS-CoV-2 [10]. This recommendation has been adopted by countries around the world by implementing face mask mandates in 2020. As some countries and regions did not introduce any face mask policies, 3 studies used this fact to identify the causal effect of mandates on SARS-CoV-2 infections. Lyu and Wehby [11] conducted a natural experiment with real-world data to show that mandates in 15 states of the United States and Washington, District of Columbia, resulted in reductions in COVID-19 growth rates of up to 19%. Another study using real-world data from the Global COVID-19 Trends and Impact Survey (CTIS) found mask mandates in the United States to be related to decreases in daily new cases, daily new deaths, daily new hospital admissions, and increases in population shares wearing masks [12]. These findings are corroborated by Mitze et al [13], who, using a synthetic control design, found a 47% reduction in daily COVID-19 case growth rates attributable to a local mask mandate in Jena, Germany. In addition, 2 other population-level studies found evidence for a direct association between wearing face masks and COVID-19 outcomes. Controlling for time-constant unobserved heterogeneity, Rader et al [14] report decreasing growth rates of COVID-19 cases with increasing state-specific proportions of face mask wearers at the US state level. Leffler et al [15] use a cross-sectional design across 196 countries and observe that longer durations of face mask use are inversely associated with COVID-19 mortality.

Objectives

To our knowledge, no real-world studies have been conducted to compare the effects of face mask mandates on SARS-CoV-2 infections internationally. Our study provides missing evidence. A prerequisite for mask mandates to be effective is that they encourage populations to wear masks [5,16]. Therefore, we exploited longitudinal variations in outcomes to first evaluate the hypothesis that mask mandates increase mask use in 51 countries worldwide. We then used the same design to test the hypothesis that mask mandates lead to a decrease in SARS-CoV-2 reproduction numbers and growth rates in COVID-19–related cases. In planning future nonpharmacological interventions, it is also important to be able to assess whether mandates are necessary to achieve the desired effect on health outcomes or whether recommendations are sufficient. Thus, the hypothesis that recommendations for face masks would lead to an increase in self-reported face mask use and a reduction in SARS-CoV-2 reproduction numbers and COVID-19 case growth rates was also evaluated. COVID-19 is a specific type of ARI. Except for rhinovirus infections, ARIs are comparable in terms of their transmission routes. Therefore, the results of our study not only provide an indication of the effectiveness of face mask interventions in the context of the COVID-19 pandemic but may also be indicative of how future ARI outbreak scenarios could potentially be addressed.
**Methods**

**Measures**

We gathered information on the proportions of face mask wearers in 105 countries from the CTIS, which is administered by the Carnegie Mellon University and the University of Maryland in partnership with Meta [17]. The CTIS uses an administrative region-stratified random sample that was drawn daily from Facebook’s active user base. Via Facebook’s news feed, the drawn individuals were invited to participate in the survey. After obtaining informed consent, users were provided with a link to a web-based questionnaire on COVID-19–related attitudes, symptoms, and behaviors. Daily population shares of self-reported face mask users were calculated from the responses of survey participants who stated that they had worn a face mask always or most of the time when in public. Before aggregation, the individual-level observations were reweighted by means of survey weights provided by Meta. Although no information collected by the survey is shared with Meta, Meta itself calculated weights that adjust for nonresponse and sampling frame mismatches with country-specific populations based on age, gender, and administrative regions [18]. Weights were made available to researchers without disclosing any user-related content.

To allow for linear model specifications, we chose 2 country-specific COVID-19 outcome measures: SARS-CoV-2 reproduction numbers and 3-day growth rates of COVID-19–associated cases per 100,000 inhabitants. To obtain information on the progression of reproduction numbers over time, we used the Our World in Data—COVID-19 data set [19]. The daily effective reproduction numbers included in the data set were derived using Kalman filter estimates of weekly case growth rates, as described by Arroyo-Marioli et al [20].

Data on nonpharmacological interventions were obtained from the Oxford COVID-19 Government Response Tracker (OxCGRT) [21]. Between January 1, 2020, and December 31, 2022, the tracker collected the start and end dates of COVID-19–related policy interventions in 186 countries. The OxCGRT data set also includes a measure that captures the scope of face mask–related policies on an ordinal scale from 1 to 4. On this scale, 1 designates official face mask recommendations. Mandates are indicated by values ranging from 2 to 4. Using this scale, we defined the levels of mandates according to the extent to which exemptions from the obligation to wear face masks in public were made. Corresponding to a value of 2, level 1 applies to all mandates that required face masks to be worn in some public places where other people were present or where social distancing was not possible, such as certain shops or public transport. Level 2 is equivalent to a value of 3, which corresponds to mandates that were more restrictive and generally required people to wear face masks in public places where other people were present or where social distancing was not possible. Level-2 mandates only allowed people not to wear masks in places such as less crowded streets or parks. Level 3 corresponds to a value of 4 and means that face masks had to be worn outdoors at all times, regardless of location or the presence of other people. We created dummy variables for each mandate level to indicate whether the mandates at a particular level were enacted at a particular time. The OxCGRT data further include an item that differentiates between policies targeted at subgroups or entire populations. A control variable for country-specific face mask mandates that applied only to subgroups was generated from this item and the indicators for level-1 face mask mandates. The variable takes the value 1 at dates when the corresponding subgroup mandates were active; otherwise, it takes the value 0. This allows for the inclusion of all subgroup mandates. None of the included countries implemented subgroup mandates at a level higher than level 1. Furthermore, we added date-specific indicators of the following additional nonpharmacological interventions: school closures, bans on events and gatherings, international travel restrictions, curfews, and measures to protect older adults. Measures to protect older adults include recommendations or restrictions on the number of visitors and hygiene practices in nursing homes. At a given time, indicators for additional nonpharmacological interventions indicate whether the intervention is recommended or mandatory, as opposed to not recommended or mandatory. Country-level mobility indicators are also included in the analyses. The mobility indicator is calculated by Google from aggregated location data that users have agreed to share through their devices. The metric indicates the daily relative changes in mobility as compared with the baseline period between January 3 and February 6, 2020 [22].

**Observational Sample**

For the analyses, we constructed a longitudinal data set by merging daily repeated cross-sections of the CTIS with the OxCGRT, Our World in Data, and Google mobility data sets. Beginning on April 23, 2020, when the first CTIS data became available, the observation period was restricted to October 31, 2020, yielding a total of 192 observation dates in calendar time. In total, 102 countries with national-level policy interventions targeted at general populations are commonly represented in all 3 data sets. Of these, 78.4% (80/102) enacted national mandates on the use of face masks in public. As the recording of the CTIS data commenced only after the interventions, the effects of face mask policies could not be identified for 43% (34/80) of the mandate countries. These countries were excluded from the analysis. Another 4% (3/80) of the mandate countries were excluded because information on SARS-CoV-2 reproduction numbers was not available at the time when mask mandates became effective.

The CTIS data and information on SARS-CoV-2 reproduction numbers and COVID-19 case growth rates were available for 100% (8/8) of the countries that did not enact any face mask policies throughout the observation period. These countries served as the control group. As can be inferred from Table 1 [17,21], this resulted in a total of 51 observed countries worldwide, out of which 84% (43/51) countries implemented level-1 face mask interventions. The countries included were Albania, Azerbaijan, Belgium, Bangladesh, Belarus, Bolivia, Costa Rica, Germany, Denmark, Egypt, Finland, France, Ghana, Greece, Honduras, Croatia, Hungary, Ireland, Jordan, Japan, Kazakhstan, South Korea, Kuwait, Libya, Sri Lanka, Moldova, Myanmar, Nigeria, Nicaragua, the Netherlands, Nepal, New Zealand, Oman, Panama, the Philippines, Portugal, Paraguay,

---

https://publichealth.jmir.org/2024/1/e49307

JMIR Public Health Surveill 2024 | vol. 10 | e49307 | p.1486

(page number not for citation purposes)
Palestine, Qatar, Romania, Saudi Arabia, Sudan, Serbia, Slovakia, Sweden, Tunisia, Turkey, Uruguay, Uzbekistan, Yemen, and South Africa.

<table>
<thead>
<tr>
<th>Table 1. Observations and CTIS respondents in mandate and nonmandate countries. Data sources: CTIS and Oxford COVID-19 Government Response Tracker.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Countries (n=51), n (%)</td>
</tr>
<tr>
<td>No mandate</td>
</tr>
<tr>
<td>Mandate</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Empirical Strategy

To evaluate face mask mandates, we follow the recommendations of Lison et al [23] for effectiveness assessments of nonpharmacological interventions. A test of our hypotheses is provided by the following potential outcomes framework: Let $y_{1,ct}$ denote a potential outcome of interest in country $c=1, 2, ..., C$ at calendar time $t=1, 2, ..., T$ if a mandate is implemented, $y_{0,ct}$ designates the potential outcome of interest in the same country if no mandate has been implemented. The average treatment effect on the outcome of interest in countries treated with mandates (average treatment effect on the treated [ATT]) is then defined as

$$\beta = \frac{E[y_{1,ct} - y_{0,ct}|D_{ct}=1]}{P(D_{ct}=1)}$$ (1)

where $D_{ct}=(0, 1)$ is a mandate indicator switching on during mandate period $s=1, 2, ..., S; S<T$. As any given country is only observed for either $D_{ct}=1$ or $D_{ct}=0$, the observed outcomes of interest $y_{ct}$ can be written

$$y_{ct} = y_{0,ct} + (y_{1,ct} - y_{0,ct}) \times D_{ct}$$ (2)

When fitted to real-world data such as those generated with the CTIS, models require additional identifying assumptions to provide causal interpretations of $\beta$. A first assumption that is being made is that, in the absence of any mandates, outcomes would exhibit parallel trends for intervention and nonintervention countries. The parallel trends assumption ensures that the estimates of the mandate effect are not biased by unobserved time-varying heterogeneity. The identification of the average treatment effect on countries with face mask mandates, $\beta$, further requires that country populations do not demonstrate any self-reported anticipatory uptake of face masks or decreases in SARS-CoV-2 reproduction numbers and COVID-19 case growth rates before mandate introduction. In terms of Granger causality, a causal interpretation can be given to estimate $\beta$ only if the mandate effects are observed after the intervention. In addition, we assume face mask mandates to homogeneously affect all intervention countries in all periods.

It must also be noted that the estimates $\beta$ are unbiased only under the strict exogeneity of $D_{ct}$. Furthermore, our empirical strategy aims to estimate changes in the levels of self-reported face mask use, SARS-CoV-2 reproduction numbers, and COVID-19 case growth rates. Countries with large differences in outcome baseline levels can therefore be included in the analysis without violating the underlying assumptions regarding the functional form of the model specifications, shown below in equation 3. On the basis of the preceding assumptions, we estimate the average mandate effects on the treated countries $\beta$ by

$$z_{ct} = \lambda_c + \phi_c + \beta_{ct} D_{ct} + \gamma X_{ct} + \epsilon_{ct}$$ (3)

whereby $y_{1,ct} = \exp(z_{ct})$ / 1+$\exp(z_{ct})$ designates the expected share of self-reported face mask users or $y_{ct}=z_{ct}$ indicates Sars-CoV-2 reproduction numbers or case growth rates, respectively. Within this specification strategy, country and calendar time fixed effects $\lambda_c$ and $\phi_c$ provide full nonparametric control for any unobserved time–constant and country-constant heterogeneity [24]. $X_{ct}$ represents a row vector of covariates containing subgroup treatment and treatment-level indicators. Measures of COVID-19 outcomes depend on country-specific testing rates [25]. Therefore, $X_{ct}$ also contains the number of country-specific COVID-19 tests per 100,000 inhabitants. Moreover, nonpharmacological interventions other than mask mandates as well as domestic mobility are a possible source of heterogeneity in SARS-CoV-2 reproduction numbers and COVID-19 case growth rates during the observation period. This may lead to biased estimates of the mandate effects. For SARS-CoV-2 reproduction numbers and COVID-19 case growth rates, $X_{ct}$ thus includes indicators for both country-level mobility and nonpharmacological interventions other than mask mandates. $\gamma$ is a column vector designating the corresponding parameters of interest. Assuming a quasi-binomial distribution of country proportions of self-reported face mask wearers, an error is given by $\epsilon_{ct} \sim \text{Logistic}(\mu, \phi)$, with mean $\mu$ and dispersion parameter $\phi$. $\epsilon_{ct} \sim (0, \sigma^2)$ is assumed for linear models of SARS-CoV-2 reproduction numbers and COVID-19 case growth rates. By specifying $r=0$, $T_0$ leads and lags of the mandate effect, with $T_0<0$ being the lowest and $T_0 \geq 0$ the highest number of leads and lags considered, we use a design that incorporates incremental mandate effects. This specification provides a test for the previously outlined assumption of no anticipatory events. As countries imposed face mask mandates at different calendar dates, using already treated countries as control groups for yet-to-be treated countries is likely to induce heterogeneities of policy effects across treated units. Therefore, estimating equation 3 with conventional two-way fixed effects methods would result in biased estimates $\beta$. Hence, to relax the treatment homogeneity assumption made in the preceding section, we group countries into cohorts depending on the calendar dates.
of mandates becoming effective, as suggested by Sun and Abraham [26]. Countries without any face mask mandates during the entire observation period are combined into a single control group. Cohort outcomes at \( r = -1 \), that is, the period before mandate implementation, are used as baseline outcomes. Estimates of average treatment effects are obtained by weighing the aggregated cohort-specific effects by cohort size.

**Ethical Considerations**

Only publicly available data aggregated at the country level have been analyzed. This eliminates any reidentification risks for individuals. The study did not require any direct or indirect interaction with humans. This means that the regulatory requirements for research involving humans are not met. Therefore, this study is exempt from institutional review board approval.

**Results**

**Descriptive Results**

Figure 1 [17,18,21] illustrates the daily average evolution of self-reported face mask use, SARS-CoV-2 reproduction numbers, and COVID-19 case growth rates in countries with mask mandates. Visual inspection of panel A reveals a substantial increase in self-reported face mask wearers around the time the regulations came into effect in these countries. As further shown in the panels B and C, no clear trends are visible for average SARS-CoV-2 reproduction numbers or growth rates of COVID-19 cases.

The timing of the introduction of face mask mandates in each country during the observation period is presented in Figure 2 [21]. There is a range of 186 days between the dates when mandates were introduced: although in Egypt and Qatar the wearing of masks in public was compulsory from 4 days after the start of the observation period, that is, from April 26, 2020, in Libya, it was only compulsory from October 29, 2020, that is, 3 days before the end of the observation period.

Effects of Face Mask Mandates

We first estimated equation 3 for proportions of self-reported face mask users, SARS-CoV-2 reproduction numbers, and growth rates of COVID-19–related cases. All postmandate period effects were aggregated to the total ATT of face mask mandates. As shown in Table 2 [17,19,21,22], the results indicate that—on average—mandates are followed by an increase in self-reported face mask wearers of 8.81 percentage points ($P = .006$).

Similarly, an average decrease of $-0.31$ units ($P = .008$) was observed for SARS-CoV-2 reproduction numbers after face mask mandates became effective. No reduction in the growth rates of COVID-19 cases was observed after mask mandates became effective ($-0.98$ percentage points; $P = .56$).

The incremental effects of pre- and postmandate implementation periods according to the event study specification of equation 3 are presented in Figure 3 [17,19,21,22,26]. As can be seen, the incremental effects of face mask mandates are demonstrated in an increase in the proportion of self-reported mask wearers at 6 days and decreases in reproduction numbers at 13 days after mandate implementation. For self-reported face mask use and reproduction numbers, the effects continuously persist for $>30$ days. Furthermore, our estimations yield single mandate effects on case growth rates on days $26$ ($-1.76$ percentage points; $P = .04$), $27$ ($-1.89$ percentage points; $P = .05$), $29$ ($-1.78$ percentage points; $P = .04$), and $30$ ($-2.14$ percentage points; $P = .02$) after mandates were enacted. We observe pretrends in self-reported face mask use on up to 4 days and on day 4 in case growth rates before the beginning of the mandate periods. Our identifying assumption of no anticipatory events is supported for a period of at least 4 days before the beginning of mandates for all outcomes.
Table 2. Event study regressions of self-reported face mask use, SARS-CoV-2 reproduction numbers, and growth rates of COVID-19 cases on face mask mandates. Percentage point changes are shown for self-reported face mask use and growth rates of COVID-19 cases, and unit changes are shown for SARS-CoV-2 reproduction numbers. Data sources: COVID-19 Trends and Impact Survey, Oxford COVID-19 Government Response Tracker, Our World in Data—COVID-19 data set, and Google COVID-19 Community Mobility Reports.

<table>
<thead>
<tr>
<th>Event</th>
<th>Self-reported face mask use</th>
<th>SARS-CoV-2 reproduction number</th>
<th>Case growth rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage points (SE) P value</td>
<td>Percentage points (SE) P value</td>
<td>Percentage points (SE) P value</td>
</tr>
<tr>
<td>ATT</td>
<td>8.81 (3.08) 0.006</td>
<td>-0.31 (0.11) 0.008</td>
<td>-0.98 (1.66) 0.56</td>
</tr>
<tr>
<td>Subgroup mandate</td>
<td>7.33 (4.76) 0.13</td>
<td>0.03 (0.09) 0.78</td>
<td>1.21 (1.54) 0.44</td>
</tr>
<tr>
<td>Level-2 mandate</td>
<td>4.49 (4.88) 0.36</td>
<td>0.16 (0.11) 0.14</td>
<td>1.06 (1.17) 0.37</td>
</tr>
<tr>
<td>Level-3 mandate</td>
<td>2.43 (4.62) 0.60</td>
<td>0.0006 (0.11) &gt;0.99</td>
<td>1.39 (2.13) 0.52</td>
</tr>
<tr>
<td>School closure</td>
<td><em>d</em></td>
<td><em>—</em></td>
<td><em>—</em></td>
</tr>
<tr>
<td>Ban on events</td>
<td><em>—</em></td>
<td><em>—</em></td>
<td>-0.11 (0.05) 0.05</td>
</tr>
<tr>
<td>Ban on gatherings</td>
<td><em>—</em></td>
<td><em>—</em></td>
<td>-0.04 (0.04) 0.32</td>
</tr>
<tr>
<td>Curfew</td>
<td><em>—</em></td>
<td><em>—</em></td>
<td>0.01 (0.06) 0.93</td>
</tr>
<tr>
<td>International travel restrictions</td>
<td><em>—</em></td>
<td><em>—</em></td>
<td>-0.13 (0.06) 0.05</td>
</tr>
<tr>
<td>Protection of older adults</td>
<td><em>—</em></td>
<td><em>—</em></td>
<td>0.06 (0.09) 0.46</td>
</tr>
<tr>
<td>Mobility</td>
<td><em>—</em></td>
<td><em>—</em></td>
<td>0.01 (0) 0.10</td>
</tr>
<tr>
<td>Tests per 100,000 inhabitants</td>
<td><em>—</em></td>
<td><em>—</em></td>
<td>0 (0) 0.07</td>
</tr>
</tbody>
</table>

Sensitivity Analyses

To assess the parsimony of the models of SARS-CoV-2 reproduction numbers and COVID-19 case growth rates, equation 3 was re-estimated by restricting the parameters of mobility and containment measures other than mask mandates to jointly equal 0. Restricted models yielded a lower mandate ATT for SARS-CoV-2 reproduction numbers (−0.32 units; P=0.01). In line with the original specification, the results do not demonstrate any relationship between mask mandates and growth rates of COVID-19 cases (−1.22 percentage points; P=0.43). In postmandate periods, the restricted specification also reveals no changes regarding incremental decreases in SARS-CoV-2 reproduction numbers, and our findings demonstrate a negative effect of mask mandates on reproduction numbers at 12 days after mandate beginning. The observed incremental effects last >30 days. A single incremental effect of mandates on COVID-19 case growth rates is observed on day 30 after mandate implementation (−1.84 percentage points; P=0.03). A Wald test was used to test the null hypothesis that the parameters of mobility and containment measures other than mask mandates jointly equal 0. The null hypothesis was rejected.
for both models of mask mandate effects on reproduction numbers (2.67; P=.01) and case growth rates (6.13; P<.001). Therefore, we retained the initial model specification including indicators of mobility and nonpharmaceutical interventions other than mask mandates.

Equation 3 was also re-estimated with counterfactual linear trends to evaluate the robustness of the initial results. Compared with the baseline specification, the observed mandate ATTs are lower for self-reported face mask use (8.66 percentage points; P=.01) and remain unchanged for SARS-CoV-2 reproduction numbers (−0.31 units; P=.01). As in the original specification, no ATT is found for the growth rates of COVID-19–related cases (−0.97 percentage points; P=.56). The incremental increases in self-reported mask use are observed on day 6 after mandate implementation. In addition, the incremental decreases in COVID-19 reproduction numbers are demonstrated on day 12. The incremental effects persist for >30 days for both models. Both results are consistent with the original specification. Significant incremental effects of mandates on COVID-19 case growth rates are found between days 26 and 30 (−1.76 percentage points; P=.04 and −2.13 percentage points; P=.02) after the mandates take effect.

As a final sensitivity check, we have considered that the selection of mandate policies must be strictly exogenous for our specification to yield unbiased estimates of mandate effects. However, the possibility that countries enacted mask mandates depending on the course of the pandemic cannot be ruled out. To rule out bias, the mandate indicators DAt were regressed on SARS-CoV-2 reproduction numbers and growth rates of COVID-19 cases, controlling for country-specific fixed effects. No significant effects were observed. In conclusion, the results of the sensitivity analyses confirm the baseline estimation results. An overview of all results from the sensitivity analyses mentioned in the preceding section is presented in Multimedia Appendix 1 [17,19,21,22,26].

**Effects of Face Mask Recommendations**

The mask mandates evaluated required the populations of the countries under study to wear masks. However, in 25 (24.5%) of the 102 countries represented in the data underlying our analysis, recommendations for voluntary mask use were issued first. Therefore, the final step was to investigate whether these recommendations might have been sufficient to prevent COVID-19 transmission. Accordingly, the baseline specification of equation 3 was re-estimated, with cohorts defined depending on the calendar dates of recommendations becoming effective. Information on self-reported mask use, SARS-CoV-2 reproduction numbers, and COVID-19 case growth rates were available for 60% (15/25) of the countries at the time of face mask recommendations. A total of 11.8% (12/102) of countries were excluded from the evaluation because they demonstrated nonparallel outcome trends in the preliminary analyses. This resulted in a total of 40% (10/25) countries with face mask recommendation being included in the analyses, yielding an analysis sample incorporating 78.4% (80/102) of the countries. The control group consisted of 88% (70/80) of the countries, of which 4% (3/70) of the countries did not implement any face mask policies, whereas 96% (67/70) of the countries enacted mask mandates only.

As presented in Table 3 [17,19,21,22], no significant ATTs of mask recommendations were found for SARS-CoV-2 reproduction numbers and COVID-19 case growth rates (−0.06 units; P=.70 and −2.45 percentage points; P=.59).

However, an ATT of 5.84 percentage points (P<.001) was shown for mask use. In addition, increases in mask use by 11.70 percentage points (P=.04) and 13.60 percentage points (P=0.02) are observed for mask mandates and subgroup mandates, respectively. Our findings also demonstrate mandate effects on reproduction numbers (−0.19 units; P=.04).

Figure 4 [17,19,21,22,26] presents the effects of face mask recommendations in 10 countries. As can be inferred, recommendations had an isolated incremental effect on self-reported use of face masks on days 11 (3.96 percentage points; P=.04), 13 (3.77 percentage points; P=.04), and 25 to 27 (4.20 percentage points; P=.048 and 5.91 percentage points; P=.01) after publication. Single incremental effects of mask recommendations are also observed for reproduction numbers on days 0 (−0.07 units; P=.03) and 1 (−0.07 units; P=.03) and between days 21 (−0.09 units; P=.04) and 28 (−0.11 units; P=.05) after publication. Case growth rates decrease incrementally between days 1 and 4 (−1.60 percentage points; P=.03 and −2.19 percentage points; P=.03) and on day 23 (−2.83 percentage points; P=.05) after recommendations were published. Pretrends are shown for >10 days and on day 9 before the recommendation for self-reported face mask use, on days 9, 5, and 4 for SARS-CoV-2 reproduction numbers, and on days 10 and 9 for growth rates of COVID-19–related cases. These findings support the assumption of no anticipatory events for at least 3 days before face mask recommendations.

<table>
<thead>
<tr>
<th></th>
<th>Self-reported face mask usea,b (observations=13,159)</th>
<th>Reproduction numbera,b (observations=8500)</th>
<th>Case growth in casesa,b (observations=8325)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage points (SE) P value Units (SE) P value</td>
<td>Percentage points (SE) P value</td>
<td></td>
</tr>
<tr>
<td>ATTc</td>
<td>5.84 (1.60) &lt;.001 −0.06 (0.16) .70 −2.45 (4.54) .59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mask mandate</td>
<td>11.70 (3.44) .001 −0.19 (0.09) .04 −2.01 (1.36) .15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subgroup mandate</td>
<td>13.60 (8.34) .002 −0.18 (0.12) .16 −3.18 (2.22) .16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2—mandate</td>
<td>1.16 (3.20) .72 0.05 (0.08) .54 2.22 (1.38) .11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3—mandate</td>
<td>−1.27 (3.88) .74 −0.08 (0.10) .42 0.76 (1.99) .70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School closure</td>
<td><em>d</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ban on events</td>
<td>—</td>
<td>−0.12 (0.05) .02 −0.38 (0.77) .62</td>
<td></td>
</tr>
<tr>
<td>Ban on gatherings</td>
<td>—</td>
<td>−0.02 (0.05) .70 −0.70 (1.02) .50</td>
<td></td>
</tr>
<tr>
<td>Curfew</td>
<td>—</td>
<td>−0.09 (0.04) .02 −0.81 (0.66) .23</td>
<td></td>
</tr>
<tr>
<td>International travel restrictions</td>
<td>—</td>
<td>−0.09 (0.06) .15 −1.58 (1.41) .27</td>
<td></td>
</tr>
<tr>
<td>Protection of older adults</td>
<td>—</td>
<td>0.04 (0.05) .42 −0.78 (1.50) .61</td>
<td></td>
</tr>
<tr>
<td>Mobility</td>
<td>—</td>
<td>−0.01 (0.01) .26 0.37 (0.998) &lt;.001</td>
<td></td>
</tr>
<tr>
<td>Tests per 100,000 inhabitants</td>
<td><em>—</em></td>
<td>0 (0) .20 0 (0) .02</td>
<td></td>
</tr>
</tbody>
</table>

aAll models have been estimated with country and time fixed effects.
bSE (in parentheses) are clustered at the country level.
cATT: Average Treatment Effect on the Treated. The ATT is obtained by averaging Sun and Abraham [26] interaction-weighted estimates of postmandate period effects.
dThe model was specified with ATT and subgroup as well as level-2 and level-3 mandate effects, only.

Discussion

Principal Findings

According to the results presented in the preceding section, the introduction of face mask mandates was associated with an increase in self-reported face mask use and a decrease in SARS-CoV-2 reproduction numbers. The sustainability of face mask mandates is demonstrated by the fact that the observed effects persist for >30 days after the mandates come into effect.

Our results are robust to different sensitivity analyses. Thus, during our observation period, face mask mandates proved to be an effective nonpharmacological measure to reduce SARS-CoV-2 reproduction numbers.

The effectiveness of mask use has occasionally been questioned in public debate [9]. This is particularly true in light of the inconclusive findings of Jefferson et al [5]. According to the authors, a key problem with the existing RCTs is that they have a wide variation in outcome measures, making it difficult to
compare results. Using real-world data, our study allows a comparison of SARS-CoV-2 reproduction numbers and COVID-19 case growth rates across 51 countries. This increases the external validity of previous findings, which are largely based on RCTs. Furthermore, our results support the findings of Ueki et al [27] and Cheng et al [28]. On the basis of aerosol and droplet transmission models, both papers show that face masks are effective in reducing the risk of transmission in low virus load environments such as public spaces.

In addition, our observations suggest inconclusive evidence on the effect of mask mandates on the growth rates of COVID-19 cases. Although Chernozhukov et al [29] demonstrate that mask mandates have contributed to a reduction in case growth rates at the US state level, we found no differences in COVID-19 cases between mandate and nonmandate countries. The lack of mask mandate effects in our study may be, at least in part, owing to the high infectiousness of the SARS-CoV-2 virus. With such high infectiousness, the implementation of mask mandates may not be sufficient to prevent clusters of COVID-19 cases in the most susceptible populations, such as older adults or those who are immunosuppressed. This may be particularly true for facilities such as older people’s care homes, where most cases occurred. Inadequate compliance with mask mandates or improper use of face masks by infected individuals in contact with these populations is hardly always avoidable. Therefore, the occurrence of such clusters is difficult to prevent. Additional strict containment measures may have been required to protect the most susceptible populations.

For masks to have a preventive effect, they must be worn by the public. Accordingly, Jefferson et al [5] discuss the lack of mask adherence as a possible cause of the inconclusive results of RCTs on the effectiveness of face masks in preventing ARIs. To increase the willingness to wear masks, public health authorities used social media to educate the public during the COVID-19 pandemic [30] or distributed masks free of charge [31]. In addition to these interventions, governments have introduced more drastic initiatives, such as mandatory masks, which penalize people who do not wear masks. Adding to the existing evidence from RCTs, our study shows that the self-reported use of masks is significantly increased because of such regulations. Although this is consistent with a previous study that found high compliance with mandatory mask use [12], our results further indicate that mandates, rather than recommendations, are required to ensure self-reported mask adherence.

**Limitations**

It cannot be completely ruled out that the estimates of self-reported face mask use are biased because of the sampling method used. First, the data on self-reported face mask use are obtained from country-specific samples of Facebook users. The distribution of these samples may not reflect the composition of each country’s population. The survey weights calculated by Meta, which are based on population- and region-specific gender and age distributions, may only partially compensate for such biases. Another possible limitation is that it can be assumed that the course of infection is determined by individual behavior. No microlevel data were available for this study that would have allowed us to test this assumption and accurately estimate behavioral effects on the effectiveness of mask mandates. A third complication is that the available data do not allow to distinguish between the type of mask worn (surgical mask vs N95 mask). Depending on mask type, the effects of mandates on reproduction numbers may differ. Such differences cannot be estimated within the framework of this study.

**Conclusions**

Our results suggest that mask mandates encourage self-reported mask use and reduce SARS-CoV-2 reproduction numbers. They may be a simple measure in a nonpharmacological strategy to control epidemics caused by respiratory-transmissible pathogens. When implementing mask mandates, special care should be taken to ensure compliance with respect to the most susceptible individuals. Given the lack of mandate effects on the growth rates of COVID-19 cases, additional containment measures may be required to ensure the adequate protection of the most susceptible populations.

**Acknowledgments**

The authors acknowledge support from the German Research Foundation (Deutsche Forschungsgemeinschaft) and the Open Access Publication Fund of Charité, Universitätsmedizin Berlin.

**Data Availability**

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

**Authors’ Contributions**

A-FN, MS-A, and MK selected, cleaned, and merged the data sets for the analysis. The statistical models for the analysis were specified by A-FN and implemented by A-FN and MS-A. The literature research was conducted by MK. MK and A-FN visualized the data and results. All authors contributed to the interpretation of the results and to the writing of the manuscript. FB provided the required resources.

**Conflicts of Interest**

None declared.
References


https://publichealth.jmir.org/2024/1/e49307


Abbreviations

ARI: acute respiratory infection
ATT: average treatment effect on the treated
CTIS: COVID-19 Trends and Impact Survey
OxCGR: Oxford COVID-19 Government Response Tracker
RCT: randomized controlled trial

Edited by A Mavragani; submitted 24.05.23; peer-reviewed by T Hirosawa, C Li; comments to author 20.08.23; revised version received 20.09.23; accepted 22.12.23; published 08.03.24.

Please cite as:

©Anatol-Fiete Näher, Matthias Schulte-Althoff, Marvin Kopka, Felix Balzer, Francisco Pozo-Martin. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 08.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Temporal Trends of Asthma Among Children in the Western Pacific Region From 1990 to 2045: Longitudinal Observational Study

Cheng-hao Yang1*, PhD; Xin-yu Li2*, PhD; Jia-jie Lv2*, PhD; Meng-jie Hou2, PhD; Ru-hong Zhang2, PhD; Hong Guo3, PhD; Chu Feng1, PhD
1Department of General Surgery, School of Medicine, Shanghai Putuo People's Hospital, Tongji University, Shanghai, China
2Department of Plastic and Reconstructive Surgery, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
3Department of Gynecology and Obstetrics, Tongji Hospital, Tongji University School of Medicine, Shanghai, China
*these authors contributed equally

Corresponding Author:
Chu Feng, PhD
Department of General Surgery, School of Medicine
Shanghai Putuo People's Hospital
Tongji University
NingJiang Road, Putuo District
Shanghai, 200001
China
Phone: 86 15610586615
Email: prs260@aliyun.com

Abstract

Background: Asthma has become one of the most common chronic conditions worldwide, especially among children. Recent findings show that the prevalence of childhood asthma has increased by 12.6% over the past 30 years, with >262 million people currently affected globally. The reasons for the growing asthma epidemic remain complex and multifactorial.

Objective: This study aims to provide an up-to-date analysis of the changing global and regional asthma prevalence, mortality, disability, and risk factors among children aged <20 years by leveraging the latest data from the Global Burden of Disease Study 2019. Findings from this study can help inform priority areas for intervention to alleviate the rising burden of childhood asthma globally.

Methods: The study used data from the Global Burden of Disease Study 2019, concentrating on children aged 0 to 14 years with asthma. We conducted an in-depth analysis of asthma, including its age-standardized prevalence, incidence, mortality, and disability-adjusted life years (DALYs), across diverse demographics, such as region, age, sex, and sociodemographic index, spanning 1990 to 2019. We also projected the future burden of the disease.

Results: Overall, in the Western Pacific Region, the age-standardized prevalence rate of asthma among children increased slightly, from 3898.4 cases per 100,000 people in 1990 to 3924 per 100,000 in 2019. The age-standardized incidence rate of asthma also increased slightly, from 979.2 to 994.9 per 100,000. In contrast, the age-standardized death rate of asthma decreased from 0.9 to 0.4 per 100,000 and the age-standardized DALY rate decreased from 234.9 to 189.7 per 100,000. At the country level, Japan experienced a considerable decrease in the age-standardized prevalence rate of asthma among children, from 6669.1 per 100,000 in 1990 to 5071.5 per 100,000 in 2019. Regarding DALYs, Japan exhibited a notable reduction, from 300.6 to 207.6 per 100,000. Malaysia also experienced a DALY rate reduction, from 188.4 to 163.3 per 100,000 between 1990 and 2019. We project that the burden of disease in countries other than Japan and the Philippines will remain relatively stable up to 2045.

Conclusions: The study indicates an increase in the prevalence and incidence of pediatric asthma, coupled with a decrease in mortality and DALYs in the Western Pacific Region between 1990 and 2019. These intricate phenomena appear to result from a combination of lifestyle shifts, environmental influences, and barriers to health care access. The findings highlight that nations such as Japan have achieved notable success in managing asthma. Overall, the study identified areas of improvement in view of persistent disease burden, underscoring the need for comprehensive collaborative efforts to mitigate the impact of pediatric asthma throughout the region.
Introduction

Asthma, which affects >14% of children globally, is one of the most common chronic respiratory diseases and poses significant public health challenges [1]. Childhood asthma often starts early in life and persists throughout life, frequently continuing into adulthood [2]. Asthma symptoms manifest as wheezing, chest tightness, and coughing that disrupt daily functioning [3]. This disease leads to impaired quality of life, loss of school days, poor sleep, and increased healthcare use and costs [4]. The global burden and impact of childhood asthma continue to increase. Recent findings from the Global Burden of Disease (GBD) Study 2019 show that over the past 30 years, the prevalence of childhood asthma has increased by 12.6%, with >262 million people currently affected worldwide [5]. The reasons for the growing asthma epidemic remain complex and multifactorial and are linked to rising air pollution, urbanization, tobacco smoke exposure, genetics, overuse of antibiotics, and other modern lifestyle factors [6,7]. Careful analysis of recent global and regional trends in childhood asthma prevalence, morbidity, mortality, and projections can inform priority areas for intervention to alleviate this rising disease burden.

The Western Pacific Region (WPR), 1 of the 6 regions classified by the World Health Organization, comprises approximately one-fourth of the world’s population and faces a substantial pediatric disease burden owing to respiratory conditions. Previous analyses of global data reflect increasing asthma prevalence and mortality rates in the WPR from 1990 to 2015, with high rates among boys [8,9]. Recent WPR data on the burden of chronic respiratory diseases also suggest an increasing prevalence of asthma, although mortality rates may be declining [10,11]. Recent global data on pediatric asthma are limited, with most large-scale studies only presenting analyses up to 2015. Adding the latest available disease estimates from 2016 to 2019 provides a contemporary 30-year perspective that is critical to uncovering current trends. Moreover, forecasting future patterns up to 2045 based on recent data can help anticipate trajectory changes and inform priority setting for pediatric asthma care in the WPR over the next 2 decades.

In this study, we leveraged the extensive global disease data within the GBD Study 2019 to analyze past trends from 1990 to 2019 and project estimates up to 2045 for key childhood asthma outcomes in the WPR. The findings can highlight success and gaps in managing this disease, whereas the future projections may guide strategic priorities for improving diagnosis, treatment access, and risk factor reduction to curb the rising prevalence of pediatric asthma across the region.

Methods

Data Source and Definitions

Our study used data from the GBD Study 2019 [12]. This extensive data set encompasses information about 369 diseases and injuries and 87 risk factors across 204 countries and territories, as described in a publication [13]. In the context of the GBD Study 2019, asthma was categorized based on the International Classification of Disease codes. Consistent with the most recent criteria set by the World Bank, countries within the WPR were categorized into 3 economic tiers: high income, upper middle income, and lower middle income. In each economic category, we selected 2 countries for our analysis, intentionally omitting islands or relatively isolated nations owing to their unique health care systems and health technology factors that might skew asthma burden data. For the high-income category, Japan and Singapore were chosen. China and Malaysia represent the upper-middle-income tier, and Vietnam and the Philippines were the choices for the lower-middle-income group.

We focused our analysis on the impact of asthma among children by dividing the patient data into 3 distinct age groups: <5 years, between 5 and 9 years, and between 10 and 14 years. We used linear regression methods to calculate the estimated annual percentage change (EAPC), providing an in-depth view of the changing trends. EAPC is commonly used in epidemiological studies to examine time-related changes in age-standardized rates (ASRs) of diseases. In this model, y represents ln (ASR) and x represents the calendar years. The EAPC, along with its 95% CI, was ascertained through the subsequent linear regression model [14]:

\[ y = \alpha + \beta x + \epsilon \]

\[ \text{EAPC} = 100 \times (\exp[\beta] - 1) \]

To understand the ASR trend, we examined the EAPC and its 95% CI. An upward trend in the ASR is inferred when both the EAPC and the lower bound of the 95% CI are positive. In contrast, a downward trend is indicated when both the EAPC and the upper bound of the 95% CI are negative [14].

Age-Period Cohort Analysis

In our study, we implemented the age-period cohort (APC) model, a highly esteemed statistical framework prevalent in health and social science disciplines. This model was instrumental in our analysis of disease prevalence across various time frames, with a particular emphasis on age, period, and birth cohort [15]. The APC model’s widespread use is especially notable in the study of noncommunicable diseases within the field of epidemiology. In this context, the age effect denotes the varying risk levels across different age groups. The period effect reflects the variations in disease incidence over time, affecting all age groups. Finally, the cohort effect pertains to the shifts in ratios among groups born in the same year, thereby affecting successive age groups over different periods. This nuanced approach enables a comprehensive understanding of disease dynamics across demographic, sociological, and epidemiological perspectives. Our approach involved using asthma data and applying the APC model to understand patterns and trends. To ensure precise analysis, we aligned 5-year age intervals with 5-year periods, spanning from 1990 to 2019.

The APC model provides a refined method for examining disease burden, revealing both broad temporal trends and age-specific nuances in the disease data. The model identifies the general temporal trend or “net drift” as the annual percentage change in the disease burden. This reflects the combined impact of time progression and successive birth cohorts. In contrast, the “local drift” or age-specific temporal trend is the annual percentage change within each age group.

Even a small drift value, when expressed as a yearly percentage change, can indicate significant changes in disease prevalence over a 30-year period. We used a Wald chi-square test to evaluate the statistical significance of these trends. In the APC framework, age effects are observed through age-specific rates over different birth cohorts, adjusted for period influences. Period and cohort effects are shown as relative risks of the disease burden, comparing age-specific rates across different periods or cohorts with a selected reference point. This reference choice is arbitrary and does not alter the interpretability of the findings.

Measurement of Health Inequalities

Total disability-adjusted life years (DALYs) and their ASRs were collected for assessing inequality. The analysis was conducted using the slope index of inequality and the concentration index. These are established measures for assessing absolute and relative gradient inequalities, respectively [16]. The slope index of inequality was calculated by performing a regression analysis on country-level, age-standardized, years of life lost rates owing to asthma among children against an income-based, social position scale. This scale was defined using the midpoint of cumulative population intervals, ranked by per capita gross domestic product. To address variations in the data and the diminishing marginal utility reflected in the data, we used a weighted regression model and applied logarithmic transformation to the relative social position values.

The health inequality concentration index was determined by plotting a Lorenz concentration curve. This curve represented the cumulative relative distribution of the population, ordered by income, against the years of life lost owing to asthma. We then quantified the inequality by numerically integrating the area beneath this curve. When the Lorenz curve is positioned above the line of equality, it indicates that the health burden predominantly affects low-income countries, a scenario reflected by a negative concentration index. It is proposed that an absolute value within the range of 0.2 to 0.3 signifies a considerably high degree of relative inequality.

This approach provided a nuanced understanding of the disparity in asthma burden among children across different income levels in the WPR [17]. The APC analysis was performed with the apc_ie package in Stata (version 17.0; StataCorp).

Future Burden

To forecast the burden from 1990 to 2045, we used a log-linear APC model, which is adept at tempering exponential growth and confining linear trend extrapolation [18]. This makes it particularly effective for aligning with recent trends. The implementation of this model was conducted using the Nordpred package in R (R Foundation for Statistical Computing) [19], which has demonstrated effective empirical performance in forecasting future trends [20]. This method facilitates a more controlled and realistic projection of future trends, using past and present data.

Ethical Considerations

The original GBD study obtained informed consent from the study participants or was granted exemptions by the institutional review board of the University of Washington. As this was a secondary analysis of existing data, no additional human participant research ethics review or informed consent was required. The data used in our study were thoroughly anonymized and deidentified to safeguard participant privacy and confidentiality. Furthermore, our analysis adheres to the established Guidelines on Accurate and Transparent Health Estimate Reporting [21-23]. All participants in this study consented to publication of the results.

Results

Temporal Trends in Pediatric Asthma Prevalence, Incidence, Deaths, and DALYs in the WPR

Overall, in the WPR, the age-standardized prevalence rate (ASPR) of asthma among children increased slightly from 3898.4 cases per 100,000 people in 1990 to 3924 per 100,000 in 2019. The age-standardized incidence rate (ASIR) of asthma also increased slightly from 979.2 to 994.9 per 100,000. In contrast, the age-standardized death rate (ASDR) of asthma decreased from 0.9 to 0.4 per 100,000, and the age-standardized DALYs rate decreased from 234.9 to 189.7 per 100,000 (Table 1; Figure 1).

At the country level, Australia had the highest ASPR of asthma in 1990 at 14,769.1 per 100,000, but it decreased to 10,974.9 per 100,000 in 2019. The age-standardized incidence rate (ASIR) of asthma also increased slightly from 979.2 to 994.9 per 100,000. In contrast, the age-standardized death rate (ASDR) of asthma decreased from 0.9 to 0.4 per 100,000, and the age-standardized DALYs rate decreased from 234.9 to 189.7 per 100,000 (Table 1: Figure 1).
Table 1. Temporal trends in pediatric asthma prevalence, incidence, deaths, and disability-adjusted life years (DALYs) in the Western Pacific Region (1990-2019).

<table>
<thead>
<tr>
<th>Category and location</th>
<th>Num_1990a (95% UI)</th>
<th>ASR_1990b (95% UI)</th>
<th>Num_2019d (95% UI)</th>
<th>ASR_2019c (95% UI)</th>
<th>EAPCf (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global</td>
<td>76,571,140.6 (55,413,704.2 to 109,195,913.6)</td>
<td>4365.6 (3159.3 to 6225.7)</td>
<td>81,720,617.8 (58,030,923.8 to 117,037,165.1)</td>
<td>4170 (2961.2 to 5972.1)</td>
<td>−0.04 (−0.28 to 0.2)</td>
</tr>
<tr>
<td>Western Pacific Region</td>
<td>16,959,863.9 (11,743,336 to 24,831,978.1)</td>
<td>3898.4 (2699.4 to 5707.9)</td>
<td>13,071,953.9 (8,866,463.2 to 19,389,884.5)</td>
<td>3924 (2661.6 to 5820.5)</td>
<td>0.13 (−0.38 to 0.64)</td>
</tr>
<tr>
<td>Australia</td>
<td>559,310.5 (481,247.9 to 632,619.3)</td>
<td>14,769.1 (12,707.7 to 16,704.8)</td>
<td>503,878 (360,666.4 to 698,506.1)</td>
<td>10,974.9 (7855.7 to 15,214.1)</td>
<td>−1.49 (−1.67 to −1.31)</td>
</tr>
<tr>
<td>Brunei Darussalam</td>
<td>3975.9 (2738.3 to 5794.5)</td>
<td>4380.9 (3020 to 6390.7)</td>
<td>4259.7 (2934.5 to 6287.9)</td>
<td>4478 (3094.4 to 6610.2)</td>
<td>−0.04 (−0.08 to 0.01)</td>
</tr>
<tr>
<td>Cambodia</td>
<td>132,886.6 (94,406.4 to 184,807.8)</td>
<td>2786 (1979.3 to 3874.6)</td>
<td>167,371.5 (115,385.3 to 247,138.7)</td>
<td>3325.3 (2292.4 to 4910.1)</td>
<td>0.59 (0.4 to 0.79)</td>
</tr>
<tr>
<td>China</td>
<td>10,454,057.7 (6,763,843.2 to 16,155,345.4)</td>
<td>3237.5 (2094.7 to 5003.1)</td>
<td>7,252,060.8 (4,725,473.7 to 11,276,601.5)</td>
<td>3226.3 (2102.3 to 5016.8)</td>
<td>0.18 (−0.73 to 1.09)</td>
</tr>
<tr>
<td>Cook Islands</td>
<td>340.7 (229.8 to 507.2)</td>
<td>5077.6 (3425.2 to 7558.7)</td>
<td>211.4 (143.2 to 316.2)</td>
<td>5053.9 (3423.3 to 7561.2)</td>
<td>0.21 (0.06 to 0.36)</td>
</tr>
<tr>
<td>Lao People’s Democratic Republic</td>
<td>59,551.2 (44,600.5 to 81,501)</td>
<td>3263.1 (2443.9 to 4465.9)</td>
<td>58,595.3 (42,118 to 81,854)</td>
<td>2604.4 (1872 to 3638.2)</td>
<td>−1.04 (−1.15 to −0.92)</td>
</tr>
<tr>
<td>Malaysia</td>
<td>193,359.9 (140,444.7 to 277,139.9)</td>
<td>2939.6 (2135.1 to 4213.2)</td>
<td>290,422.7 (195,058 to 437,283.4)</td>
<td>3774.5 (2535.1 to 5683.1)</td>
<td>1.12 (0.89 to 1.35)</td>
</tr>
<tr>
<td>Marshall Islands</td>
<td>804.7 (574.6 to 1122.8)</td>
<td>3628.4 (2591.1 to 5063)</td>
<td>671.8 (460.3 to 943.3)</td>
<td>3641.8 (2495.2 to 5113.6)</td>
<td>−0.07 (−0.4 to 0.25)</td>
</tr>
<tr>
<td>Micronesia</td>
<td>1668.2 (1208.3 to 2302.7)</td>
<td>3545.9 (2568.3 to 4894.6)</td>
<td>1231.8 (846.6 to 1773.8)</td>
<td>3921.6 (2695.3 to 5647.2)</td>
<td>0.31 (0.06 to 0.57)</td>
</tr>
<tr>
<td>Mongolia</td>
<td>20,810.5 (13,946.9 to 30,760.5)</td>
<td>2317.5 (1553.1 to 3425.5)</td>
<td>25,252.5 (16,220.4 to 38,449.7)</td>
<td>2494.4 (1602.2 to 3798)</td>
<td>0.13 (0.02 to 0.24)</td>
</tr>
<tr>
<td>Nauru</td>
<td>173.2 (119.6 to 246.5)</td>
<td>3897.4 (2692.3 to 5548.8)</td>
<td>165.2 (112.7 to 246.4)</td>
<td>4217.9 (2878.8 to 6292.8)</td>
<td>0.14 (−0.29 to 0.57)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>90,992.3 (61,774.7 to 131,546.3)</td>
<td>11,373.9 (7721.7 to 16,443.1)</td>
<td>82,994.7 (55,753.8 to 118,776.2)</td>
<td>9200.2 (6248.2 to 13,311.1)</td>
<td>−0.91 (−1.06 to −0.76)</td>
</tr>
<tr>
<td>Niue</td>
<td>38.3 (26.6 to 55.6)</td>
<td>4622.7 (3204.8 to 6704.2)</td>
<td>19.2 (13.3 to 27.9)</td>
<td>4755.2 (3304 to 6935.6)</td>
<td>0.42 (0.18 to 0.65)</td>
</tr>
<tr>
<td>Palau</td>
<td>220.1 (153.7 to 317.5)</td>
<td>4643.7 (3242.7 to 6697)</td>
<td>153.1 (105.7 to 221.9)</td>
<td>4544.8 (3138 to 6586)</td>
<td>−0.09 (−0.25 to 0.07)</td>
</tr>
<tr>
<td>Papua New Guinea</td>
<td>90,389.7 (68,468.4 to 117,497.3)</td>
<td>5385.8 (4079.6 to 7000.9)</td>
<td>161,968.8 (123,526.9 to 209,990.9)</td>
<td>4401.1 (3356.1 to 5705.3)</td>
<td>−0.89 (−1.05 to −0.72)</td>
</tr>
<tr>
<td>Philippines</td>
<td>2,337,086.3 (1,695,954.1 to 3,260,644.5)</td>
<td>9192.4 (6670.7 to 12,825.1)</td>
<td>2,485,339.2 (1,817,569.4 to 3,472,416.2)</td>
<td>6987.9 (5110.4 to 9763.2)</td>
<td>−1.03 (−1.17 to −0.9)</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>456,884.3 (336,467.5 to 624,604.2)</td>
<td>3995.3 (2942.3 to 5462)</td>
<td>298,549.8 (200,075.2 to 448,074.1)</td>
<td>4353.2 (2917.3 to 6533.5)</td>
<td>0.8 (0.61 to 0.99)</td>
</tr>
<tr>
<td>Samoa</td>
<td>2419.2 (1684.8 to 3508.9)</td>
<td>3672.6 (2557.7 to 5326.9)</td>
<td>2777.5 (1915.5 to 4117)</td>
<td>3770.3 (2600.2 to 5588.6)</td>
<td>0.1 (−0.11 to 0.31)</td>
</tr>
</tbody>
</table>
## Incidence

<table>
<thead>
<tr>
<th>Category and location</th>
<th>Num_{1990}^a (95% UI)^b</th>
<th>ASR_{1990}^a (95% UI)</th>
<th>Num_{2019}^d (95% UI)</th>
<th>ASR_{2019}^d (95% UI)</th>
<th>EAPC^c (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>31,180.5 (23,196.2 to 42,797.1)</td>
<td>4801.6 (3572 to 6590.4)</td>
<td>35,086.6 (22,982.6 to 51,478.3)</td>
<td>4486.8 (2939 to 6583)</td>
<td>−0.26 (−0.36 to −0.15)</td>
</tr>
<tr>
<td>Solomon Islands</td>
<td>5537.9 (3822 to 7886.2)</td>
<td>3523.5 (2431.8 to 5017.7)</td>
<td>10,113.8 (7048.9 to 14,748.4)</td>
<td>3915.3 (2728.8 to 5709.4)</td>
<td>0.4 (0.06 to 0.73)</td>
</tr>
<tr>
<td>Tonga</td>
<td>1983.5 (1354.7 to 2930.9)</td>
<td>4990.1 (3408.2 to 7373.6)</td>
<td>1709.7 (1132.1 to 2545.2)</td>
<td>4764.5 (3154.9 to 7092.7)</td>
<td>−0.26 (−0.47 to −0.05)</td>
</tr>
<tr>
<td>Tuvalu</td>
<td>118.4 (83.5 to 166.5)</td>
<td>3587.8 (2532 to 5047.3)</td>
<td>155.3 (105.7 to 233.3)</td>
<td>4593.4 (3090.3 to 6821.3)</td>
<td>0.69 (0.35 to 1.02)</td>
</tr>
<tr>
<td>Vanuatu</td>
<td>2310.1 (1618.8 to 3262.5)</td>
<td>3452.8 (2419.6 to 4876.4)</td>
<td>4238.6 (2771.2 to 6182)</td>
<td>3824.1 (2500.2 to 5577.4)</td>
<td>0.17 (−0.3 to 0.63)</td>
</tr>
<tr>
<td>Vietnam</td>
<td>800,414.4 (552,891.5 to 1,200,763.8)</td>
<td>3055.4 (2110.6 to 4583.7)</td>
<td>765,718.1 (512,976.9 to 1,147,161.2)</td>
<td>3591.3 (2405.9 to 5380.3)</td>
<td>0.31 (−0.19 to 0.8)</td>
</tr>
</tbody>
</table>

### Footnotes

- ^a^: Incidence
- ^b^: Number of cases
- ^c^: Annual percentage change
- ^d^: Number of cases (year)
- ^e^: Age-standardized rate (year)
JMIR PUBLIC HEALTH AND SURVEILLANCE

Yang et al

Num_1990a (95% UIb) ASR_1990c (95% UI)

Num_2019d (95% UI)

ASR_2019e (95% UI)

EAPCf (95% CI)

Palau

55.1 (36.1 to 77.1)

1162.9 (761.7 to 1626)

36.6 (24.1 to 51)

1085.4 (715.3 to
1514.9)

−0.25 (−0.36 to −0.14)

Papua New
Guinea

22,420.8 (16,186.6 to
29,807.9)

1335.9 (964.5 to
1776.1)

41,888.7 (30,890.2 to
55,842.8)

1138.1 (839.3 to
1517.2)

−0.75 (−0.87 to −0.63)

Philippines

539,771.3 (381,497.1 to 2123.1 (1500.5 to 2974) 599,984.4 (418,410.1 to 1686.9 (1176.4 to
756,103.8)
845,919.5)
2378.4)

Category and location

−0.83 (−0.99 to −0.66)

Republic of Ko- 102,439.2 (71,056 to
rea
141,887.1)

895.8 (621.4 to 1240.8) 69,002.6 (44,555.3 to
100,627.1)

1006.1 (649.7 to
1467.3)

Samoa

615.6 (417 to 861.7)

934.5 (633.1 to 1308.2) 682.4 (444.4 to 980.7)

926.3 (603.3 to 1331.2) −0.05 (−0.25 to 0.14)

Singapore

6739.1 (4680.5 to
9182.1)

1037.8 (720.8 to 1414)

1087.2 (710.6 to
1566.4)

−0.07 (−0.24 to 0.09)

Solomon Islands

1483.4 (989.2 to 2097)

943.8 (629.4 to 1334.3) 2632.6 (1794.4 to
3780.3)

1019.1 (694.6 to
1463.4)

0.3 (0.04 to 0.56)

Tonga

488 (324.7 to 687.4)

1227.8 (816.9 to
1729.3)

1183.7 (760.3 to
1676.8)

−0.21 (−0.4 to −0.02)

Tuvalu

32.4 (22.3 to 44.9)

981.8 (676.2 to 1359.6) 37.7 (24.9 to 53.4)

1103.2 (727 to 1560.3)

0.26 (−0.05 to 0.56)

Vanuatu

629.3 (438.4 to 879.6)

940.6 (655.3 to 1314.7) 1089.1 (696.2 to
1543.1)

982.6 (628.1 to 1392.2) −0.01 (−0.4−0.39)

Vietnam

208,363.9 (137,308.5 to 795.4 (524.2 to 1150.9) 189,272 (123,981.6 to
301,488.3)
270,239.5)

887.7 (581.5 to 1267.5) 0.31 (−0.08 to 0.7)

27,949.7 (18,502.8 to
35,060.2)

9911.2 (7949.9 to
12,337)

0.5 (0.4 to 0.6)

−3.62 (−3.73 to −3.52)

Western Pacific 4013 (2666.9 to 5029.9) 0.9 (0.6 to 1.2)
Region

1228 (975.7 to 1472.8)

0.4 (0.3 to 0.4)

−2.44 (−2.77 to −2.11)

Australia

15.8 (13.3 to 18.6)

0.4 (0.4 to 0.5)

6.8 (5.3 to 8.7)

0.1 (0.1 to 0.2)

−3.86 (−4.33 to −3.38)

Brunei Darussalam

0.4 (0.3 to 0.6)

0.5 (0.3 to 0.7)

0.1 (0.1 to 0.1)

0.1 (0.1 to 0.2)

−5.46 (−5.9 to −5.01)

Cambodia

165.1 (68.1 to 248.5)

3.5 (1.4 to 5.2)

33.6 (22.8 to 51.5)

0.7 (0.5 to 1)

−5.4 (−5.54 to −5.26)

China

1299 (775.3 to 1681.3)

0.4 (0.2 to 0.5)

52.3 (39.8 to 73.1)

0 (0 to 0)

−9.95 (−10.44 to −9.45)

Cook Islands

0 (0 to 0)

0.5 (0.3 to 0.7)

0 (0 to 0)

0 (0 to 0.1)

−9.27 (−9.87 to −8.66)

Fiji

6.8 (4.9 to 9.6)

2.4 (1.7 to 3.4)

3.9 (2.7 to 5.3)

1.5 (1 to 2)

−2.68 (−3.28 to −2.08)

Japan

88.8 (70.7 to 98.7)

0.4 (0.3 to 0.4)

4.7 (3.3 to 5.9)

0 (0 to 0)

−9.85 (−10.31 to −9.39)

Kiribati

1.6 (1 to 2.3)

5.6 (3.5 to 8)

0.8 (0.5 to 1.4)

2 (1.3 to 3.3)

−3.52 (−3.73 to −3.31)

Lao People’s
248.2 (98.6 to 403.9)
Democratic Republic

13.6 (5.4 to 22.1)

45.3 (27.7 to 68.7)

2 (1.2 to 3.1)

−6.35 (−6.49 to −6.21)

Malaysia

55.9 (37.2 to 77.7)

0.8 (0.6 to 1.2)

10.3 (6.8 to 16.1)

0.1 (0.1 to 0.2)

−6.73 (−7.13 to −6.33)

Marshall Islands

0.3 (0.2 to 0.5)

1.3 (0.8 to 2.1)

0.1 (0.1 to 0.2)

0.7 (0.5 to 1.1)

−2.4 (−3.07 to −1.72)

Micronesia

1 (0.6 to 1.4)

2 (1.4 to 3)

0.2 (0 to 0.3)

0.5 (0.1 to 0.8)

−5.14 (−5.3 to −4.97)

Mongolia

5 (1.1 to 12.4)

0.6 (0.1 to 1.4)

0.6 (0.3 to 1.1)

0.1 (0 to 0.1)

−8.01 (−8.48 to −7.54)

Nauru

0.1 (0.1 to 0.2)

2.5 (1.2 to 4.1)

0 (0 to 0.1)

0.9 (0.5 to 1.4)

−3.45 (−4.6 to −2.28)

New Zealand

4.3 (3.7 to 4.9)

0.5 (0.5 to 0.6)

1.6 (1.3 to 2)

0.2 (0.1 to 0.2)

−4.23 (−4.53 to −3.93)

Niue

0 (0 to 0)

1 (0.6 to 1.5)

0 (0 to 0)

0.5 (0.2 to 0.8)

−2.98 (−3.45 to −2.52)

Palau

0 (0 to 0.1)

0.7 (0.3 to 1.4)

0 (0 to 0)

0.2 (0.1 to 0.3)

−3.69 (−4.09 to −3.3)

Papua New
Guinea

86.6 (33.8 to 152)

5.2 (2 to 9.1)

102.4 (48.8 to 181.5)

2.8 (1.3 to 4.9)

−1.84 (−2.1 to −1.57)

8501.5 (5557 to
12,249.4)

424.8 (272.8 to 601.7)

0.68 (0.59 to 0.77)

Deaths
Global

https://publichealth.jmir.org/2024/1/e55327

XSL• FO
RenderX

1.6 (1.1 to 2)

JMIR Public Health Surveill 2024 | vol. 10 | e55327 | p.1501
(page number not for citation purposes)


JMIR PUBLIC HEALTH AND SURVEILLANCE

Yang et al

Num_1990a (95% UIb) ASR_1990c (95% UI)

Num_2019d (95% UI)

1670.4 (1021.5 to
2195.8)

6.6 (4 to 8.6)

900.8 (652.1 to 1102.5) 2.5 (1.8 to 3.1)

−2.35 (−2.71 to −1.98)

Republic of Ko- 59.6 (46.4 to 73.3)
rea

0.5 (0.4 to 0.6)

1.2 (0.7 to 2.2)

0 (0 to 0)

−12.37 (−12.82 to
−11.92)

Samoa

0.9 (0.6 to 1.4)

1.4 (0.9 to 2.1)

0.2 (0.1 to 0.4)

0.3 (0.2 to 0.5)

−4.6 (−4.8 to −4.4)

Singapore

4.5 (3.5 to 6)

0.7 (0.5 to 0.9)

0.3 (0.2 to 0.6)

0 (0 to 0.1)

−10.58 (−11.09 to
−10.06)

Solomon Islands

2.1 (1.3 to 3.1)

1.3 (0.8 to 2)

1.6 (1.1 to 2.3)

0.6 (0.4 to 0.9)

−2.34 (−2.67 to −2.02)

Tonga

0.2 (0.2 to 0.3)

0.6 (0.4 to 0.9)

0.1 (0.1 to 0.1)

0.2 (0.2 to 0.4)

−3.17 (−3.65 to −2.69)

Tuvalu

0.1 (0.1 to 0.3)

4.5 (2.2 to 7.8)

0 (0 to 0)

0.4 (0.3 to 0.7)

−8 (−8.23 to −7.76)

Vanuatu

1.1 (0.7 to 1.8)

1.7 (1 to 2.7)

1.1 (0.7 to 1.7)

1 (0.6 to 1.6)

−1.96 (−2.34 to −1.57)

Vietnam

286.1 (153.9 to 446.1)

1.1 (0.6 to 1.7)

59.3 (37.2 to 88.2)

0.3 (0.2 to 0.4)

−4.25 (−4.47 to −4.02)

5,433,991.4
(3,840,150.8 to
7,609,681.4)

309.8 (218.9 to 433.9)

4,116,909.5
(2,702,283.8 to
6,170,009.6)

210.1 (137.9 to 314.8)

−1.19 (−1.4 to −0.98)

Category and location
Philippines

ASR_2019e (95% UI)

EAPCf (95% CI)

DALYs
Global

Western Pacific 1,021,929.5 (690,976.1 234.9 (158.8 to 346.1)
Region
to 1,505,850.2)

631,850.6 (400,999.9 to 189.7 (120.4 to 294.5)
980,983.6)

−0.46 (−0.92 to 0)

Australia

23,892.2 (15,923.8 to
34,403.6)

630.9 (420.5 to 908.5)

20,906.3 (12,111.7 to
33,982.2)

455.4 (263.8 to 740.2)

−1.58 (−1.77 to −1.39)

Brunei Darussalam

196.8 (125.4 to 303.9)

217.1 (138.3 to 335.2)

180.2 (105.2 to 299.2)

189.5 (110.6 to 314.5)

−0.55 (−0.63 to −0.47)

Cambodia

19,258.5 (10,582.7 to
27,408.2)

403.8 (221.9 to 574.6)

9547.6 (6473.9 to
14,223.4)

189.7 (128.6 to 282.6)

−2.6 (−2.83 to −2.37)

China

531,512.5 (333,439 to
832,409.3)

164.6 (103.3 to 257.8)

298,487.3 (166,283.3 to 132.8 (74 to 225.4)
506,739.3)

−0.57 (−1.53 to 0.39)

Cook Islands

16.3 (10.1 to 25.8)

243.4 (150.8 to 384.3)

8.7 (5 to 14.2)

207.2 (118.9 to 339.3)

−0.41 (−0.58 to −0.23)

Fiji

970.2 (732.1 to 1300.9) 344 (259.5 to 461.2)

660.6 (476 to 893.5)

248.8 (179.3 to 336.5)

−1.83 (−2.24 to −1.42)

Japan

69,344.6 (42,600 to
110,193.4)

300.6 (184.7 to 477.7)

32,395.6 (18,423.8 to
54,912.4)

207.6 (118 to 351.8)

−2.27 (−2.75 to 1.79)

Kiribati

193.6 (135.4 to 259.6)

663.2 (463.8 to 889.3)

127.7 (90.5 to 179)

304.8 (216.1 to 427.2)

−2.81 (−2.95 to −2.67)

Lao People’s
23,364.8 (10,409.3 to
Democratic Re- 36,578.7)
public

1280.3 (570.4 to
2004.3)

6125.7 (4222.4 to
8678.5)

272.3 (187.7 to 385.7)

−5.29 (−5.38 to −5.21)

Malaysia

12,392.2 (8756.2 to
17,637.8)

188.4 (133.1 to 268.1)

12,563.3 (7558.1 to
20,841)

163.3 (98.2 to 270.9)

−0.37 (−0.6 to −0.15)

Marshall Islands

55.9 (39.4 to 81.7)

252.2 (177.6 to 368.2)

37.3 (25 to 54.7)

202.4 (135.5 to 296.7)

−1 (−1.22 to −0.77)

Micronesia

145.3 (105.1 to 198.1)

308.8 (223.3 to 421)

62 (39 to 95.4)

197.3 (124.1 to 303.7)

−1.8 (−2.09 to −1.52)

Mongolia

1254.3 (740.5 to
2033.9)

139.7 (82.5 to 226.5)

1068.3 (611.3 to
1777.2)

105.5 (60.4 to 175.5)

−1.27 (−1.5 to −1.04)

Nauru

16.1 (10.1 to 23.3)

362.9 (227.6 to 525.3)

9.6 (6.3 to 14.3)

244.9 (161.5 to 365.9)

−1.71 (−2.23 to −1.18)

New Zealand

4006.8 (2424.5 to
6422.7)

500.8 (303.1 to 802.8)

3441.8 (2050.7 to
5584.2)

385.7 (229.8 to 625.8)

−1.1 (−1.25 to −0.95)

Niue

2.2 (1.4 to 3.3)

266.1 (174.6 to 400.4)

0.9 (0.6 to 1.4)

229.6 (142.2 to 350.8)

−0.47 (−0.62 to −0.32)

Palau

11.8 (7.5 to 18)

248.3 (158.2 to 379.6)

6.7 (4 to 10.6)

199.2 (119.1 to 315.2)

−0.63 (−0.81 to −0.44)

Papua New
Guinea

10,851.6 (6147.3 to
16,546.1)

646.6 (366.3 to 985.9)

15,023.5 (9532 to
21,930.5)

408.2 (259 to 595.8)

−1.48 (−1.68 to −1.28)

https://publichealth.jmir.org/2024/1/e55327

XSL• FO
RenderX

JMIR Public Health Surveill 2024 | vol. 10 | e55327 | p.1502
(page number not for citation purposes)


New Zealand had the highest ASIR of asthma in 1990 at 2086.4 per 100,000. By 2019, the incidence in New Zealand decreased to 1785.4 per 100,000 (EAPC = 0.31, 95% CI: −0.08 to 0.7). The countries with the lowest ASIR of asthma in 2019 were the Lao People’s Democratic Republic at 692 per 100,000 and Mongolia at 723.7 per 100,000.

Papua New Guinea had the highest ASDR of asthma in 2019 at 2.8 per 100,000. By 2019, the ASDR of asthma in the Lao People’s Democratic Republic decreased from 13.6 per 100,000 to 2 per 100,000 (EAPC = −6.35, 95% CI: −6.49 to −6.21). In terms of DALYs owing to asthma, there was a significant overall decrease among children in the WPR. The figures decreased from 234.9 per 100,000 in 1990 to 189.7 per 100,000 in 2019, indicating an EAPC of −0.46, with the 95% CI ranging from −0.92 to 0. This trend mirrors the mortality patterns, with Japan exhibiting a notable reduction in pediatric asthma DALYs during this period. In Japan, the rate decreased from 300.6 to 207.6 per 100,000 (EAPC = −2.27, 95% CI: −2.75 to −1.79). Malaysia also experienced a reduction, from 188.4 to 163.3 per 100,000, between 1990 and 2019 (EAPC = −0.37, 95% CI: −0.60 to −0.15), as shown in Table 1 and Figure 1.

<table>
<thead>
<tr>
<th>Category and location</th>
<th>Num_1990&lt;sup&gt;a&lt;/sup&gt; (95% UI)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>ASR_1990&lt;sup&gt;c&lt;/sup&gt; (95% UI)</th>
<th>Num_2019&lt;sup&gt;d&lt;/sup&gt; (95% UI)</th>
<th>ASR_2019&lt;sup&gt;e&lt;/sup&gt; (95% UI)</th>
<th>EAPC&lt;sup&gt;f&lt;/sup&gt; (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philippines</td>
<td>234,992.5 (162,218.2 to 310,470)</td>
<td>924.3 (638.1 to 1221.2)</td>
<td>175,739.6 (129,991.3 to 238,440.4)</td>
<td>494.1 (365.5 to 670.4)</td>
<td>−1.71 (−1.95 to −1.48)</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>23,353.3 (15,712.7 to 34,903.5)</td>
<td>204.2 (137.4 to 305.2)</td>
<td>12,181.8 (6904.1 to 20,592.3)</td>
<td>175.4 (109 to 282.5)</td>
<td>−0.01 (−0.21 to 0.19)</td>
</tr>
<tr>
<td>Samoa</td>
<td>169.9 (117.4 to 241.1)</td>
<td>257.9 (178.3 to 366)</td>
<td>129.2 (80.3 to 208.1)</td>
<td>175.4 (109 to 282.5)</td>
<td>−1.22 (−1.48 to −0.96)</td>
</tr>
<tr>
<td>Singapore</td>
<td>1632.1 (1101.8 to 2404.2)</td>
<td>251.3 (169.7 to 370.2)</td>
<td>1444.6 (824.5 to 2402.1)</td>
<td>184.7 (105.4 to 307.2)</td>
<td>−1.1 (−1.32 to −0.88)</td>
</tr>
<tr>
<td>Solomon Islands</td>
<td>391.6 (269.3 to 554.7)</td>
<td>249.2 (171.3 to 352.9)</td>
<td>540.9 (365.9 to 809.9)</td>
<td>209.4 (141.6 to 313.5)</td>
<td>−0.63 (−0.85 to −0.41)</td>
</tr>
<tr>
<td>Tonga</td>
<td>99.8 (63.7 to 155.8)</td>
<td>251.1 (160.4 to 392)</td>
<td>76.3 (45.5 to 122.2)</td>
<td>212.6 (126.9 to 340.6)</td>
<td>−0.72 (−0.94 to −0.49)</td>
</tr>
<tr>
<td>Tuvalu</td>
<td>17.3 (10.3 to 27.2)</td>
<td>524.6 (311.8 to 825.4)</td>
<td>7.4 (4.6 to 11.7)</td>
<td>216.2 (134.3 to 340.8)</td>
<td>−3.17 (−3.71 to −2.63)</td>
</tr>
<tr>
<td>Vanuatu</td>
<td>184.8 (126 to 264.9)</td>
<td>276.2 (188.3 to 396)</td>
<td>257.7 (174.3 to 386.7)</td>
<td>232.5 (157.3 to 348.9)</td>
<td>−0.81 (−1.06 to −0.55)</td>
</tr>
<tr>
<td>Vietnam</td>
<td>56,279.1 (37,947.7 to 83,399.1)</td>
<td>214.8 (144.9 to 318.4)</td>
<td>35,963.9 (22,611.9 to 57,339.3)</td>
<td>168.7 (106.1 to 268.9)</td>
<td>−0.89 (−1.21 to −0.57)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Num_1990: number in 1990.
<sup>b</sup>UI: uncertainty interval.
<sup>c</sup>ASR_1990: age-standardized rate in 1990.
<sup>d</sup>Num_2019: number in 2019.
<sup>e</sup>ASR_2019: age-standardized rate in 2019.
<sup>f</sup>EAPC: estimated annual percentage change.
Figure 1. Temporal trends in pediatric asthma prevalence, incidence, deaths, and disability-adjusted life years (DALYs) in the Western Pacific Region (1990-2019). ASIR: age-standardized incidence rate; ASPR: age-standardized prevalence rate.

Trends in Asthma Burden Among Children Across the Sociodemographic Index Quintiles in the WPR

Over the 3 decades under study, the WPR experienced significant changes in the ASPR, ASIR, and ASDR for asthma burden among children across all countries categorized by the sociodemographic index (SDI) level (Figure 2). We found that as the SDI increased, the ASPR and ASIR gradually increased, whereas the ASDR gradually decreased.
Temporal Trends in Asthma Burden Among Children Across Age Groups

The annual percentage change in asthma prevalence for each age group, that is, the local trends in prevalence calculated using the APC model, is presented in Figure S1 in Multimedia Appendix 1. In the WPR, asthma prevalence decreased in the 0-to-4 year and 10-to-14 year age groups. The upper middle-income group (China and Malaysia) showed the same trend. Japan and Singapore also showed favorable changes, with Japan showing a gradual downward trend and Singapore maintaining a stable trend, with no significant upward trend. In Vietnam, asthma prevalence demonstrated decreasing trends in the 0-to-4 year and 5-to-9-year age groups and showed a flat trend in the 5-to-9-year and 10-to-14-year age groups. The local trends in prevalence are shown in Multimedia Appendix 2.

Temporal changes in the age distribution of asthma prevalence are illustrated in Figure S2 in Multimedia Appendix 1. The distribution of age around the world and across the WPR has remained relatively stable over the past 30 years.

Age, Period, and Birth Cohort Effects on Asthma Prevalence Across Different Age Groups

The age, period, and birth cohort effects on asthma prevalence derived from the APC model are illustrated in Figure 3, Multimedia Appendix 3, and Multimedia Appendix 4, respectively. Overall, the pattern of age effects was similar across the selected regions or countries, with the lowest risk among those aged between 0 and 4 years and the risk increasing, decreasing, and then increasing with age. In addition, we found that, in all the regions, the prevalence was higher among men than among women. Overall, period effects presented an initially decreasing and then increasing risk of prevalence across different regions or countries, except in Vietnam. The Vietnam region had generally low period risks over the study period, whereas others had more unfavorable period risks most of the time. Compared with individuals in the reference period from 2000 to 2004, the relative period risk for individuals in the 2015 to 2019 period ranged from 1.25 (95% CI 1.19-1.31) in the WPR to 0.84 (95% CI 0.82-0.86) in Japan and 1.47 (95% CI 1.36-1.58) in China (Multimedia Appendix 5). Regarding birth cohort effects, there was an initially decreasing and then increasing risk of prevalence in successive birth cohorts in the WPR (Multimedia Appendix 4). Japan and Singapore (high-income countries) had favorable prevalence improvements in successive birth cohorts, whereas China, Malaysia, and Vietnam had progressive prevalence deteriorations (Multimedia Appendix 4). Compared with individuals born in the reference cohort from 1990 to 1999, the relative cohort risk for individuals born in the 2010 to 2019 cohort ranged from 1.12 (95% CI 1.06-1.20) in the WPR to 1.35 (95% CI 1.22-1.50) in China (Multimedia Appendix 6).
Health Inequality and Future Burden

The slope index of inequality was $-294$ and $-56$ DALYs per 100,000 in 1990 and 2019, respectively, showing a negative association between the age-standardized DALY rate and gross domestic product per capita. This reduction indicates that the inequality in the age-standardized burden of asthma among children between high-income and low-income countries narrowed during this time (Figure 4). In contrast, a relative inequality analysis showed that the concentration index has increased by 0.06 over 30 years, showing disproportionate concentration of the burden among the richer half of the population (Figure 5).

The projected ASPR and ASIR of childhood asthma for specific representative countries suggest an ongoing increase in both the Philippines and Japan. In contrast, the rates in other selected countries are expected to remain comparatively stable (as illustrated in Multimedia Appendices 7 and 8). In addition, DALYs associated with childhood asthma are anticipated to maintain a relatively steady trend in the future, as shown in Multimedia Appendix 9.
Figure 4. Slope index of inequality in pediatric asthma, presented using regression lines (1990 vs 2019). Regression of age-standardized disability-adjusted life year (DALY) rate on the relative rank by per capita gross domestic product. The slope index of inequality, shown as the slope of the regression line, represents the absolute difference in blindness and vision loss burden between countries or territories with the highest and lowest incomes. Dots represent countries or territories, with different sizes representing the population sizes (1990 in blue and 2019 in red). SDI: sociodemographic index.

Figure 5. Concentration index in pediatric asthma, presented using concentration curves (1990 vs 2019). Concentration curves (Lorenz curves) graphed by plotting the cumulative fraction of population ranked by national per capita gross domestic product (x-axis) against the cumulative fraction of pediatric asthma burden ranked by age-standardized disability-adjusted life year (DALY; y-axis). The concentration index, calculated as twice the area between the 45° diagonal line and the Lorenz curve, represents the relative extent to which the pediatric asthma burden is concentrated among the poor (negative value) or the rich (positive value). Dots represent country or territories, with sizes representing population sizes (1990 in blue and 2019 in red). Trendline (in green) of the hollow rhombi demonstrates the trend in concentration index from 1990 to 2019. SDI: sociodemographic index.
Discussion

Principal Findings

Our study used comprehensive data from the GBD Study 2019 to examine the trends in crucial childhood asthma outcomes across countries in the WPR from 1990 to 2019 and to project future patterns up to 2045. The study revealed increases in both the ASPR and ASIR, alongside decreases in mortality rates and DALYs related to childhood asthma in the WPR over the past 3 decades. This pattern persisted across nations with varying income levels, hinting at universal contributing factors. Notably, Japan showed a significant reduction in pediatric asthma DALYs, reflecting effective management strategies. According to our projections, both the Philippines and Japan might experience continued increases in the ASPR and ASIR. In summary, although the WPR is making progress in reducing fatal asthma outcomes, it continues to face increasing instances of the condition. Focusing on risk factors and ensuring equal access to health care could be key in altering these projected trends.

The gradual upward trend in the ASPR likely results from complex interactions among genetic predisposition, lifestyle changes, and environmental exposure. Key factors influencing the increasing pediatric asthma risk across nations in the WPR include increasing air pollution and urbanization, which promotes triggers such as vehicle emissions, industrial fumes, and household air pollution [24,25]; reduced rates of protective elements such as breastfeeding, which bolsters immune maturation [26]; overseuse of antibiotics, linked to microbiome disruption and heightened immune reactions [27]; limited exposure to microbial diversity, particularly in urban versus rural areas [15]; and unhealthy modern lifestyles marked by poor diets, minimal exercise, and increasing obesity [26]. In addition, dietary shifts, sedentary lifestyles, and increasing obesity may contribute [28]. Hereditary factors also increase the susceptibility in Asian populations. Meanwhile, the increase in the ASIR could arise from superior diagnostic capabilities and awareness over time. Hence, although part of the escalating incidence is because of new-onset disease, improved case identification may also play a role.

Our study indicates that from 1990 to 2019, the ASPR and ASIR for pediatric asthma in countries in the WPR have notably increased, whereas the ASDR has decreased. This trend suggests that although the prevalence and incidence of asthma among children are increasing, efforts to reduce mortality after intervention measurement have been successful. These findings are consistent with those of previous studies of long-term asthma patterns. For example, a global data analysis from 1990 to 2015 showed an increase in asthma prevalence in countries in the WPR, despite a decrease in mortality rates [4]. Moreover, in Organisation for Economic Co-operation and Development countries, there has been a surge in asthma prevalence in recent decades, but child mortality rates have decreased [1]. Thus, advancements in asthma management appear to be effective in reducing fatalities among patients who have been diagnosed. However, the continual increase in asthma cases across nations points to the existing limitations in disease prevention.

This study’s per-country data offer additional insights into pediatric asthma patterns in different WPR settings. From 1990 to 2019, Australia exhibited a reduction in the ASPR while still retaining the highest ASPR, signaling scope for further progress. The substantial reduction in childhood asthma mortality and DALYs in Japan is likely owing to its cutting-edge health care system centered on universal coverage, easy access, technology integration, and personalized care [29]. This facilitates early diagnosis and evidence-based treatment. In addition, nationwide asthma control initiatives in Japan, such as the Practical Asthma Management Guideline, offer standardized diagnostic approaches and globally recognized therapies [30]. Malaysia witnessed a marginal reduction in the DALY rate; however, its burden remains higher than those of neighboring countries with comparable income levels. Thus, Malaysia may glean best practices from Japan’s robust asthma care models to enhance outcomes.

The gradual increase in the SDI has been accompanied by a corresponding increase in both the ASPR and ASIR for pediatric asthma over this study’s time frame. The SDI is indicative of socioeconomic development with factors such as income, education, and fertility rates [31,32]. It appears that pediatric populations in the WPR are still witnessing a growing burden of the disease alongside developmental indexes. Implementing targeted prevention strategies to reduce these avoidable risk factors could help in altering this trend for different socioeconomic groups.

Our age-specific analysis in the WPR reveals a decrease in asthma prevalence across various age groups, starting from ≤5 years and extending to older age groups. This trend can be attributed to the natural maturation of the respiratory and immune systems in very young children, with approximately 75% of early-onset asthma cases resolving by midchildhood [33]. However, for those with moderate to severe asthma persisting beyond the age of 10 years, there is an increased risk of impaired lung function later in life [34]. Therefore, although prevalence decreases with age, the subset struggling with ongoing asthma beyond childhood needs adequate control and follow-up to prevent long-term pulmonary deficits.

In terms of temporal trends, the age distributions within the WPR have remained fairly stable, yet we observed fluctuating asthma risks across the decades. Following an initial reduction, there has been a recent increase in risk. This fluctuation suggests that time-related factors significantly influence asthma risk among different pediatric age groups in the region. Environmental factors are likely the primary drivers of these varying period effects. The successive birth cohorts between 2000 and 2009 exhibited escalating asthma risks compared with the preceding generation. Rapidly evolving environmental and lifestyle changes appear to influence disease susceptibility over generations differently. For example, factors such as shifting microbiome composition, increased exposure to antibiotics, and reduced number of older siblings may contribute to risk amplification in more recent cohorts [35]. In contrast, favorable cohort effects in Japan and Singapore reflect the strengths of those health systems. However, overarching regional trends highlight the need for ready-to-implement pediatric asthma strategies to break the intergenerational risk acceleration.
Our inequality assessment revealed that absolute socioeconomic disparities in childhood asthma DALYs have reduced slightly, but relative inequality persists. Low-income nations in the WPR still disproportionately face asthma burden compared with high-income economies. Barriers related to health awareness, low access to quality care, unaffordable therapies, and high out-of-pocket costs may underpin inequities. Cross-regional cooperation and localized health policy initiatives aimed at marginalized pediatric groups are required to tackle these demands.

Recent global data about pediatric asthma are limited, with most large-scale studies only presenting analyses up to 2015. Adding the latest estimates from 2016 to 2019 provides an up-to-date, 30-year perspective that is critical to capturing the current trends and trajectory changes. Forecasting future patterns up to 2045 based on recent data rather than old estimates can help anticipate potential increases or decreases in childhood asthma morbidity. This long-range projection up to 2045 is key for priority setting, resource planning, and targeted interventions by western Pacific health systems aimed at managing pediatric asthma over the next 2 decades. Specifically, updating the data up to 2019 revealed a reversal from the previous decreasing trends in some nations. These emerging increases signal to policy makers that renewed efforts are needed against pediatric asthma amid the possibly worsening patterns. The value of the 2045 projections lies in helping countries pre-empt potential resurgences and proactively formulate control strategies tailored to local risk factors. From a resource allocation perspective, the expected trajectories can direct preventive efforts and capacity building, where the impacts of pediatric asthma are likely to escalate. Overall, augmenting the data scope offers granular insights to strategically mitigate the barriers confronting children with asthma across the western Pacific area based on where the disease burdens are headed.

Limitations
Our study had some limitations. First, it is vital to recognize the inherent constraints of GBD research. The validity and accuracy of our findings depend on the quality of the population and disease data collected, which may lead to discrepancies between our results and the actual conditions across regions. Second, the lack of comprehensive registry systems for recording mortality in many countries is another limitation. This deficiency often leads to the underestimation of actual figures. Finally, despite the multiple methodologies used in GBD studies for calculations, rectifying the disease classification errors, and reclassifying ambiguous codes, the potential inaccuracies within the data cannot be disregarded. These inherent inaccuracies can affect the reliability of our findings.

Conclusions
Our study offers a comprehensive 30-year analysis of pediatric asthma trends in the WPR. It highlights an increase in disease prevalence alongside a decrease in mortality, amid moderate advancements in development. These trends reflect a dynamic interplay of lifestyle changes, generational transitions, obstacles in accessing health care, and demographic challenges. Addressing these issues requires focused strategies for enhancing awareness, improving prevention, ensuring equitable diagnosis and treatment, and minimizing risks. Ultimately, fostering overall population health and mitigating the risk factors are key to altering the projected trend shifts. Insights from the regional achievements in effective care models can guide future strategies.

Data Availability
The data sets generated during and analyzed during this study are available from the Global Health Data Exchange query tool on the web [12].

Authors’ Contributions
RHZ and HG designed the study. CHY, XYL, MJH, RHZ, JJL, and HG drafted and revised the paper. JJL and CF collected and cleaned the data. JJL checked the accuracy of the data and prepared the tables. CHY, XYL, and CF performed the statistical analysis. All authors reviewed the paper. CHY, XYL, and JJL are credited as co–first authors for their significant contributions, whereas RHZ, CF, and HG are recognized as co–corresponding authors.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Local drift and age distribution of prevalence from 1990 to 2019 for pediatric asthma. Local drift of prevalence from 1990 to 2019 for pediatric asthma. The dots and shaded areas represent the local drift (ie, annual percentage change of age-specific prevalence; percentage per year) and their corresponding 95% CIs. Temporal changes in age distribution of pediatric asthma–related disability-adjusted life years from 1990 to 2019.

[PDF File (Adobe PDF File), 5503 KB - publichealth_v10i1e55327_app1.pdf ]

Multimedia Appendix 2
Local drift of pediatric asthma prevalence in the Western Pacific Region (1990-2019).

[DOCX File, 25 KB - publichealth_v10i1e55327_app2.docx ]
Multimedia Appendix 3
Period effects on disability-adjusted life years for pediatric asthma. This figure illustrates the period relative risk, calculated as the ratio of age-specific rates from the 1990 to 1994 period to the 2015 to 2019 period, with the reference period set at 2000 to 2004.

Multimedia Appendix 4
Birth cohort effects on disability-adjusted life years for pediatric asthma. This figure illustrates the cohort relative risk, calculated as the ratio of age-specific rates from the 1975 to 1984 birth cohort to the 2010 to 2019 cohort, with the reference cohort set at 1990 to 1999.

Multimedia Appendix 5
Period cohort effects of pediatric asthma prevalence in the Western Pacific Region across different age groups (1990-2019).

Multimedia Appendix 6
Birth cohort effects of pediatric asthma prevalence in the Western Pacific Region (1990-2019).

Multimedia Appendix 7
Projected age-standardized prevalence rates of childhood asthma in selected representative countries from 1990 to 2045. Age-standardized prevalence rate shows ongoing increase in both the Philippines and Japan over the period.

Multimedia Appendix 8
Projected age-standardized incidence rates of childhood asthma in selected representative countries from 1990 to 2045. Age-standardized incidence rate demonstrates ongoing increase in both the Philippines and Japan over the period.

Multimedia Appendix 9
Projected disability-adjusted life years associated with childhood asthma in selected representative countries from 1990 to 2045.

References
8. GBD 2015 Chronic Respiratory Disease Collaborators. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic


- https://publichealth.jmir.org/2024/1/e55327

JMIR Public Health Surveill 2024 | vol. 10 | e55327 | p.1511
(page number not for citation purposes)


Abbreviations

APC: age-period cohort
ASDR: age-standardized death rate
ASIR: age-standardized incidence rate
ASPR: age-standardized prevalence rate
ASR: age-standardized rate
DALLY: disability-adjusted life year
EAPC: estimated annual percentage change
GBD: Global Burden of Disease
SDI: sociodemographic index
WPR: Western Pacific Region

Edited by A Mavragani, T Sanchez; submitted 09.12.23; peer-reviewed by XT Yang, ZL Guo; comments to author 16.12.23; revised version received 19.12.23; accepted 23.01.24; published 14.03.24.

Please cite as:
Yang CH, Li XY, Lv JJ, Hou MJ, Zhang RH, Guo H, Feng C
Temporal Trends of Asthma Among Children in the Western Pacific Region From 1990 to 2045: Longitudinal Observational Study
JMIR Public Health Surveill 2024;10:e55327
URL: https://publichealth.jmir.org/2024/1/e55327
doi:10.2196/55327
PMID:36483459

©Cheng-hao Yang, Xin-yu Li, Jia-jie Lv, Meng-jie Hou, Ru-hong Zhang, Hong Guo, Chu Feng. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 14.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Investigating Nonspecific Effects of the Live-Attenuated Japanese Encephalitis Vaccine on Lower Respiratory Tract Infections in Children Aged 25-35 Months: Retrospective Cohort Study

Siyi Zhan¹, MPH; Hongbo Lin², MD; Yingying Yang¹, MPH; Tao Chen¹, MPH; Sheng Mao¹, MPH; Chuanxi Fu¹, PhD

¹The Institute of Infectious Disease and Vaccine, School of Public Health, Zhejiang Chinese Medical University, Hangzhou, China
²Center for Disease Control and Prevention of Yinzhou District, Ningbo, China

Corresponding Author:
Chuanxi Fu, PhD
The Institute of Infectious Disease and Vaccine
School of Public Health
Zhejiang Chinese Medical University
548 Binwen Road
Binjiang District
Hangzhou, 310053
China
Phone: 86 13675863596
Email: fuchuanxi@zcmu.edu.cn

Abstract

Background: Live attenuated vaccines may be used to prevent nontargeted diseases such as lower respiratory tract infections (LRTIs) due to their nonspecific effects (NSEs).

Objective: We aimed to analyze the NSEs of the Japanese encephalitis vaccine on pediatric LRTIs in children aged 25 months to 35 months.

Methods: A retrospective cohort study was conducted by using a population-based electronic health record database in Zhejiang, China. Enrolled participants were children born from January 1, 2017, to December 31, 2017, and who were inoculated with the live-attenuated Japanese encephalitis vaccine (JE-L) or inactivated Japanese encephalitis vaccine (JE-I) as the most recent vaccine at 24 months of age. The study was carried out between January 1, 2019, and December 31, 2019. All inpatient and outpatient hospital visits for LRTIs among children aged 25 months to 35 months were recorded. The Andersen-Gill model was used to assess the NSEs of JE-L against LRTIs in children and compared with those of JE-I as the most recent vaccine.

Results: A total of 810 children born in 2017 were enrolled, of whom 585 received JE-L (JE-L cohort) and 225 received JE-I (JE-I cohort) as their last vaccine. The JE-L cohort showed a reduced risk of LRTIs (adjusted hazard ratio [aHR] 0.537, 95% CI 0.416-0.693), including pneumonia (aHR 0.501, 95% CI 0.393-0.638) and acute bronchitis (aHR 0.525, 95% CI 0.396-0.698) at 25 months to 35 months of age. The NSEs provided by JE-L were especially pronounced in female children (aHR 0.305, 95% CI 0.198-0.469) and children without chronic diseases (aHR 0.553, 95% CI 0.420-0.729), without siblings (aHR 0.361, 95% CI 0.255-0.511), with more than 30 inpatient and outpatient hospital visits prior to 24 months of age (aHR 0.163, 95% CI 0.091-0.290), or with 5 to 10 inpatient and outpatient hospital visits due to infectious diseases prior to 24 months old (aHR 0.058, 95% CI 0.017-0.202).

Conclusions: Compared with JE-I, receiving JE-L as the most recent vaccine was associated with lower risk of inpatient and outpatient hospital visits for LRTIs among children aged 25 months to 35 months. The nature of NSEs induced by JE-L should be considered for policymakers and physicians when recommending JE vaccines to those at high risk of infection from the Japanese encephalitis virus.

(JMIR Public Health Surveill 2024;10:e53040) doi:10.2196/53040
KEYWORDS
nonspecific effect of vaccines; Japanese encephalitis vaccine; Anderson Gill model

Introduction

Vaccination is widely recognized as one of the most effective measures to prevent infectious diseases and has been listed as the top priority for public health precautions by governments worldwide. According to estimates provided by the World Health Organization, 2-3 million children younger than 5 years old could be saved through vaccination against target diseases [1].

Recently, several observational studies and randomized controlled trials have suggested that, in addition to their specific effects, vaccines may have nonspecific effects (NSE) against nontargeted diseases [2]. Live vaccines such as the bacillus Calmette-Guérin (BCG) vaccine [3]; measles, mumps, and rubella (MMR) vaccine [4]; and oral polio vaccine (OPV) [5] have been observed to decrease all-cause mortality and nontargeted infectious disease hospitalizations in children, while inactivated vaccines such as the inactivated polio vaccine (IPV) [5] and diphtheria, tetanus, and pertussis (DTaP) vaccine [6] have been associated with a possible increase in all-cause mortality, on average [7,8]. Consequently, it could be extrapolated that the nature of NSEs produced by a vaccine may be associated with the type of vaccine (ie, NSEs may be more likely to be produced by a live vaccine rather than an inactivated vaccine). However, the conclusions of these previous studies are drawn from studies with a high risk of bias [6] and must be treated with caution. A high risk of bias includes selection bias and information bias arising from, for example, misclassification of vaccination status, confounding at baseline, and selective reporting (and nonreporting) of results. Currently, the immunological mechanisms underlying this phenomenon are unclear but may be related to the trained immunity generated by live vaccines, which is defined as the induction of innate immunity to generate immune memory while clearing pathogens by innate immune cells [9].

In addition, the NSEs of a live vaccine are affected by the sequence of vaccinations. Sørup et al [4] reported that receiving the inactivated DTaP-IPV-Haemophilus influenzae type b (Hib) vaccine as the most recent vaccine (the type of the last vaccine administered to the child) could increase the risk of hospital admissions by 62% during the first 11 months to 24 months after vaccination, while the risk decreased by 14% if MMR was the most recent vaccine. In the United States, Bardenheier et al [10] replicated the study by Sørup et al [4] and demonstrated that receiving a live vaccine as the most recent vaccine was associated with a lower risk of hospitalization for nontargeted infectious diseases from 16 months through 24 months of age compared with an inactivated vaccine (hazard ratio [HR] 0.83, 95% CI 0.72-0.94) as well as concurrent receipt compared with inactivated vaccine (HR 0.91, 95% CI 0.72-0.94).

Japanese encephalitis (JE) vaccines have been included in China’s National Immunization Program (NIP) since 2008. According to the vaccination schedule, children need be inoculated with 2 doses of the live-attenuated JE vaccine (JE-L) at 8 months and 24 months of age or with 4 doses of the inactivated JE vaccine (JE-I) at 8 months, 8 months (7-10 days after the previous dose), 24 months, and 72 months of age [11]. In China, all school-aged children who have reached the age at which they should be vaccinated are vaccinated in accordance with the current NIP vaccination procedures. The type of vaccine (attenuated or inactivated) can be chosen by the child's guardians. Otherwise, the local health authority randomly replenishes JE vaccines, and the providers administer one after another. In Hangzhou, for example, failure to receive immunization vaccines affects children’s enrollment in day care and school. In contrast, children in the United States are administered only the JE-I due to safety considerations [12]. Based on the conclusions of previous studies, we hypothesized that JE-L might also have NSEs. Therefore, in this study, we aimed to evaluate whether the risk of inpatient and outpatient hospital visits for a lower respiratory tract infection (LRTI) differed among children who received the JE-L as their most recent vaccination in comparison with the JE-I.

Methods

Data Source and Participants

Yinzhou is an urban district of Ningbo located in the southeast coastal region of China with an estimated area of 812.40 km² and 1.61 million permanent residents in 2020. A previous study [13] conducted a comprehensive surveillance of JE in Ningbo and showed that the dominant mosquito species in Ningbo was Culex tritaeniorhynchus and that neither JE virus nor dengue virus was detected in mosquitoes throughout the year, suggesting that the rate of local mosquito vectors carrying the virus was relatively low. In addition, another study [14,15] found the vaccine coverage against JE for children aged 1 year to 3 years in Yinzhou District in 2015 was 99.66%.

In 2005, the Yinzhou District Center for Disease Control and Prevention (CDC) developed a population-based electronic health record (EHR) database that collects information from hospitals and community health service centers in the region. This database includes general demographic characteristics, health care information, inpatient and outpatient electronic medical records, health insurance information, disease surveillance, vaccination information, management information, and death certificates [16,17]. As of 2015, it contained the health records of 1.19 million people [18].

We performed this analysis using the aforementioned database as our source of data from 2017 to 2019. Children born in Yinzhou between January 1, 2017, and December 31, 2017, who had received the second dose of JE-L or the third dose of JE-I at 24 months of age as their most recent vaccine were enrolled in this study.
Data Collection
We collected information on demographic characteristics, vaccination (type of vaccine, vaccine name, vaccine dose, and vaccination date), inpatient and outpatient visits (visit time, admission time, discharge time), and disease diagnosis for all included participants. Data pertaining to other variables including maternal parity, birth weight, chronic diseases (including malformations of the respiratory system, other conditions associated with respiratory symptoms, neuromuscular disease, congenital diseases of the heart and urinary system, chromosomal abnormalities, and acquired chronic conditions; see Table S1 in Multimedia Appendix 1 for the specific chronic disease types and International Statistical Classification of Diseases, Tenth Revision [ICD-10] codes), and the number of inpatient and outpatient visits for infectious diseases or for all causes (see Table S2 in Multimedia Appendix 1) prior to 24 months of age were also obtained.

Outcomes
From January 1, 2019, to December 31, 2019, we recorded the primary or secondary discharge diagnosis of LRTIs including influenza, pneumonia, pertussis, acute bronchitis, and others encoded under the diseases of the respiratory system by the ICD-10 as the outcomes of interest (see Table S3 in Multimedia Appendix 1).

Design
Children who had received vaccines other than JE-L or JE-I at the age of 25 months to 35 months were excluded from this study to limit the possibility of bias due to other vaccines. Further, participants with any incomplete or missing information were also excluded. The remaining participants were divided into two groups by the type of JE vaccines received at 24 months of age. Children who received JE-L as their most recent vaccine were included in the JE-L cohort, while those who received JE-I as their most recent vaccine were included in the JE-I cohort. Participants in both groups were followed from the date of administration of JE-L or JE-I to 35 months of age, death, or migration.

We calculated the incidence density of LRTIs and constructed a model to estimate the HR between the JE-L cohort and JE-I cohort.

Statistical Analysis
We used mean (SD) or frequencies with constituent ratios to report the distribution of the children’s age, sex, chronic diseases, maternal parity, and other variables. The chi-square test and Fisher exact test were used to compare baseline characteristics between the 2 cohorts.

Since a child could have had recurring hospital visits due to an LRTI, all hospital visits needed to be involved in the analysis. We calculated the incidence density of LRTIs by using the sum of the number of inpatient and outpatient visits at 25 months to 35 months of age and divided this by the number of person-years of observation. At the same time, we assessed the average length of hospital stay by using the total number of days of hospitalization divided by the number of patients.

The Cox regression hazards model is primarily applied to a single outcome. Therefore, we decided to construct an Andersen-Gill model to assess the HRs and 95% CIs. This model is used for recurrent data and allows every participant to be presented only once and not be compared with him or herself [4]. Schoenfeld residuals were evaluated to assume the proportionality of hazards. If violations were detected, the normal Andersen-Gill model was changed to a time-dependent Andersen-Gill model. We used gender, age group, chronic diseases, birth weight, and the number of outpatient visits with or without infectious diseases prior to 24 months of age as potential covariates in the adjusted analyses. In addition, the results were stratified by sex and other variables.

Statistical analyses for this study were performed using SAS 9.4 (SAS Institute Inc). All tests were 2-sided, and P<.05 was considered significant.

Sensitivity Analyses
Given that the JE-L needs to be administered in 2 doses and JE-I needs to be administered in 4 doses, sequential immunization may be undertaken in clinical practice. Meanwhile, studies have shown that the NSEs of live vaccines might be influenced by the concurrent or successive administration of inactivated vaccines. Consequently, to limit the bias caused by receiving 2 types of JE vaccine by a single child, we repeated our analyses by dividing the children into 4 groups according to the immunization course, as follows: (1) children who received JE-L only, (2) children who received JE-I after JE-L, (3) children who received JE-L after JE-I, and (4) children who received JE-I only. We calculated the incidence density of LRTIs in all 4 groups. At the same time, the Andersen-Gill model was constructed to estimate the HR for the cohort that received JE-L only compared with the cohort that received only JE-I. As per previous reports, if JE-L has NSE, the HR should be <1 [2,10].

Ethical Considerations
The protocol for this study was approved by Zhejiang Chinese Medical University Ethics Committee (No. 20200515-1).

Results
Population
The study initially included 8447 children born in Yinzhou in 2017, of whom 1 child lacked vaccination records and 405 had missing weight information. Further, 7075 children who received vaccines other than those for JE at 25 months to 35 months of age were also excluded. Among the remaining children, 156 were excluded because JE was not administered as their most recent vaccine. The JE-L cohort included the 585 remaining children were administered JE-L as the most recent vaccine, and the JE-I cohort included the 225 children who were administered JE-I as the most recent vaccine, for a total of 810 participants (Figure 1).

Of the 810 children included in the final analysis, 53.1% (430/810) were male, 22.5% (182/810) had a birth weight <3000 g, 5.2% (42/810) had chronic diseases, 63.5% (514/810) had a mother with 1 parity, 30% (243/810) had visited the hospital...
more than 15 times prior to 24 months of age, and 6.5% (53/810) had visited the hospital more than 5 times due to infectious diseases prior to 24 months of age. Both the study cohorts were comparable in terms of sex ($P=.81$), birth weight ($P=.25$), number of hospital visits with or without infectious diseases prior to 24 months of age ($P=.38$ and $P=.79$, respectively), and chronic diseases ($P=.41$). Notably, the JE-I cohort had a higher coverage of the 13-valent pneumonia vaccine, enterovirus 71 vaccine, and rotavirus vaccine and were statistically more likely to have no siblings than the JE-L cohort (see Table S4 in Multimedia Appendix 1).

Figure 1. Flowchart of participant inclusion. JE-I: inactivated Japanese encephalitis vaccine; JE-L: live-attenuated Japanese encephalitis vaccine.

**Vaccination Status**

Among the 8041 children with complete demographic and medical records in the database, 8029 children received a JE vaccine, for a coverage of 99.85%. A total of 36.26% (2916/8041) had completed the first dose of JE-I, and 63.59% (5113/8041) had completed the first dose of JE-L. The age at vaccination administration was consistent with the recommended age in the NIP. Of the 810 children included in final analysis, 27.8% (225/810) received JE-I at an average age of 24.08 months, and 72.2% (585/810) received JE-L at an average age of 24.18 months (see Table S5 in Multimedia Appendix 1).

**Inpatient and Outpatient Hospital Visits for LRTIs**

Among the 810 children who received the JE vaccine as their most recent vaccine at 24 months of age, 85 reported LRTIs during the period of 25 months to 35 months of age, with 310 inpatient and outpatient hospital visits in total. The incidence density was 0.383 (95% CI 0.349-0.417) person-years. Of these 85 children, 73 had acute bronchitis, with an incidence density of 0.331 (95% CI 0.299-0.344) person-years; 8 experienced pneumonia, with an incidence density of 0.033 (95% CI 0.022-0.048) person-years; and 4 reported influenza, with an incidence density of 0.019 (95% CI 0.010-0.030) person-years. Due to LRTIs, 33 children were admitted only once, whereas 52 children were admitted ≥2 times between 25 months and 35 months of age, with a maximum of 28 admissions. The onset age was 29.25 (95% CI 28.70-29.80) months in the JE-I cohort and 29.24 (95% CI 28.78-29.69) months in the JE-L cohort, with no statistical difference ($P=.97$).

Of the 85 children with inpatient or outpatient hospital visits due to LRTIs, 5 were hospitalized for 33 days, with the shortest hospital stay being 5 days and the longest being 7 days. There were average lengths of hospitalization of 7 days in the JE-I cohort and 4.5 days in the JE-L cohort (see Table S6 in Multimedia Appendix 1).

**Risk of Inpatient and Outpatient Hospital Visits for LRTIs**

In the JE-I cohort, there were 128 visits attributed to LRTIs during 225 person-years, with an incidence density of 0.568 (95% CI 0.501-0.634) person-years. In the JE-L cohort, there were 182 visits attributed to LRTIs during 585 person-years, with an incidence density of 0.311 (95% CI 0.274-0.350) person-years. Further, we performed a subgroup analysis, which revealed that the incidence density of the JE-I cohort was lower than that of the JE-L cohort (Table 1).
Table 1. Incidence density and hazard ratios (HRs) of inpatient and outpatient hospital visits for lower respiratory tract infections in different cohorts (N=810): live-attenuated Japanese encephalitis vaccine (JE-L; n=585), inactivated Japanese encephalitis vaccine (JE-I; n=225).

<table>
<thead>
<tr>
<th>Characteristics of the cohorts</th>
<th>Incidence density, person-years (95% CI)</th>
<th>Admissions/person-years</th>
<th>Unadjusted HR (95% CI)</th>
<th>P value</th>
<th>Adjusted HRa (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All characteristics</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.568 (0.501-0.634)</td>
<td>128/225</td>
<td>1.000</td>
<td>.001</td>
<td>1.000</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.311 (0.274-0.350)</td>
<td>182/585</td>
<td>0.547 (0.436-0.685)</td>
<td>.537</td>
<td>0.537 (0.416-0.693)</td>
<td>.537</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
<td>.18</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.529 (0.436-0.620)</td>
<td>64/121</td>
<td>1.000</td>
<td>.537</td>
<td>0.764 (0.534-1.093)</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.430 (0.374-0.488)</td>
<td>133/309</td>
<td>0.814 (0.604-1.096)</td>
<td>.547</td>
<td>0.764 (0.534-1.093)</td>
<td>.547</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.615 (0.515-0.709)</td>
<td>64/104</td>
<td>1.000</td>
<td>.547</td>
<td>0.305 (0.198-0.469)</td>
<td>.547</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.177 (0.134-0.228)</td>
<td>49/276</td>
<td>0.288 (0.199-0.419)</td>
<td>.537</td>
<td>0.305 (0.198-0.469)</td>
<td>.537</td>
</tr>
<tr>
<td><strong>Age (months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>25-27</strong></td>
<td></td>
<td></td>
<td></td>
<td>.37</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.071 (0.020-0.173)</td>
<td>4/56</td>
<td>1.000</td>
<td>.537</td>
<td>1.000</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.119 (0.068-0.187)</td>
<td>15/127</td>
<td>1.654 (0.549-4.982)</td>
<td>.537</td>
<td>1.384 (0.345-5.556)</td>
<td>.537</td>
</tr>
<tr>
<td><strong>28-31</strong></td>
<td></td>
<td></td>
<td></td>
<td>.006</td>
<td>.34</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.474 (0.310-0.642)</td>
<td>18/38</td>
<td>1.000</td>
<td>.537</td>
<td>0.536 (0.150-1.908)</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.208 (0.148-0.282)</td>
<td>31/149</td>
<td>0.439 (0.246-0.785)</td>
<td>.537</td>
<td>0.536 (0.150-1.908)</td>
<td>.537</td>
</tr>
<tr>
<td>≥<strong>32</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>1.140 (0.943-1.362)</td>
<td>106/93</td>
<td>1.000</td>
<td>.537</td>
<td>1.000</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.504 (0.442-0.565)</td>
<td>136/270</td>
<td>0.442 (0.343-0.570)</td>
<td>.537</td>
<td>0.454 (0.336-0.612)</td>
<td>.537</td>
</tr>
<tr>
<td><strong>Chronic diseases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>No</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.474 (0.405-0.544)</td>
<td>100/211</td>
<td>1.000</td>
<td>.537</td>
<td>1.000</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.289 (0.252-0.329)</td>
<td>161/557</td>
<td>0.610 (0.475-0.783)</td>
<td>.537</td>
<td>0.553 (0.420-0.729)</td>
<td>.537</td>
</tr>
<tr>
<td><strong>Yes</strong></td>
<td></td>
<td></td>
<td></td>
<td>.001</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>2.000 (1.372-2.759)</td>
<td>28/14</td>
<td>1.000</td>
<td>.537</td>
<td>1.000</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.750 (0.551-0.893)</td>
<td>21/28</td>
<td>0.375 (0.213-0.660)</td>
<td>.537</td>
<td>0.472 (0.175-1.276)</td>
<td>.537</td>
</tr>
<tr>
<td><strong>Maternal parity</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td><strong>1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.567 (0.486-0.646)</td>
<td>89/157</td>
<td>1.000</td>
<td>.537</td>
<td>1.000</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.244 (0.200-0.292)</td>
<td>87/357</td>
<td>0.430 (0.320-0.578)</td>
<td>.537</td>
<td>0.361 (0.255-0.511)</td>
<td>.537</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td>.83</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.574 (0.448-0.693)</td>
<td>39/68</td>
<td>1.000</td>
<td>.537</td>
<td>1.000</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.424 (0.359-0.492)</td>
<td>95/224</td>
<td>0.739 (0.509-1.073)</td>
<td>.537</td>
<td>0.952 (0.603-1.505)</td>
<td>.537</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td></td>
<td></td>
<td></td>
<td>.b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0</td>
<td>0/0</td>
<td>1.000</td>
<td>.537</td>
<td>1.000</td>
<td>.537</td>
</tr>
<tr>
<td>JE-L</td>
<td>0</td>
<td>0/0</td>
<td>—</td>
<td>.537</td>
<td>—</td>
<td>.537</td>
</tr>
</tbody>
</table>

aAndersen-Gill model adjusted for sex, birth weight, age, maternal parity, chronic diseases, number of hospital visits prior to 24 months of age, number of hospital visits prior to 24 months of age due to infectious diseases, and nonimmunization program vaccines administered before 24 months of age.

bNot applicable.
There was no violation detected between the 2 cohorts, as evaluated via Schoenfeld residuals to test the assumption of a proportional hazard. Therefore, we constructed a normal Andersen-Gill model to calculate the HR. The results showed the JE-L cohort had a lower risk of LRTIs than the JE-I cohort (adjusted HR [aHR] 0.537, 95% CI 0.416-0.693) in the adjusted analyses. Likewise, there were statistical differences within the subgroups (sex, age, chronic diseases, maternal parity), with a lower risk of LRTIs for female children (aHR 0.305, 95% CI 0.198-0.469), children aged 32 months to 35 months (aHR 0.454, 95% CI 0.336-0.612), children without chronic diseases (aHR 0.553, 95% CI 0.420-0.729), and children without siblings (aHR 0.361, 95% CI 0.255-0.511; Table 1).

### Types of LRTIs

In the adjusted analyses, the JE-L cohort had a lower risk of pneumonia (aHR 0.501, 95% CI 0.393-0.638) and acute bronchitis (aHR 0.525, 95% CI 0.396-0.698) than the JE-I cohort (Table 2).

#### Table 2. Incidence density and hazard ratios (HRs) of inpatient and outpatient hospital visits for different types of lower respiratory tract infection (LRTIs) in the cohorts: live-attenuated Japanese encephalitis vaccine (JE-L; n=585), inactivated Japanese encephalitis vaccine (JE-I; n=225).

<table>
<thead>
<tr>
<th>Types of LRTIs by cohorts</th>
<th>Incidence density, person-years (95% CI)</th>
<th>Admissions/ person-years</th>
<th>Adjusted HR (^a) (95% CI)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Influenza</strong></td>
<td></td>
<td></td>
<td>(P) value</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.013 (0.003-0.038)</td>
<td>3/225</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>JE-L</td>
<td>0.021 (0.011-0.036)</td>
<td>12/585</td>
<td>0.501 (0.102-2.451)</td>
<td></td>
</tr>
<tr>
<td><strong>Pneumonia</strong></td>
<td></td>
<td></td>
<td>(P) value</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.040 (0.018-0.075)</td>
<td>9/225</td>
<td>1.000</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.031 (0.018-0.048)</td>
<td>18/585</td>
<td>0.501 (0.393-0.638)</td>
<td></td>
</tr>
<tr>
<td><strong>Acute bronchitis</strong></td>
<td></td>
<td></td>
<td>(P) value</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>0.516 (0.448-0.582)</td>
<td>116/225</td>
<td>1.000</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>JE-L</td>
<td>0.260 (0.225-0.297)</td>
<td>152/585</td>
<td>0.525 (0.396-0.698)</td>
<td></td>
</tr>
</tbody>
</table>

\(a\) Andersen-Gill model adjusted for sex, birth weight, age, maternal parity, chronic diseases, number of hospital visits prior to 24 months of age, number of hospital visits prior to 24 months of age due to infectious diseases, and nonimmunization program vaccines administered before 24 months of age.

### Association With Timing

In the adjusted analyses, no association was observed between the time elapsed after the most recent vaccine and the occurrence of LRTIs among children aged 25 months to 35 months (Table 3).

#### Table 3. Incidence density and hazard ratios (HRs) of inpatient or outpatient hospital visits according to the time elapsed since the most recent vaccine in the cohorts: live-attenuated Japanese encephalitis vaccine (JE-L; n=585), inactivated Japanese encephalitis vaccine (JE-I; n=225).

<table>
<thead>
<tr>
<th>Time after the most recent vaccine by cohort (months)</th>
<th>Incidence density, person-years (95% CI)</th>
<th>Admissions/ person-years</th>
<th>Unadjusted HR (95% CI)</th>
<th>(P) value</th>
<th>Adjusted HR (^a) (95% CI)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>0-89</strong></td>
<td></td>
<td></td>
<td>(P) value</td>
<td></td>
<td>(P) value</td>
<td></td>
</tr>
<tr>
<td>JE-I</td>
<td>5.667 (3.743-7.454)</td>
<td>17/3</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JE-L</td>
<td>2.882 (2.215-3.626)</td>
<td>49/17</td>
<td>0.509 (0.293-0.883)</td>
<td>0.208 (0.027-1.581)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>90-179</strong></td>
<td></td>
<td></td>
<td>(P) value</td>
<td>&lt;.001</td>
<td>(P) value</td>
<td>b</td>
</tr>
<tr>
<td>JE-I</td>
<td>4.750 (3.151-6.388)</td>
<td>19/4</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JE-L</td>
<td>1.538 (0.966-2.276)</td>
<td>20/13</td>
<td>0.324 (0.173-0.607)</td>
<td>0.681 (0.279-1.659)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>180-269</strong></td>
<td></td>
<td></td>
<td>(P) value</td>
<td>.09</td>
<td>(P) value</td>
<td>.40</td>
</tr>
<tr>
<td>JE-I</td>
<td>3.714 (2.589-4.952)</td>
<td>26/7</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JE-L</td>
<td>2.450 (1.871-3.106)</td>
<td>49/20</td>
<td>0.659 (0.410-1.061)</td>
<td>0.681 (0.279-1.659)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>≥270</strong></td>
<td></td>
<td></td>
<td>(P) value</td>
<td>.01</td>
<td>(P) value</td>
<td>.03</td>
</tr>
<tr>
<td>JE-I</td>
<td>8.250 (7.238-9.009)</td>
<td>66/8</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JE-L</td>
<td>5.333 (4.401-6.249)</td>
<td>64/12</td>
<td>0.646 (0.451-0.899)</td>
<td>4.295 (1.200-15.375)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(a\) Andersen-Gill model adjusted for sex, birth weight, age, maternal parity, chronic diseases, number of hospital visits prior to 24 months of age, number of hospital visits prior to 24 months of age due to infectious diseases, and nonimmunization program vaccines administered before 24 months of age.

\(b\) Not applicable.
Association With the Number of Hospital Visits

In the adjusted analyses, we found that children with more hospital visits in the JE-L cohort had a lower risk of LRTIs than the children in the JE-I cohort. Prior to 24 months of age, children who visited the hospital more than 30 times (aHR 0.163, 95% CI 0.091-0.290) and those who visited the hospital 5 to 10 times owing to infectious diseases (aHR 0.058, 95% CI 0.017-0.202) had fewer hospital visits due to LRTIs (Table 4).

Table 4. Incidence density and hazard ratios (HRs) of inpatient and outpatient hospital visits for lower respiratory tract infections (LRTIs) according to hospital visits in the cohorts: live-attenuated Japanese encephalitis vaccine (JE-L; n=585), inactivated Japanese encephalitis vaccine (JE-I; n=225).

<table>
<thead>
<tr>
<th>Number of hospital visits prior to 24 months of age by cohort</th>
<th>Incidence density, person-years (95% CI)</th>
<th>Admissions/person-years</th>
<th>Unadjusted HR (95% CI)</th>
<th>P value</th>
<th>Adjusted HRa (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any reason</td>
<td>0</td>
<td>—</td>
<td>0/42</td>
<td>1.000</td>
<td>1.000</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>1-14</td>
<td>0.042 (0.014-0.096)</td>
<td>5/118</td>
<td>—</td>
<td>—</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>15-29</td>
<td>0.196 (0.127-0.282)</td>
<td>22/112</td>
<td>1.000</td>
<td>1.000</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>≥30</td>
<td>0.149 (0.111-0.195)</td>
<td>44/295</td>
<td>0.759 (0.455-1.267)</td>
<td>0.790 (0.459-1.361)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Due to infectious causes</td>
<td>0</td>
<td>0.078 (0.036-0.143)</td>
<td>9/115</td>
<td>1.000</td>
<td>1.000</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>1-4</td>
<td>0.076 (0.048-0.112)</td>
<td>22/291</td>
<td>0.966 (0.445-2.097)</td>
<td>0.939 (0.398-2.217)</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>5-9</td>
<td>0.347 (0.253-0.452)</td>
<td>33/95</td>
<td>1.000</td>
<td>1.000</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>≥10</td>
<td>0.430 (0.368-0.493)</td>
<td>110/256</td>
<td>1.237 (0.838-1.825)</td>
<td>1.053 (0.705-1.573)</td>
<td>.003</td>
</tr>
</tbody>
</table>

aAndersen-Gill model adjusted for sex, birth weight, age, maternal parity, chronic diseases, number of hospital visits prior to 24 months of age, number of hospital visits prior to 24 months of age due to infectious diseases, and nonimmunization program vaccines administered before 24 months of age.

b Not applicable.

Sensitivity Analyses

During the entire course of JE immunization, 212 children received JE-I only, 573 children received JE-L only, 13 children received JE-I after JE-L, and 12 children received JE-L after JE-I. The incidence density of those who received JE-L only was lower than those who received JE-I only (0.316 vs 0.604 person-years), which is similar to the results of the main analysis (see Table S7 in Multimedia Appendix 1). In addition, the association between the vaccination course of JE and the risk of LRTIs was also similar to that of the main analysis. Children who received JE-L only had a lower risk of inpatient and outpatient hospital visits due to LRTIs than those who received JE-I only (aHR 0.512, 95% CI 0.396-0.662), and this effect was especially pronounced in female children (aHR 0.292, 95% CI 0.189-0.449), those aged 32 months to 35 months (aHR 0.434, 95% CI 0.321-0.587), children with no siblings (aHR 0.355, 95% CI 0.251-0.502), children without chronic diseases (aHR 0.524, 95% CI 0.397-0.692), children with >30
hospital visits prior to 24 months of age (aHR 0.163, 95% CI 0.091-0.290), and children with 5 to 10 hospital visits due to infectious diseases prior to 24 months of age (aHR 0.062, 95% CI 0.017-0.218; see Table S8 in Multimedia Appendix 1).

Discussion

Principal Findings

Most previous studies on NSEs of vaccines have been carried out in resource-limited countries, and several have been conducted in high-income countries, such as the United States [10], Denmark [4], and The Netherlands [19]. To the best of our knowledge, this is the first study to explore the NSEs of JE vaccines in China.

This retrospective cohort study was based on information derived from an integrated clinical database to explore the NSEs of JE-L and their relevant factors. Our findings show that receiving JE-L as the most recent vaccine is associated with a lower risk of hospital admission due to LRTIs (including pneumonia and acute bronchitis) among children aged 25 months to 35 months. In addition, this study also illustrated that the NSEs of JE vaccines may be influenced by sex, health status, maternal parity, number of inpatient or outpatient hospital visits prior to 24 months of age, and the absence of siblings. There was no statistical significance between the time elapsed since the last vaccination and the onset of LRTI-related hospital admissions.

We believe that our findings are reliable since the information used in this study was acquired from Yinzhou EHR databases, which have been proved to be authentic and reliable based on our previous herpes zoster study and other studies [16,18,20-23]. In addition, we had access to the complete immunization records of the children included in this study (vaccination information was obtained from the Immunization Administration Registry and the CDC Adverse Event Following Immunization Information System).

Among other studies, vaccines currently shown to have NSEs include BCG, MMR, measles, live polio, and DPT. In China, these vaccines are all NIP vaccines, which can be given to children free of charge. Some studies have reported that the vaccine coverage of NIP vaccines in China is consistently above 80% [24], and in Yinzhou, the coverage rate is even higher (96.22%) [14]. Almost all children were vaccinated. Therefore, the NSEs effect found in this study cannot be attributed to BCG, MMR, and other related vaccines.

Consistent with the findings of previous studies, our study showed that the incidence density of LRTIs at 25 months to 35 months of age among the 810 participants was much higher among male children and those with a mother with more than 1 parity. Unlike other countries, China’s one-child policy, which was in place for nearly 40 years (1979-2015), fundamentally altered the country’s demographic and social structure. The fact that most of our study participants were a single child is also influenced by the change in policy [25]. One study reported that the risk of hospital admissions for children with respiratory syncytial virus was related to maternal parity and gender of the children in the family and that children with siblings (adjusted odds ratio [OR] 1.96) and male children (adjusted OR 1.57) were at higher risk [26]. A meta-analysis also identified male sex (OR 1.23) and having siblings (OR 1.60) as risk factors for developing LRTIs [27]. Thus, having 1 child may protect against secondary LRTIs from siblings. Meanwhile, we also found that the incidence density of LRTIs was higher in children with chronic diseases and those of older age, which is in agreement with the findings by Sørup et al [4].

Several reports have shown that administration of live vaccines such as BCG [3,6,28], MMR [4,19,29], and OPV [5,30,31] as the most recent vaccine may have NSEs that decrease all-cause mortality or the risk of hospital admissions due to nontarget infectious diseases. In their study, Sørup et al [4] found that receiving MMR as the most recent vaccine, compared with DTaP-IPV-Hib, was associated with a 20% (95% CI 16%-24%) lower risk of LRTIs. In addition, Sørup et al [5] also found that OPV could reduce the risk of LRTIs by 15% (95% CI 5%-23%) compared with DTaP-IPV-Hib as the most recent vaccine. Hollm-Delgado et al [32] found that the BCG vaccine was associated with 17% to 35% risk reduction of LRTIs. A systematic review [2] also suggested that live vaccines could lower the risk of developing LRTIs. Furthermore, Bardenheier et al [10] reported that live vaccines could reduce hospitalizations due to nontarget LRTIs by 44% to 64%. Therefore, our results indicating that JE-L has NSEs on LRTIs are consistent with those of the aforementioned reports.

Although the possible biological mechanisms to support our findings have not been identified, trained immunity [9,33] can be assumed to play a major role. It is now believed that trained immunity is the result of the interaction between immunity, metabolism, and epigenetics. Alterations in the metabolism of intrinsic immune cells affect epigenetics, which can further influence metabolic pathways and cytokine production [34]. For example, the epigenetic modifications in the body after BCG vaccination led to an increase of H3K4me3 at the promoter regions of certain genes, producing protection against unrelated pathogens [35]. Further, it can be conjectured that heterologous protective immunity (cross-protection via T cells) might have a minor role in producing the NSEs of vaccines. In addition, compared with JE-I, JE-L has been reported to produce much more interferon-γ spot-forming cells and interleukin-2 spot-forming cells, both of which exert important antiviral effects [36].

Studies have also found that the NSEs of live vaccines are sex-specific to some extent and that the NSEs of BCG [37,38] and the measles vaccine [39,40] favor the female sex, whereas the NSEs of OPV [30,41] favor the male sex in all-cause mortality and hospital admissions. These results are similar to our findings showing that the NSEs of JE-L favors female children. This observation may be related to the fact that female children are more likely to have a stronger and faster innate and adaptive immune response to viruses than male children due to the higher number, activity, and inflammatory immune response of innate immune cells, including monocytes, macrophages, and dendritic cells in female children than male children [42-44]. The other possible reasons might be the hormone-mediated pro-inflammatory effects of low-dose
estradiol and anti-inflammatory effects of testosterone and progesterone in those of the female sex [2,45,46].

Notably, our results did not elicit protective effects of JE-L on those with chronic diseases, which may be related to the limited number of children with chronic diseases included in the study. Regarding the association between the NSEs of JE-L and the number of hospital visits, our study showed that more hospital visits had a greater protective effect of JE-L in children. This was in agreement with the results of the study by Sørup et al [4], which reported that the NSEs of MMR were better in children with a higher frequency of hospital admissions for infectious diseases before the age of 11 months (0 vs 1 vs 2 vs 3 times and above: HR 1.00 vs 0.96 vs 0.92 vs 0.86, respectively) and children who had been admitted before 11 months of age (0 vs 1 times and above: HR 1.00 vs 0.80). Meanwhile, similar to our findings, the study by Sørup et al [4] also reported that the NSEs of MMR are better in children with mothers with 1 parity than with mothers with 2 parities (HR 1.00 vs 1.03).

Moreover, we also performed a sensitivity analysis in our study to ensure the reliability of our results. The sensitivity analysis in this study revealed similar results to the main analysis, which further supported the finding that JE-L has NSEs on LRTIs. Additionally, it verified that the NSEs of JE-L could be diminished by JE-I.

**Limitations**

This study had several limitations. First, the limited number of participants who met the inclusion criteria resulted in a wide confidence interval for the results, which may have reduced the reliability of our study. Second, we cannot exclude some underreporting, which may bias the results toward uncorrelation. Third, due to the limitations of the database, our study failed to include potential confounders that may have influenced the results. For example, studies have demonstrated that parental smoking [4] and socioeconomic status [47] can influence hospital admissions and the highest educational level of the child’s guardian can influence vaccination choices [48]. Consequently, more studies on the NSEs of JE-L based on EHR databases with sufficient variables are needed in the future. Fourth, the low number of cases of LRTIs found in this study may be related to the vaccination of children against pneumonia [14] and influenza [49]. This may mask the NSEs of JE vaccines.

It is noteworthy that most countries recommend JE-I rather than JE-L for those with an immunocompromised status, such as patients with rheumatic diseases [50], HIV, or solid organ transplant [51,52] owing to possible safety concerns (ie, JE-L was contraindicated in immunocompromised individuals in Australia) [53]. However, inoculating healthy children with JE-L is recommended for its stronger protection [54,55] and may be associated with fewer side effects than JE-I [56,57]. According to our data, the nature of NSEs induced by JE-L should be considered by policymakers and physicians when recommending JE vaccines to those at high risk of infection from the JE virus.

Finally, as this study was conducted just prior to the COVID-19 epidemic, it would be valuable to repeat the study after COVID-19 to validate the robustness of the results in further studies, as COVID-19–related isolation and societal changes may have influenced susceptibility to LRTIs in the study cohort.

**Conclusions**

Compared with JE-I, receiving JE-L as the most recent vaccine was associated with a lower risk of hospital visits for LRTIs among children aged 25 months to 35 months.

**Acknowledgments**

We appreciate the participation of all enrollees in this study. We wish to give special thanks to Yexiang Sun from the Yinzhou Center for Disease Prevention and Control, who helped considerably with this project.

**Data Availability**

The data that support the findings of this study are available from the Yinzhou Center for Disease Prevention and Control (CDC). Restrictions apply to the available of these data, which were used under license for this study. Data are available with permission from the Yinzhou CDC.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Supplementary tables.

[DOCX File, 62 KB - publichealth_v10i1e53040_app1.docx ]

**References**


Abbreviations

aHR: adjusted hazard ratio
BCG: bacillus Calmette-Guérin
CDC: Center for Disease Control and Prevention
DTaP: diphtheria, tetanus, and pertussis
EHR: electronic health record
Hib: Haemophilus influenzae type b
HR: hazard ratio
ICD-10: International Statistical Classification of Diseases, Tenth Revision
IPV: inactivated polio vaccine
JE: Japanese encephalitis
JE-I: inactivated Japanese encephalitis vaccine
JE-L: live-attenuated Japanese encephalitis vaccine
LRTI: lower respiratory tract infection
MMR: measles, mumps, and rubella
NIP: National Immunization Program
NSE: nonspecific effects
OPV: oral polio vaccine
OR: odds ratio
|------------|--------------------------------------------------------------------------------------------------|

©Siyi Zhan, Hongbo Lin, Yingying Yang, Tao Chen, Sheng Mao, Chuanxi Fu. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 18.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Association Between Long-Term Exposure to Air Pollution and the Rate of Mortality After Hip Fracture Surgery in Patients Older Than 60 Years: Nationwide Cohort Study in Taiwan

Shu-Han Chuang¹, MD; Yi-Jie Kuo²,³*, MD, PhD; Shu-Wei Huang⁴, PhD; Han-Wei Zhang⁵,⁶,⁷,⁸, PhD; Hsiao-Ching Peng⁵, MS; Yu-Pin Chen²,³, MD, PhD

¹Division of General Practice, Department of Medical Education, Changhua Christian Hospital, Changhua, Taiwan
²Department of Orthopedics, Wan Fang Hospital, Taipei Medical University, Taipei City, Taiwan
³Department of Orthopedics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
⁴Department of Applied Science, National Taitung University, Taitung City, Taitung County, Taiwan
⁵MetaTrial Research Center, Biomedica Corporation, New Taipei, Taiwan
⁶Program for Aging, China Medical University, Taichung, Taiwan
⁷Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan
⁸Department of Electrical and Computer Engineering, Institute of Electrical Control Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
* these authors contributed equally

Corresponding Author: Yu-Pin Chen, MD, PhD
Department of Orthopedics
Wan Fang Hospital
Taipei Medical University
No. 111, Sec. 3, Xinglong Rd, Wenshan District
Taipei City, 116079
Taiwan
Phone: 886 933296411
Email: 99231@w.tmu.edu.tw

Abstract

Background: To enhance postoperative patient survival, particularly in older adults, understanding the predictors of mortality following hip fracture becomes paramount. Air pollution, a prominent global environmental issue, has been linked to heightened morbidity and mortality across a spectrum of diseases. Nevertheless, the precise impact of air pollution on hip fracture outcomes remains elusive.

Objective: This retrospective study aims to comprehensively investigate the profound influence of a decade-long exposure to 12 diverse air pollutants on the risk of post–hip fracture mortality among older Taiwanese patients (older than 60 years). We hypothesized that enduring long-term exposure to air pollution would significantly elevate the 1-year mortality rate following hip fracture surgery.

Methods: From Taiwan’s National Health Insurance Research Database, we obtained the data of patients who underwent hip fracture surgery between July 1, 2003, and December 31, 2013. Using patients’ insurance registration data, we estimated their cumulative exposure levels to sulfur dioxide (SO₂), carbon dioxide (CO₂), carbon monoxide (CO), ozone (O₃), particulate matter having a size of <10 μm (PM₁₀), particulate matter having a size of <2.5 μm (PM₂.₅), nitrogen oxides (NOₓ), nitrogen monoxide (NO), nitrogen dioxide (NO₂), total hydrocarbons (THC), nonmethane hydrocarbons (NMHC), and methane (CH₄). We quantified the dose-response relationship between these air pollutants and the risk of mortality by calculating hazard ratios associated with a 1 SD increase in exposure levels over a decade.

Results: Long-term exposure to SO₂, CO, PM₁₀, PM₂.₅, NOₓ, NO, NO₂, THC, NMHC, and CH₄ demonstrated significant associations with heightened all-cause mortality risk within 1 year post hip fracture surgery among older adults. For older adults, each 1 SD increment in the average exposure levels of SO₂, CO, PM₁₀, PM₂.₅, NOₓ, NO, NO₂, THC, NMHC, and CH₄ corresponded to a substantial escalation in mortality risk, with increments of 14%, 49%, 18%, 12%, 41%, 33%, 38%, 20%, 9%, and 26%,
respectively. We further noted a 35% reduction in the hazard ratio for O\textsubscript{3} exposure suggesting a potential protective effect, along with a trend of potentially protective effects of CO\textsubscript{2}.

Conclusions: This comprehensive nationwide retrospective study, grounded in a population-based approach, demonstrated that long-term exposure to specific air pollutants significantly increased the risk of all-cause mortality within 1 year after hip fracture surgery in older Taiwanese adults. A reduction in the levels of SO\textsubscript{2}, CO, PM\textsubscript{10}, PM\textsubscript{2.5}, NO\textsubscript{X}, NO, NO\textsubscript{2}, THC, NMHC, and CH\textsubscript{4} may reduce the risk of mortality after hip fracture surgery. This study provides robust evidence and highlights the substantial impact of air pollution on the outcomes of hip fractures.

(JMIR Public Health Surveill 2024;10:e46591) doi:10.2196/46591

KEYWORDS

air pollution; hip fracture; mortality; exposure; older adult; environmental hazard; hazard; morbidity

Introduction

Hip fracture, a major concern for older adults, is associated with high mortality and morbidity as well as a high economic burden on the health care system [1,2]. Hip fracture surgery often leads to poor functional outcomes, such as reduced quality of life, substantial care demands, and limited mobility [3-9]. Several studies have reported an increased rate of mortality (12% to 40%) within 1 year after hip fracture in older adults [10-15]. Pneumonia and cardiovascular diseases are the predominant causes of mortality after hip fracture in these individuals [16-19]. Perioperative factors associated with the risk of 1-year mortality after hip fracture surgery include high baseline Charlson Comorbidity Index (CCI) scores, high American Society of Anesthesiologists grades, and prolonged surgery delay [4,20-22]. To adopt stratified care approaches, such as prioritizing geriatric patients with hip fractures at high risk of mortality for intensive care and to reduce the rate of mortality, clinicians must be aware of the relevant prognostic factors [15].

Since the 1990s, air pollution has been recognized as a major global environmental problem, which leads to increased mortality and morbidity rates [23-26]. Exposure to air pollution is associated with the risk of mortality from various conditions, such as cancer [27], critical illness [28], chronic obstructive pulmonary disease [29], COVID-19 [30], and major surgery [31-33]. However, few studies have investigated the effects of air pollution on the outcomes of hip fractures.

A recent meta-analysis revealed that high risks of osteoporosis and hip fracture were positively associated with increased exposure to air pollutants [34]. However, most studies on air pollution have focused only on bone marrow density and hip fracture incidence, but not on further prognosis or outcome. In 2022, Shi et al [35] reported that exposure to particulate matter having a size of <2.5 \mu m (PM\textsubscript{2.5}), particulate matter having a size of <10 \mu m (PM\textsubscript{10}), and nitrogen dioxide (NO\textsubscript{2}) may increase the risk of mortality within 30 days after hip fracture. Despite the limited samples of air pollutants collected and the lack of long-term evaluation, the aforementioned study may be the only one on the effects of air pollution on mortality after hip fracture surgery.

Emerging evidence underscores the need for a more detailed understanding of the relationship between air pollution and post–hip fracture mortality, particularly given the complex interplay of factors involved in the health outcomes of older adults. Beyond the direct respiratory effects, air pollution has been implicated in systemic inflammation [36-42] and oxidative stress [43-52], both of which are central to the pathophysiological processes underlying hip fracture complications and mortality [53,54]. In other words, this complex interplay might go beyond immediate respiratory implications and include the broader landscape of systemic health. Besides, the older population, particularly those with preexisting comorbidities, may be disproportionately vulnerable to the adverse effects of air pollution due to reduced physiological reserves and impaired reparative mechanisms [55]. It is essential to recognize the holistic impact of air pollution, not only as a respiratory hazard but as a systemic disruptor influencing the delicate balance of physiological processes in older adults.

Furthermore, hip fracture recovery is a complex process that involves various mechanisms that can be influenced by systemic factors. Investigations into the specific impact of long-term exposure to a comprehensive array of air pollutants on the 1-year postoperative mortality risk in this population have been limited, necessitating a deeper exploration into the biological underpinnings. As we consider the theoretical framework of hip fracture pathogenesis, we must delve into the ways air pollution may intricately contribute to the vulnerability and subsequent mortality risk in older adults following hip fracture surgery. By expanding our focus beyond the immediate postoperative period, we can unravel the enduring effects of air pollution on recovery trajectories and long-term outcomes in this vulnerable population.

Therefore, we conducted this retrospective study to investigate the effects of decade-long exposure to various air pollutants on the risk of mortality after hip fracture surgery in older Taiwanese patients. We hypothesized that long-term exposure to air pollution would have an impact on the rate of 1-year mortality after hip fracture surgery. This study investigated the effects of decade-long exposure to 12 air pollutants—sulfur dioxide (SO\textsubscript{2}), carbon dioxide (CO\textsubscript{2}), carbon monoxide (CO), ozone (O\textsubscript{3}), PM\textsubscript{10}, PM\textsubscript{2.5}, nitrogen oxides (NO\textsubscript{X}), nitrogen monoxide (NO), NO\textsubscript{2}, total hydrocarbons (THC), nonmethane hydrocarbons (NMHC), and methane (CH\textsubscript{4})—on older Taiwanese adults who underwent hip fracture surgery. This study seeks to address these critical knowledge gaps by using a population-based approach and comprehensive air pollutant exposure data.
shedding light on the intricate relationship between air pollution and post–hip fracture outcomes. This knowledge has the potential to offer evidence that may serve as a foundation for government authorities to develop strategies aimed at mitigating air pollution. Such strategies could significantly alleviate the strain on the health care system.

**Methods**

**Data Source**

In this study, we collected relevant data from Taiwan’s National Health Insurance Research Database (NHIRD), which was launched by the Taiwanese government in 1995. This database contains the comprehensive medical information of approximately 98.29% of the population of Taiwan (approximately 23 million insured individuals) [56]; these data include the patient’s sex, date of birth, employment status, inpatient and outpatient diagnoses, medical procedures, drug use, treatment duration, and medical costs [57]. We obtained the patients’ baseline information from the Longitudinal Health Insurance Database 2000, a subset of the NHIRD, containing the data of 1 million randomly selected patients.

All data generated or analyzed during this study are included in this published paper and its Multimedia Appendix files. More detailed data sets are not publicly available due to restrictions set by the Taiwan Ministry of Health and Welfare regarding the NHIRD in Taiwan. Researchers interested in accessing the data from the NHIRD must obtain approval from the Health and Welfare. For further information on data availability and access, please refer to the “Data Availability” section.

**Study Population**

Only patients with complete available data were included in this study. Patients with missing, inconsistent, or unknown records of baseline information, such as sex and birth year, were excluded from this study. The cohort included patients who underwent hip fracture surgery between January 1, 2000, and December 31, 2012. Additional exclusion criteria were as follows: unreasonable age (0 years old); younger than 60 years at baseline; with the outcome diagnosis prior to the start of the study (to prevent reverse causation bias); history of pathological fracture, open fracture, or major traffic accident before the initiation of the study; whose follow-up start date was the same as the follow-up end date. At last, participants were excluded if their survival date was before July 01, 2003, to ensure that the study population had a minimum of 10 years of exposure to air pollution. This criterion was established because the Environmental Protection Administration (EPA) data for this study was only available starting from July 1, 1993. Therefore, if a participant’s follow-up ended before July 1, 2003, their exposure to air pollution would be less than 10 years. The ICD-9 (International Classification of Diseases, Ninth Revision) codes established in the inclusion and exclusion criteria were listed in Multimedia Appendix 1.

**Exposure Data Collection**

We obtained information on the cumulative daily average levels of the aforementioned 12 air pollutants from 76 monitoring stations maintained by the Taiwan EPA, Executive Yuan. Data were collected for the period between July 1, 1993, and December 31, 2013. The cumulative daily average level of each pollutant was calculated over the 10 years before the end of follow-up. Then, the data were integrated with the patients’ residential postal codes obtained from their insurance registration data. Any change in residence during the assessment period was also considered. Specifically, changes in a patient’s place of residence during the study period were tracked by their information with insurance registration records, ensuring that any shifts in location were accounted for in the study’s data analysis to maintain accuracy and reliability.

**Outcomes and Confounders**

All-cause mortality was the major outcome of interest. We followed up on each patient until we reached the primary end point, which was determined by 2 conditions: death (withdrawal from the National Health Insurance program) or end of the study period (December 31, 2013). The survival duration was calculated in months. Several confounders were identified and adjusted for, such as age, urbanization level, insurance amount, CCI score, hip fracture procedure, comediations, antosteoporosis medication, ambient temperature, season, and lag 0–1. Patient data such as age, insurance amount, CCI score, hip fracture procedure, comediations, and antosteoporosis medication were also obtained from the NHIRD database. Specifically, the insurance amount was evaluated as an average value for the assessment period of air pollution exposure. Comorbidities were defined as those occurring before the survival date. ICD-9 codes or Anatomical Therapeutic Chemical Classification codes for the definition of hip fracture procedure, comediations, and antosteoporosis medication were detailed in Multimedia Appendix 1. Information on the meteorological factor of ambient temperature was collected from the EPA. Data on the urbanization level were recorded according to the patients’ residence at the beginning of the follow-up period in accordance with the classification proposed by Liu et al [58]. Furthermore, the season was defined on the basis of the date. We also included the 2-day moving average of current- and previous-day levels of air pollutants before the primary end point (lag 0–1) as a confounder, considering the largest effect estimate reported in the literature [59-61].

**Statistical Analysis**

To identify the patient characteristics associated with air pollution, the cohort was divided into 3 tertiles on the basis of the level of exposure to each pollutant. The tertiles were compared using the chi-square test or 1-way ANOVA. Post hoc tests were performed to estimate between-tertile differences if significance was indicated. We plotted crude cumulative incidence curves of mortality within 1 year after hip fracture surgery for the 3 tertiles; between-tertile differences were assessed through log-rank tests. Hazard ratios (HRs) for exposure at 1 SD increment for 10 years were calculated using Cox regression models to evaluate the dose-response effects between air pollutants and mortality risk. The regression models were adjusted for the aforementioned confounders. All tests were 2-sided; statistical significance was set at P<.05. All analyses were performed using the MetaTrial Research Platform (Biomedica Corp).
Ethical Considerations

The Research Ethics Committee of Taipei Medical University, Taiwan, approved this study (TMU-JIRB N202203088). Because of the anonymity of patient data in the NHIRD, the requirement of informed consent was waived.

Results

Study Population

After the inclusion of patients with complete available data (n=882,391), those who underwent hip fracture surgery during the study period were identified (n=9286). Patients who met the exclusion criteria (1349, 75, and 473 for the aforementioned 3 criteria, respectively) were further excluded from the analysis. Finally, 7426 patients were included in the analysis. Figure 1 illustrates the selection process.

Figure 1. Flowchart for patient selection. NHI: National Health Insurance. CH\textsubscript{4}: methane; CO: carbon monoxide; CO\textsubscript{2}: carbon dioxide; NMHC: nonmethane hydrocarbons; NO: nitrogen monoxide; NO\textsubscript{2}: nitrogen dioxide; NO\textsubscript{X}: nitrogen oxides; O\textsubscript{3}: ozone; PM\textsubscript{10}: particulate matters having a size of <10 μm; PM\textsubscript{2.5}: particulate matters having a size of <2.5 μm; SO\textsubscript{2}: sulfur dioxide; THC: total hydrocarbons.

Characteristics and Descriptive Results

Table 1 summarizes the characteristics of the included cohort (n=7426). The incidence of mortality was 929 (12.51%). The average age of the cohort was 78.54 years, and 2926 (39.40%) patients were men. The characteristics of patients exposed to each pollutant were assessed by dividing the cohort into 3 tertiles (Multimedia Appendices 2-13). Table 2 presents the mean and distribution values of each pollutant over the 10-year exposure period.
Table 1. Baseline characteristics of the study cohort.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deaths, n (%)</td>
<td>929 (12.51)</td>
</tr>
<tr>
<td>Men, n (%)</td>
<td>2926 (39.40)</td>
</tr>
<tr>
<td>Age (years), n (%)</td>
<td></td>
</tr>
<tr>
<td>60-79</td>
<td>3866 (52.06)</td>
</tr>
<tr>
<td>≥80</td>
<td>3560 (47.94)</td>
</tr>
<tr>
<td>Age (years); mean (SD)</td>
<td>78.54 (8.07)</td>
</tr>
<tr>
<td>Urbanization level, n (%)</td>
<td></td>
</tr>
<tr>
<td>1 (highest)</td>
<td>3272 (44.06)</td>
</tr>
<tr>
<td>2</td>
<td>2765 (37.23)</td>
</tr>
<tr>
<td>3</td>
<td>711 (9.57)</td>
</tr>
<tr>
<td>4 (lowest)</td>
<td>112 (1.51)</td>
</tr>
<tr>
<td>Unknown</td>
<td>566 (7.62)</td>
</tr>
<tr>
<td>Insurance amount (US $), n (%)</td>
<td></td>
</tr>
<tr>
<td>Financially dependent</td>
<td>24 (0.32)</td>
</tr>
<tr>
<td>0.032 to 631.23</td>
<td>3537 (47.63)</td>
</tr>
<tr>
<td>631.26 to 1262.50</td>
<td>2373 (31.96)</td>
</tr>
<tr>
<td>≥1262.53</td>
<td>119 (1.60)</td>
</tr>
<tr>
<td>Unknown</td>
<td>1373 (18.49)</td>
</tr>
<tr>
<td>CCI² score; mean (SD)</td>
<td>4.57 (2.97)</td>
</tr>
<tr>
<td>Hip fracture procedure, n (%)</td>
<td></td>
</tr>
<tr>
<td>Closed reduction of fracture with internal fixation</td>
<td>448 (6.03)</td>
</tr>
<tr>
<td>Open reduction of fracture with internal fixation</td>
<td>3957 (53.29)</td>
</tr>
<tr>
<td>Partial hip replacement</td>
<td>3021 (40.68)</td>
</tr>
<tr>
<td>Comedications, n (%)</td>
<td>6344 (85.43)</td>
</tr>
<tr>
<td>Antiosteoporosis medication, n (%)</td>
<td></td>
</tr>
<tr>
<td>Alendronate</td>
<td>752 (10.13)</td>
</tr>
<tr>
<td>Risedronate</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Ibandronate</td>
<td>11 (0.15)</td>
</tr>
<tr>
<td>Zoledronic</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Denosumab</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Raloxifene</td>
<td>236 (3.18)</td>
</tr>
</tbody>
</table>

²CCI score: Charlson Comorbidity Index.
Table 2. Mean and distribution of air pollutants over the exposure period.

<table>
<thead>
<tr>
<th></th>
<th>SO₂ ¹b</th>
<th>CO₂ ³c</th>
<th>CO ⁴d</th>
<th>O₃ ⁵e</th>
<th>PM₁₀ ⁶f</th>
<th>PM₂.₅ ⁷b</th>
<th>NOₓ ⁸i</th>
<th>NO ⁹j</th>
<th>NO₂ ¹k</th>
<th>THC ¹l</th>
<th>NMHC ²m</th>
<th>CH₄ ³n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>4.14 (1.35)</td>
<td>398.46 (11.50)</td>
<td>0.55 (0.14)</td>
<td>28.08 (2.56)</td>
<td>55.68 (9.78)</td>
<td>33.7 (6.85)</td>
<td>25.43 (8.16)</td>
<td>7.41 (4.41)</td>
<td>18.02 (4.16)</td>
<td>2.31 (0.17)</td>
<td>0.30 (0.11)</td>
<td>2.01 (0.11)</td>
</tr>
<tr>
<td>T₁ ¹o</td>
<td>3.59</td>
<td>393.97</td>
<td>0.47</td>
<td>27.31</td>
<td>50.25</td>
<td>29.98</td>
<td>20.59</td>
<td>4.69</td>
<td>15.87</td>
<td>2.22</td>
<td>0.25</td>
<td>1.96</td>
</tr>
<tr>
<td>T₂ ¹p</td>
<td>4.06</td>
<td>401.51</td>
<td>0.61</td>
<td>28.83</td>
<td>58.45</td>
<td>36.49</td>
<td>28.32</td>
<td>7.76</td>
<td>20.29</td>
<td>2.36</td>
<td>0.32</td>
<td>2.05</td>
</tr>
</tbody>
</table>

¹SO₂: sulfur dioxide.
²ppb: parts per billion.
³CO₂: carbon dioxide.
⁴ppm: parts per million.
⁵CO: carbon monoxide.
⁶O₃: ozone.
⁷PM₁₀: particulate matter <10 μm in size.
⁸PM₂.₅: particulate matter <2.5 μm in size.
⁹NOₓ: nitrogen oxides.
¹₀NO: nitrogen monoxide.
¹¹NO₂: nitrogen dioxide.
¹²THC: total hydrocarbons.
¹³NMHC: nonmethane hydrocarbons.
¹⁴CH₄: methane.
¹⁵T₁: 33.33rd percentile.
¹⁶T₂: 66.66th percentile.

Cumulative Mortality Incidence

Table 3 presents the incidence of mortality within 1 year after hip fracture surgery across the tertiles, as well as the corresponding P value (1-way analysis of variance and post hoc tests). The cumulative incidence curves revealed substantial differences in the direction and strength of the associations between each air pollutant and the incidence of mortality (Figures 2-13).
Table 3. Incidence of mortality within 1 year after hip fracture surgery across tertiles.

<table>
<thead>
<tr>
<th>Pollutants</th>
<th>Tertiles of average daily exposure, n/n (%)</th>
<th>P values</th>
<th>Total, n/n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1 (lowest)</td>
<td>T2</td>
<td>T3 (highest)</td>
</tr>
<tr>
<td>SO$_2$\textsuperscript{a}</td>
<td>229/2391 (9.58)</td>
<td>264/2541 (10.39)</td>
<td>436/2494 (17.48)</td>
</tr>
<tr>
<td>CO$_2$\textsuperscript{b}</td>
<td>252/1271 (19.83)</td>
<td>86/1269 (6.78)</td>
<td>64/1272 (5.03)</td>
</tr>
<tr>
<td>CO\textsuperscript{c}</td>
<td>180/2438 (7.38)</td>
<td>300/2512 (11.94)</td>
<td>449/2476 (18.13)</td>
</tr>
<tr>
<td>O$_3$\textsuperscript{d}</td>
<td>574/2343 (24.50)</td>
<td>178/2607 (6.83)</td>
<td>177/2476 (7.15)</td>
</tr>
<tr>
<td>PM$_{10}$\textsuperscript{e}</td>
<td>274/2475 (11.07)</td>
<td>266/2335 (11.39)</td>
<td>389/2616 (14.87)</td>
</tr>
<tr>
<td>PM$_{2.5}$\textsuperscript{f}</td>
<td>240/2417 (9.93)</td>
<td>245/2417 (10.14)</td>
<td>361/2418 (14.93)</td>
</tr>
<tr>
<td>NO$_X$\textsuperscript{g}</td>
<td>199/2439 (8.16)</td>
<td>329/2511 (13.10)</td>
<td>401/2476 (16.20)</td>
</tr>
<tr>
<td>NO\textsuperscript{h}</td>
<td>215/2409 (8.92)</td>
<td>325/2541 (12.79)</td>
<td>389/2476 (15.71)</td>
</tr>
<tr>
<td>NO$_2$\textsuperscript{i}</td>
<td>197/2475 (7.96)</td>
<td>311/2475 (12.57)</td>
<td>421/2476 (17.00)</td>
</tr>
<tr>
<td>THC\textsuperscript{j}</td>
<td>118/2442 (4.83)</td>
<td>231/2441 (9.46)</td>
<td>573/2442 (23.46)</td>
</tr>
<tr>
<td>NMHC\textsuperscript{k}</td>
<td>215/2442 (8.80)</td>
<td>305/2441 (12.49)</td>
<td>402/2442 (16.46)</td>
</tr>
<tr>
<td>CH$_4$\textsuperscript{l}</td>
<td>100/2442 (4.10)</td>
<td>229/2441 (9.38)</td>
<td>593/2442 (24.28)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}SO$_2$: sulfur dioxide.  
\textsuperscript{b}CO$_2$: carbon dioxide.  
\textsuperscript{c}CO: carbon monoxide.  
\textsuperscript{d}O$_3$: ozone.  
\textsuperscript{e}PM$_{10}$: particulate matters having a size of <10 μm.  
\textsuperscript{f}PM$_{2.5}$: particulate matters having a size of <2.5 μm.  
\textsuperscript{g}NO$_X$: nitrogen oxides.  
\textsuperscript{h}NO: nitrogen monoxide.  
\textsuperscript{i}NO$_2$: nitrogen dioxide.  
\textsuperscript{j}THC: total hydrocarbons.  
\textsuperscript{k}NMHC: nonmethane hydrocarbons.  
\textsuperscript{l}CH$_4$: methane.
Figure 2. Cumulative incidence curves of mortality across tertiles of sulfur dioxide.

![Cumulative incidence curves of mortality across tertiles of sulfur dioxide.](image1)

Concentration, ppb: T1 (< 3.59), T2 (≥ 3.59 and < 4.06), T3 (≥ 4.06)

Figure 3. Cumulative incidence curves of mortality across tertiles of carbon dioxide.

![Cumulative incidence curves of mortality across tertiles of carbon dioxide.](image2)

Concentration, ppm: T1 (< 393.97), T2 (≥ 393.97 and < 401.51), T3 (≥ 401.51)
Figure 4. Cumulative incidence curves of mortality across tertiles of carbon monoxide.

Concentration, ppm: T1 (< 0.47), T2 (≥ 0.47 and < 0.61), T3 (≥ 0.61)

Figure 5. Cumulative incidence curves of mortality across tertiles of ozone.

Concentration, ppb: T1 (< 27.31), T2 (≥ 27.31 and < 28.83), T3 (≥ 28.83)
Figure 6. Cumulative incidence curves of mortality across tertiles of particulate matters having a size of <10 μm.

Figure 7. Cumulative incidence curves of mortality across tertiles of particulate matters having a size of <2.5 μm.
Figure 8. Cumulative incidence curves of mortality across tertiles of nitrogen oxides.

Concentration, ppb: T1 (< 20.59), T2 (≥ 20.59 and < 28.32), T3 (≥ 28.32)

Figure 9. Cumulative incidence curves of mortality across tertiles of nitrogen monoxide.

Concentration, ppb: T1 (< 4.69), T2 (≥ 4.69 and < 7.76), T3 (≥ 7.76)
Figure 10. Cumulative incidence curves of mortality across tertiles of nitrogen dioxide.

Concentration, ppb; T1 (< 15.87), T2 (≥ 15.87 and < 20.29), T3 (≥ 20.29)

Figure 11. Cumulative incidence curves of mortality across tertiles of total hydrocarbons.

Concentration, ppm; T1 (< 2.22), T2 (≥ 2.22 and < 2.36), T3 (≥ 2.36)
Figure 12. Cumulative incidence curves of mortality across tertiles of nonmethane hydrocarbons.

Figure 13. Cumulative incidence curves of mortality across tertiles of methane.
Dose-Response Effects Between Air Pollutants and Mortality Risk

To evaluate the dose-response effects between the levels of air pollutants and the risk of mortality, we used Cox regression models and calculated the HRs for exposure at 1 SD increment for 10 years. Table 4 presents the corresponding HRs. The findings suggest that a 1 SD increment in the average exposure levels of \( \text{SO}_2 \), \( \text{CO} \), \( \text{PM}_{10} \), \( \text{PM}_{2.5} \), \( \text{NO}_x \), \( \text{NO}, \text{NO}_2, \text{THC}, \text{NMHC}, \text{and CH}_4 \) was associated with 14\%, 49\%, 18\%, 12\%, 41\%, 33\%, 38\%, 20\%, 9\%, and 26\% significant increases in mortality risk, respectively. However, a significantly negative association was observed for \( \text{O}_3 \), with a 35\% reduction in mortality risk. \( \text{CO}_2 \) tended to exhibit only a trend but not a statistical significance to reduced mortality risk.

### Table 4. Hazard ratios for mortality associated with long-term exposure to air pollution at 1 SD increment for 10 years.

<table>
<thead>
<tr>
<th>Pollutants</th>
<th>Adjusted(^a) HR(^b) (95% CI)</th>
<th>( P ) values</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{SO}_2 ) (^c) (ppb)</td>
<td>1.14 (1.06-1.23)</td>
<td>&lt;.001</td>
<td>4.14 (1.35)</td>
</tr>
<tr>
<td>( \text{CO}_2 ) (^d) (ppm)</td>
<td>0.88 (0.76-1.02)</td>
<td>.09</td>
<td>398.46 (11.50)</td>
</tr>
<tr>
<td>( \text{CO} ) (^e) (ppm)</td>
<td>1.49 (1.38-1.59)</td>
<td>&lt;.001</td>
<td>0.55 (0.14)</td>
</tr>
<tr>
<td>( \text{O}_3 ) (^h) (ppb)</td>
<td>0.65 (0.60-0.69)</td>
<td>&lt;.001</td>
<td>28.08 (2.56)</td>
</tr>
<tr>
<td>( \text{PM}_{10} ) (^i) (μg/m(^3))</td>
<td>1.18 (1.08-1.29)</td>
<td>&lt;.001</td>
<td>55.68 (9.78)</td>
</tr>
<tr>
<td>( \text{PM}_{2.5} ) (^j) (μg/m(^3))</td>
<td>1.12 (1.01-1.23)</td>
<td>.03</td>
<td>33.7 (6.85)</td>
</tr>
<tr>
<td>( \text{NO}_x ) (^k) (ppb)</td>
<td>1.41 (1.29-1.53)</td>
<td>&lt;.001</td>
<td>25.43 (8.16)</td>
</tr>
<tr>
<td>( \text{NO} ) (^l) (ppb)</td>
<td>1.33 (1.23-1.44)</td>
<td>&lt;.001</td>
<td>7.41 (4.41)</td>
</tr>
<tr>
<td>( \text{NO}_2 ) (^m) (ppb)</td>
<td>1.38 (1.26-1.50)</td>
<td>&lt;.001</td>
<td>18.02 (4.16)</td>
</tr>
<tr>
<td>( \text{THC} ) (^n) (ppm)</td>
<td>1.20 (1.11-1.29)</td>
<td>&lt;.001</td>
<td>2.31 (0.17)</td>
</tr>
<tr>
<td>( \text{NMHC} ) (^o) (ppm)</td>
<td>1.09 (1.01-1.17)</td>
<td>.03</td>
<td>0.30 (0.11)</td>
</tr>
<tr>
<td>( \text{CH}_4 ) (^p) (ppm)</td>
<td>1.26 (1.16-1.36)</td>
<td>&lt;.001</td>
<td>2.01 (0.11)</td>
</tr>
</tbody>
</table>

\(^a\) Cox regression models were adjusted for age, urbanization level, insurance amount, Charlson Comorbidity Index score, hip fracture procedure, comedications, antosteoporosis medication, ambient temperature, season, and lag 0-1.

\(^b\) HR: hazard ratio.

\(^c\) \( \text{SO}_2 \): sulfur dioxide.

\(^d\) ppb: parts per billion.

\(^e\) \( \text{CO}_2 \): carbon dioxide.

\(^f\) ppm: parts per million.

\(^g\) \( \text{CO} \): carbon monoxide.

\(^h\) \( \text{O}_3 \): ozone.

\(^i\) \( \text{PM}_{10} \): particulate matters having a size of <10 μm.

\(^j\) \( \text{PM}_{2.5} \): particulate matters having a size of <2.5 μm.

\(^k\) \( \text{NO}_x \): nitrogen oxides.

\(^l\) \( \text{NO} \): nitrogen monoxide.

\(^m\) \( \text{NO}_2 \): nitrogen dioxide.

\(^n\) THC: total hydrocarbons.

\(^o\) NMHC: nonmethane hydrocarbons.

\(^p\) \( \text{CH}_4 \): methane.

**Discussion**

**Principal Findings**

In this study, long-term exposure to \( \text{SO}_2 \), \( \text{CO} \), \( \text{PM}_{10} \), \( \text{PM}_{2.5} \), \( \text{NO}_x \), \( \text{NO}, \text{NO}_2, \text{THC}, \text{NMHC}, \text{and CH}_4 \) was associated with an increased risk of all-cause mortality within 1 year after hip fracture surgery. In older individuals, a 1 SD increment in the average exposure levels of \( \text{SO}_2 \), \( \text{CO} \), \( \text{PM}_{10} \), \( \text{PM}_{2.5} \), \( \text{NO}_x \), \( \text{NO}, \text{NO}_2, \text{THC}, \text{NMHC}, \text{and CH}_4 \) was associated with 14\%, 49\%, 18\%, 12\%, 41\%, 33\%, 38\%, 20\%, 9\%, and 26\% increases in mortality risk, respectively. By contrast, we found a significant reduction (35\%) in the HR for \( \text{O}_3 \) and a trend of potentially protective effects of \( \text{CO}_2 \).
Comparison With Prior Work

In this study, long-term exposure to SO₂, CO, PM₁₀, PM₂.₅, NOₓ, NO, NO₂, THC, NMHC, and CH₄ was consistently found to be risk factors for mortality within 1 year after hip fracture surgery. Previous studies have found an association between osteoporosis and the levels of exposure to air pollutants, including PM₂.₅, PM₁₀, CO, and NOₓ [62-65]. The same trends were observed in the risk of hip fracture associated with PM₂.₅, SO₂, NO, NO₂, and O₃ [66,67]. Shi et al [35] were the first to investigate the effects of air pollution on mortality after hip fracture. In their nationwide cohort study in the United Kingdom, the levels of specific air pollutants, such as PM₂.₅, PM₁₀, and NO₂ exhibited a moderately positive association with the increases in the incidence and 30-day mortality rates of hip fracture. Thus far, their study was the only one on the negative effects of air pollution on mortality after hip fracture. However, the focus on 30-day mortality limited the scope of the long-term prognosis assessment. Moreover, the representativeness and extrapolation of their findings were limited because of the lack of age-standardization of the incidence; the possibility of errors in air pollution assessment owing to the use of data acquired only from the hospital region; and the presence of key unadjusted confounders, such as sex, season, temperature, socioeconomic, and CCI score. The aforementioned limitations of their study and the complexity of air pollution resulted in an unclear dose-response effect. To overcome this problem, we analyzed many pollutants; the robust evidence obtained in this study may help policy makers assess the economic loss and burden associated with air pollution and devise effective prevention strategies.

Possible Mechanisms

Although the precise mechanisms underlying the association between air pollution and mortality remain unclear, we hypothesized the following mechanism: air pollution contributes to frailty in older adults, thus increasing their susceptibility to mortality after hip fracture surgery. Air pollution is an independent risk factor for mortality from respiratory and cardiac diseases [25,68]. Furthermore, air pollution modifies the aging process by interfering with biological pathways; these effects are supported by toxicological evidence indicating that air pollution enhances oxidative stress [43-52], activates systemic inflammation [36-42], and causes metabolic disorders [69-72], genetic and epigenetic alterations [73-81], and vector-mediated pathogen transmission [30]. Zanobetti and Schwartz [82] suggested that air pollution influences the frail population in the following manners: increasing the mortality rate, increasing recruitment into the frail group, and delaying recovery. This finding indicates that air pollution modifies aging through frailty. Moreover, air pollution may aggravate age-related decline and functional deterioration at the cellular, tissue, and organ levels [83], thus increasing individuals’ vulnerability to disease development and mortality incidence. Given the association of exposure to ambient air pollution with all-cause and specific morbidity and mortality, our hypothesis and the obtained significant HRs appear to be highly conceivable and concomitant.

We found a negative association between the risk of mortality after hip fracture and exposure to O₃ pollution. Various conclusions have been reported regarding the effects of O₃ exposure [84-88]. Studies reporting protective effects, which are consistent with our findings, have provided the following explanations. First, sunlight facilitates O₃ formation; O₃ is a seasonal pollutant because its level is higher in summer than in winter. Therefore, seasonal effects should be considered when analyzing the effects of O₃ on health [84,89]. In this study, season served as a confounder in our regression models for potentially adjusting for the seasonal effects. Second, O₃ appears to be associated with other air pollutants; hence, the effects of O₃ on health should be considered together with those of other pollutants. Thus, in this study, we performed a Pearson correlation analysis of air pollutants (Multimedia Appendix 1). To address the confounding impact of other pollutants and to prevent potential issues related to multicollinearity, we used potential associations among air pollutants to evaluate the impacts of multiple pollutants, selectively regulating those based on their weak correlations with other air pollutants. We defined the absolute value of correlation coefficients of <0.3 to indicate a low strength of correlation. However, the negative association between mortality and O₃ pollution persisted (Multimedia Appendix 1) in the multiple-pollutant models of the targeted pollutants; thus, the second explanation was rejected. Thus, the precise mechanisms underlying the protective effects of O₃ remain unclear. Nevertheless, other theories may provide some insights into these mechanisms, as follows: the enhanced expression levels of endothelial and inducible nitric oxide synthase [90-92], improvement of hemorheological parameters and oxygen delivery [92,93], and neuroprotective effects [92,94].

CO₂ emission exhibits a nonsignificant negative correlation with mortality after hip fracture. For the correlation between exposure to CO₂ pollution and the risk of mortality within 1 year after hip fracture surgery, the HR was 0.88 (95% CI 0.76-1.02), with a P value of .09. This finding is similar to previous findings that CO₂ emission exerted positive but nonsignificant effects on longevity [95-97]. After adjustment for the urbanization level and insurance amount, our analysis revealed a positive but nonsignificant trend in the health effects of CO₂ exposure, which is consistent with the literature. However, to the best of our knowledge, this has so far not been explained nor recognized in the literature. The need for more extensive research in the future is evident, as there remain numerous unexplored facets and complexities within this subject that warrant thorough investigation to advance our knowledge and drive progress in the field.

Strengths

Our study has several strengths. First, to the best of our knowledge, this study is the first to indicate that long-term exposure to air pollution contributes to mortality within 1 year after hip fracture surgery. Second, we analyzed many key air pollutants (n=12). Third, we assessed the mortality within 1 year of surgery, thus focusing on the long-term outcomes of hip
fracture. Finally, we adjusted for most of the major confounders, such as sex, season, temperature, socioeconomic, and CCI score, which further strengthened our findings and provided robustness for our primary conclusion.

**Limitations**

This study has some limitations. First, although smoking is a key factor influencing ambient air quality, we could not include it in the analysis because of data unavailability; this might have led to a potential bias. Second, the NHIRD does not include data on disease severity, clinical manifestations, laboratory findings, and mortality causes. Hence, we could not determine the specific causes of mortality. Third, the possibility of errors in the assessment of air pollution cannot be ignored. Fourth, because we regarded death as withdrawal from the National Health Insurance program, bias might have been introduced into our findings. Fifth, while we considered any changes in residence during the assessment period, it is worth noting that the residential postcodes may not always be updated accurately in the insurance registration data. This could result in inaccuracies when reflecting patient characteristics. Sixth, although we adjusted for several confounders, unmeasured or unknown confounders might have introduced bias into our results. Given the limitations of retrospective health insurance database cohorts, which may not encompass all potential confounding factors, the possibility of residual bias due to unmeasured variables remains. Seventh, although we aimed to analyze individual molecules separately, it is possible that we neglected the potential for nonlinear relationships or the possibility of mixed effects of other molecules. Finally, there is a potential that the current levels of air pollution may differ from the data used in our study. However, due to the availability of data, this is the most up-to-date information we can analyze. Furthermore, our paper highlights the health and societal hazards of air pollution, and we hope that our research findings can provide empirical support for air pollution control policies across various times and situations. Despite the aforementioned limitations, this study provides adequate, high-quality evidence to support our conclusion that long-term exposure to SO₂, CO, PM₁₀, PM₂.₅, NOₓ, NO, NO₂, THC, NMHC, and CH₄ is associated with an increased risk of all-cause mortality within 1 year after hip fracture surgery.

**Conclusions**

In this nationwide population-based retrospective study, we discovered that long-term exposure to SO₂, CO, PM₁₀, PM₂.₅, NOₓ, NO, NO₂, THC, NMHC, and CH₄ is correlated with an increased risk of all-cause mortality within 1 year after hip fracture surgery in Taiwanese adults older than 60 years. Furthermore, 1 SD increment in the average exposure levels of SO₂, CO, PM₁₀, PM₂.₅, NOₓ, NO, NO₂, THC, NMHC, and CH₄ was associated with 14%, 49%, 18%, 12%, 41%, 33%, 38%, 20%, 9%, and 26% increases in mortality risk, respectively. However, O₃ exposure was associated with a significant reduction (35%) in HR, whereas CO₂ exhibited a nonsignificant trend of potentially protective effects. Thus, our findings provide robust evidence that can be used by the government to devise and implement air pollution prevention strategies for reducing the burden on the health care system. Nonetheless, future studies with well-adjusted confounders are warranted to investigate the correlation between air pollution and hip fracture mortality and the underlying pathophysiologic mechanisms.

**Acknowledgments**

The authors thank the programmer and data technologists from the MetaTrial Platform, a cloud-based solution using the R language, for data mining, sorting, merging, algorithm building, and statistical application for various study designs. The authors also would like to acknowledge the Laboratory Animal Center at Taipei Medical University for technical support in English editing. The authors attest that there was no use of generative artificial intelligence (AI) technology in the generation of text, figures, or other informational contexts of this manuscript. The authors are grateful to Wan Fang Hospital (Grant numbers 113-wf-eva-31 and 113-wf-swf-03) for financially supporting this research.

**Data Availability**

All data generated or analyzed during this study are included in this published paper and its Multimedia Appendix files. More detailed data sets are not publicly accessible due to the requirement of obtaining approval from the Taiwan Ministry of Health and Welfare. Researchers interested in obtaining access to this data set may initiate the process by submitting an application form to the Ministry of Health and Welfare. For further guidance and assistance, please reach out to the MOHW staff via email at stcarolwu@mohw.gov.tw. The address of the Taiwan Ministry of Health and Welfare is as follows: No.488, Sec. 6, Zhongxiao E. Rd., Nangang Dist., Taipei City 115, Taiwan. You can also contact them by phone at +886-2-8590-6848.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

The International Classification of Diseases, Ninth Revision, (ICD-9) codes and Anatomical Therapeutic Chemical (ATC) Classification code established in the inclusion criteria, exclusion criteria, and definition of confounders.

[DOCX File, 20 KB - publichealth_v10i1e46591_app1.docx]
Multimedia Appendix 2
Characteristics of the study population across the tertiles of SO\textsubscript{2} exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app2.docx ]

Multimedia Appendix 3
Characteristics of the study population across the tertiles of CO\textsubscript{2} exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app3.docx ]

Multimedia Appendix 4
Characteristics of the study population across the tertiles of CO exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app4.docx ]

Multimedia Appendix 5
Characteristics of the study population across the tertiles of O\textsubscript{3} exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app5.docx ]

Multimedia Appendix 6
Characteristics of the study population across the tertiles of PM\textsubscript{10} exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app6.docx ]

Multimedia Appendix 7
Characteristics of the study population across the tertiles of PM\textsubscript{2.5} exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app7.docx ]

Multimedia Appendix 8
Characteristics of the study population across the tertiles of NO\textsubscript{x} exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app8.docx ]

Multimedia Appendix 9
Characteristics of the study population across the tertiles of NO exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app9.docx ]

Multimedia Appendix 10
Characteristics of the study population across the tertiles of NO\textsubscript{2} exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app10.docx ]

Multimedia Appendix 11
Characteristics of the study population across the tertiles of THC exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app11.docx ]

Multimedia Appendix 12
Characteristics of the study population across the tertiles of NMHC exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app12.docx ]

Multimedia Appendix 13
Characteristics of the study population across the tertiles of CH\textsubscript{4} exposure.
[DOCX File, 25 KB - publichealth_v10i1e46591_app13.docx ]

Multimedia Appendix 14
Pearson correlation analysis for air pollutants detected over the exposure period.
[DOCX File, 23 KB - publichealth_v10i1e46591_app14.docx ]

Multimedia Appendix 15
Hazard ratios for long-term ozone exposure at 1 standard deviation increment associated with mortality rate.
References


44. Andersson H, Piras E, Demmu J, Hellman B, Brittebo E. Low levels of the air pollutant 1-nitropyrene induce DNA damage, increased levels of reactive oxygen species and endoplasmic reticulum stress in human endothelial cells. Toxicology 2009;262(1):57-64. [doi: 10.1016/j.tox.2009.05.008] [Medline: 19460413]


Abbreviations

CCI: Charlson Comorbidity Index
CH4: methane
CO: carbon monoxide
CO2: carbon dioxide
EPA: Environmental Protection Administration
HR: hazard ratio
ICD-9: International Classification of Diseases, Ninth Revision
NHIRD: National Health Insurance Research Database
NMHC: nonmethane hydrocarbons
NO: nitrogen monoxide
NO2: nitrogen dioxide
NOX: nitrogen oxides
O3: ozone
PM10: particulate matters having a size of <10 μm
PM2.5: particulate matters having a size of <2.5 μm
SO2: sulfur dioxide
THC: total hydrocarbons
Abstract

Background: Human papillomavirus (HPV) infection causes nearly all cervical cancer cases and is a cause of anogenital and oropharyngeal cancers. The incidence of HPV-associated cancers is inequitable, with an increased burden on marginalized groups in high-income countries. Understanding how immunization status varies by material and social deprivation, health system, and geospatial factors is valuable for prioritizing and planning HPV immunization interventions.

Objective: The objective of this study was to describe school-based HPV immunization rates by individual and geospatial determinants of health in Alberta, Canada.

Methods: Health administrative data for male and female individuals born in 2004 in Alberta were used to determine HPV immunization status based on age and the number of doses administered in schools during the 2014/2015-2018/2019 school years. Immunization status and its relationship with material and social deprivation and health system factors were assessed by a logistic regression model. Geospatial clustering was assessed using Getis-Ord Gi* hot spot analysis. Mean scores of material and social deprivation and health system factors were compared between hot and cold spots without full HPV immunization using independent samples t tests. A multidisciplinary team comprising researchers and knowledge users formed a co-design team to design the study protocol and review the study results.

Results: The cohort consisted of 45,207 youths. In the adjusted model, the odds of those who did not see their general practitioner (GP) within 3 years before turning 10 years old and not being fully immunized were 1.965 times higher (95% CI 1.855-2.080) than those who did see their GP. The odds of health system users with health conditions and health system nonusers not being fully immunized were 1.092 (95% CI 1.006-1.185) and 1.831 (95% CI 1.678-1.998) times higher, respectively, than health system users without health conditions. The odds of those who lived in areas with the most material and social deprivation not being fully immunized were 1.287 (95% CI 1.200-1.381) and 1.099 (95% CI 1.029-1.174) times higher, respectively, than those who lived in areas with the least deprivation. The odds of those who lived in rural areas not being fully immunized were 1.428 times higher (95% CI 1.359-1.501) than those who lived in urban areas. Significant hot spot clusters of individuals without full HPV immunization exist in rural locations on the northern and eastern regions of Alberta. Hot spots had significantly worse mean material deprivation scores ($P=0.008$) and fewer GP visits ($P=0.001$) than cold spots.
Conclusions: Findings suggest that material and social deprivation, health system access, and rural residency impact HPV immunization. Such factors should be considered by public health professionals in other jurisdictions and will be used by the Alberta co-design team when tailoring programs to increase HPV vaccine uptake in priority populations and regions.

(JMIR Public Health Surveill 2024;10:e45508) doi:10.2196/45508

KEYWORDS
co-design; geospatial; human papillomavirus; immunization; population-based; vaccine

Introduction

Background Information
Human papillomavirus (HPV) is the most common sexually transmitted agent in Canada, with upwards of 75% of Canadians experiencing at least 1 infection in their lifetime [1]. While most infections resolve on their own, some infections cause cell abnormalities, which can lead to cancer [2]. Infection with HPV causes nearly all cases of cervical cancer and is a cause of anogenital and oropharyngeal cancers in both men and women [3]. In Canada, cervical cancer is the 13th most common cancer diagnosed in women and the third most common among women aged 20–44 years [4].

Cervical cancer tends to be a disease of inequity, with an increased burden in high-income countries observed in disadvantaged groups [1,5]. A meta-analysis of 57 studies found that the relative risk of cervical dysplasia and invasive cervical cancer increases with decreasing socioeconomic status, with the relationship strongest in low- to middle-income countries and in North America [6]. The authors postulate that these results are due to relationships between socioeconomic status, HPV exposure, and access to cervical cancer screening programs. In addition to socioeconomic status, studies from Canada and the United States have found higher cervical cancer incidence among women living in rural rather than urban areas, which could be related to health care access [7,8]. HPV-associated cancer incidence can be addressed by increasing the uptake of the HPV vaccine, a safe and effective vaccine available since 2006 [9], but there is a need to explore factors associated with the uptake of the vaccine to close the gap in the inequities of cervical cancer incidence.

In 2018, the World Health Organization announced a global call to action for the elimination of cervical cancer, with a global strategy involving improved immunization, screening, and treatment [10]. The Canadian Partnership Against Cancer (CPAC) developed a corresponding action plan, with the immunization priority targeting having 90% of 17-year-olds in Canada fully immunized with the HPV vaccine by 2025 [11]. If this target is achieved, we could expect to see a 23% reduction in cervical cancer cases and a 21% decrease in cervical cancer–related mortalities [12]. Despite the HPV vaccine being publicly available for school-aged children, HPV coverage rates across most of Canada remain below the 90% target [1]. To address this coverage gap, it is imperative that we identify individual and geospatial determinants of health associated with being underimmunized with the HPV vaccine so that public health professionals can tailor public health planning and interventions toward priority populations.

Objective
The primary objective of this study was to describe school-based HPV immunization rates by individual and geospatial determinants of health in Alberta. Factors were selected through an equity-informed lens to appropriately describe HPV immunization coverage by the determinants of health. To achieve this objective, the study had 2 aims:

1. To understand variation in HPV immunization uptake in relation to determinants of health within individual and geographic area-level analyses.
2. To identify clusters of regions with a higher percent of individuals without full HPV immunization (hot spots) and those with a lower percent of individuals (cold spots) and compare the distribution of individual determinants of health across those regions.

Methods

Setting
This study took place in Alberta, Canada, which has a fully integrated, publicly funded health care system that delivers care to nearly 4.4 million people [13]. Alberta is separated into 5 health zones based on population, geography, and the distribution of health services: Calgary, Central, Edmonton, North, and South. This study investigated Alberta’s school-based HPV immunization program, which includes HPV vaccines administered by public health in schools and in Well Child Clinics across the province [14].

In Alberta, the Gardasil vaccine was originally offered to female individuals in grade 5 in 2008 (Figure 1 [15]) [16]. The original schedule consisted of 3 doses administered over a 6-month period. In 2014, the program was extended to include grade 5 male individuals, and in 2016, the Gardasil vaccine was replaced with the Gardasil-9 vaccine. In 2018, a total of two policy changes were implemented to the school-based program: (1) administration was pushed from grade 5 to 6, and (2) the number of doses was decreased from 3 to 2 for immunocompetent individuals aged 14 years and younger. As of 2020, all residents aged 9–26 years are eligible to receive the HPV vaccine free of charge.
Participants
This was a cohort study looking at male and female individuals born in 2004 in Alberta. The index event was the date on which the individual turned 10 years old, with the assumption that they were in grade 5 during the 2014-2015 school year. Individuals who were 9 years old in September-December 2014 were also included in the cohort. The cohort was followed up for 5 years after their index date, from the 2014-2015 school year when they would have first become eligible to receive the HPV vaccine in grade 5, to the 2018-2019 school year when they would have been in grade 9 and reoffered the HPV vaccine in school if they hadn’t previously declined and were not up to date.

The exclusion criteria for this study included: (1) missing age, sex, or postal code (n=0); (2) not having valid Alberta health care coverage 3 years before index (necessary for complete covariate data), at index, or during follow-up (including deaths and emigration; n=8381); and (3) living outside of Alberta (n=144).

Outcome
The outcome of interest was whether individuals in the cohort were fully or not fully immunized with the HPV vaccine. Doses were only counted toward immunization status if they occurred during the follow-up period (2014-2015 to 2018-2019 school years). Immunization status was based on age and the number of doses administered. Individuals were assessed as fully immunized after 2 doses for those 14 years of age and younger and after 3 doses for those 15 years of age and older. Individuals were considered not fully immunized if there was no record of them being immunized over the follow-up period, or they received 1 dose at 14 years of age and younger, or 1-2 doses at 15 years of age and older (ie, partially immunized). Individuals who were partially immunized were combined with those who were not immunized during the follow-up period, as they only represented 1.65% (n=746) of the cohort, and the objective of the study was to focus on those not fully immunized to inform public health planning.

Since 2008, individuals were assessed as fully immunized after 3 doses for all individuals regardless of age; however, in September 2018, the number of doses was reduced for this group after a clinical trial found that both the 2- and 3-dose schedules produced similar vaccine efficacy [17]. Immune status was not taken into consideration for immunization status in this study, as we were not able to accurately ascertain immunocompetency or HIV status from health codes associated with the Canadian Institute for Health Information (CIHI) grouper. Immunization status for this study was classified based on post-2018 age and dose number standards throughout the entire follow-up period.

In order to achieve an optimal immune response, the first and last dose of either a 2- or 3-dose HPV vaccine series is meant to be administered over a 6-month interval [4]. The time between the first and last dose was not considered in this analysis as we deviated from the historical vaccine schedule and instead based immunization status on post-2018 age and dose number standards (ie, individuals were assessed as fully immunized at 2 doses for those 14 years of age and younger, or 3 doses for those 15 years of age and older).

Data Sources
The study cohort was created using administrative data extracted from the Alberta Health Services Enterprise Data Warehouse. The Alberta Health Care Insurance Plan (AHCIP) Provincial Registry contains population demographics for all persons in Alberta covered for basic medical and hospital insurance during a given fiscal year and was used to define our study cohort. Using a provincial health care number, the cohort was linked to other population-based data sets, including the Alberta Health Postal Code Translator File, the 2016 Pampalon Index, physician claims, the Discharge Abstract Database (DAD), and the National Ambulatory Care Reporting System (NACRS). The Alberta Health Postal Code Translator File was used to establish geographical boundaries to classify local geographic areas (LGAs), health zones, and urban and rural areas. The 2016 Pampalon Index uses Census data at the dissemination area level to present socioeconomic disparities among the population. The index is made up of material and social components, stratifying the population into 5 quintiles, with the first level being the least deprived, and the fifth being the most deprived. The material component is characterized by low income, an insecure job situation, and low education, while the social component is defined by being a single parent, being separated, divorced, or widowed, and living alone [18]. Physician claims capture data for fee-for-service and shadow-billed (ie, bills

Figure 1. Evolution of Alberta’s human papillomavirus immunization program.
submitted for reporting purposes, but practitioners compensated through alternate methods) claims by physicians and provides information on service dates, number of visits, and services provided [19]. Physician claims data cover the 3 years before index (ie, the date an individual turned 10 years old). DAD and NACRS capture acute care hospitalizations and community-based and hospital-based ambulatory care, including emergency department visits, respectively, in Alberta, and include records for demographic, diagnostic, procedural, and treatment information for each visit [20]. DAD and NACRS were used to define the CIHI grouper, which categorizes individuals into health profile groups based on the history of health services received and the presence of health conditions. Individuals from the study cohort were assigned a health profile group based on diagnoses from a variety of health sectors, including but not limited to hospitals, emergency rooms, urgent care centers, and specialist and family practice settings. A total of 226 health conditions from 16 categories (eg, minor acute, major chronic, and other mental health) were used to synthesize an individual’s clinical profile to the most complex and clinically relevant health conditions present. Youths were placed into 1 of the following 3 categories: health system nonuser, health system user without health conditions (ie, without 1 of the 226 health conditions), and health system user with health conditions (ie, with 1 of the 226 health conditions). Groups were based on data from the 2013-2014 fiscal year [20,21]. The data set was then merged with the Meditech Health Information System, which contains information on all HPV vaccines administered by public health in schools and in Well Child Clinics across the province. After merging the Meditech Health Information System data set with the AHCIP Provincial Registry and linked population-based data sources and applying the exclusion criteria, the final merged data set consisted of 45,094 observations (Figure 2).
Figure 2. Creation of the study cohort data set. The (n=2751) surplus in the initial cohort when compared to the original Alberta Health Care Insurance Plan (AHCIP) Provincial Registry and linked data sources study cohort was due to the “did not have Alberta coverage at index” exclusion criterion being pre-applied to the AHCIP Provincial Registry but not the Meditech data set. This surplus is removed in the next exclusion step. HPV: Human Papillomavirus.

Individual postal code was geocoded using the Alberta Health Postal Code Translator File and estimates of percent of individuals without full HPV immunization were aggregated at the LGA. The LGAs were chosen for this spatial analysis as they are the smallest geographic area that Alberta Health Services and Alberta Health use as official geographies [22], and they determine how health services are distributed across Alberta. In Alberta, there are 132 LGAs of varying sizes divided across the 5 health zones.

Analysis
Variation in HPV immunization status and its relationship with determinants of health were assessed by analytical, data visualization, and geospatial techniques. Variables for inclusion in this analysis were selected a priori based on our conceptual framework (Figure 3; adapted from Dahlgren and Whitehead [23]) and data availability. Our conceptual framework was based on determinants of health documented in the literature (for example [5,24-28]) that impact student HPV immunization coverage.
Characteristics were summarized by HPV immunization status using counts and percentages. The use of mean and SD was determined based on the distribution of the data. Before their inclusion in the logistic regression analysis, correlations between covariates were checked using chi-square and Spearman rank-order correlation tests, where appropriate. A logistic regression model was used to estimate the odds ratios (ORs) comparing not fully immunized with fully immunized status for the characteristics of interest. ORs, adjusted odds ratios (aORs), and 95% CIs are reported. Observations were excluded from the logistic regression if they had missing covariate values (n=1694). No variables had more than 5% missing observations.

Global Moran’s $I$ was used to detect whether spatial autocorrelation exists, meaning that individuals without full HPV immunization are spatially clustered in Alberta. Moran’s $I$ ranges from +1 to −1 with positive values indicating spatial autocorrelation (ie, clustering of similar rates), while negative values indicate similar rates are located far away from each other (ie, dispersion). A $P$ value that is smaller than .05 means we can reject the null hypothesis of complete spatial randomness and accept that spatial autocorrelation exists and there is clustering [29].

To detect the location and magnitude of spatial clustering, we used local Getis-Ord Gi* optimized hot-spot analysis. Getis-Ord Gi* hot-spot analysis was used to identify areas of statistically significant clusters of LGAs with high rates of individuals without full HPV immunization are spatially clustered in Alberta. Moran’s $I$ ranges from +1 to −1 with positive values indicating spatial autocorrelation (ie, clustering of similar rates), while negative values indicate similar rates are located far away from each other (ie, dispersion). A $P$ value that is smaller than .05 means we can reject the null hypothesis of complete spatial randomness and accept that spatial autocorrelation exists and there is clustering [29].

Ethical Considerations
This study received ethics approval by the Health Research Ethics Board of Alberta (HREBA.CC-20-0425).

Co-Design Team
A multidisciplinary team based in Alberta formed a co-design team comprising experts in cancer prevention, epidemiology, immunization programming, biostatistics, public health planning, and communicable disease control to collaborate on this study. Co-design, or collaborative design, involves the active engagement of stakeholders involved with an issue who work together to design solutions to address the problem [30,31]. By drawing on the collective perspectives, experiences, and strengths of the group, there can be improved idea generation, better interventions, and enhanced outcomes [30]. The co-design team designed the study protocol, reviewed the study results, and will use the findings to identify priority populations and regions that should be engaged when designing interventions to improve HPV immunization coverage across Alberta. This collaboration facilitated the creation of relevant questions,
appropriate methods, and how the results could be used for public health planning.

**Results**

**Descriptive Statistics**

Descriptive statistics summarizing characteristics by HPV immunization status are displayed in Table 1. While the majority of youths who were fully immunized received their doses in grade 5, nearly 10% (2750/31,388) of youths received all of their doses in grade 9. A lesser proportion of youths who were fully immunized were from the Central, North, and South zones, compared to those in the Calgary and Edmonton zones. Those who were fully immunized tended to live in urban rather than rural areas. Youths who were not fully immunized tended to live in areas with more material and social deprivation than those who were fully immunized. Youths who were fully immunized were more likely to have had a GP visit in the 3 years before the index and tended to have a higher number of GP visits compared with youths who were not fully immunized. Those who were not fully immunized were more likely to be health system nonusers or health system users with health conditions.
Table 1. Characteristics of Albertans by human papillomavirus immunization status (n=45,207).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Fully immunized (n=31,388)</th>
<th>Not fully immunized (n=13,819)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>14,924 (47.55)</td>
<td>7187 (52.01)</td>
</tr>
<tr>
<td>Male</td>
<td>16,464 (52.45)</td>
<td>6632 (47.99)</td>
</tr>
<tr>
<td><strong>Grade received, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 5</td>
<td>27,632 (88.03)</td>
<td>211 (28.28)</td>
</tr>
<tr>
<td>Grade 9</td>
<td>2750 (8.76)</td>
<td>474 (63.54)</td>
</tr>
<tr>
<td>Both</td>
<td>823 (2.62)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Neither</td>
<td>183 (0.58)</td>
<td>61 (8.18)</td>
</tr>
<tr>
<td>Not applicable&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0 (0)</td>
<td>13,073 (94.6)</td>
</tr>
<tr>
<td><strong>Health zone, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calgary</td>
<td>12,365 (39.39)</td>
<td>4352 (31.49)</td>
</tr>
<tr>
<td>Central</td>
<td>3389 (10.80)</td>
<td>2153 (15.58)</td>
</tr>
<tr>
<td>Edmonton</td>
<td>9600 (30.58)</td>
<td>3815 (27.61)</td>
</tr>
<tr>
<td>North</td>
<td>3633 (11.57)</td>
<td>2252 (16.30)</td>
</tr>
<tr>
<td>South</td>
<td>2401 (7.65)</td>
<td>1247 (9.02)</td>
</tr>
<tr>
<td><strong>Area, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>24,593 (78.35)</td>
<td>9531 (68.97)</td>
</tr>
<tr>
<td>Rural</td>
<td>6795 (21.65)</td>
<td>4288 (31.03)</td>
</tr>
<tr>
<td><strong>Material deprivation, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (least deprived)</td>
<td>4995 (16.46)</td>
<td>1810 (13.74)</td>
</tr>
<tr>
<td>2</td>
<td>5895 (19.43)</td>
<td>2312 (17.55)</td>
</tr>
<tr>
<td>3</td>
<td>5914 (19.49)</td>
<td>2511 (19.06)</td>
</tr>
<tr>
<td>4</td>
<td>6725 (22.16)</td>
<td>2893 (21.96)</td>
</tr>
<tr>
<td>5 (most deprived)</td>
<td>6812 (22.45)</td>
<td>3646 (27.68)</td>
</tr>
<tr>
<td><strong>Social deprivation, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (least deprived)</td>
<td>6797 (22.40)</td>
<td>2717 (20.63)</td>
</tr>
<tr>
<td>2</td>
<td>5386 (17.75)</td>
<td>2167 (16.45)</td>
</tr>
<tr>
<td>3</td>
<td>5974 (19.69)</td>
<td>2540 (19.28)</td>
</tr>
<tr>
<td>4</td>
<td>6452 (21.26)</td>
<td>2906 (22.06)</td>
</tr>
<tr>
<td>5 (most deprived)</td>
<td>5732 (18.89)</td>
<td>2842 (21.58)</td>
</tr>
<tr>
<td><strong>GP&lt;sup&gt;b&lt;/sup&gt; Visits (3 years before index), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>27,259 (86.85)</td>
<td>9813 (71.01)</td>
</tr>
<tr>
<td>No</td>
<td>4129 (13.15)</td>
<td>4006 (28.99)</td>
</tr>
<tr>
<td><strong>Number of GP visits (3 years before index), mean (SD)</strong></td>
<td>5.92 (5.42)</td>
<td>5.19 (4.86)</td>
</tr>
<tr>
<td><strong>CIHI&lt;sup&gt;c&lt;/sup&gt; grouper, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health system nonuser</td>
<td>7343 (23.39)</td>
<td>5641 (40.82)</td>
</tr>
<tr>
<td>Health system user with no health conditions</td>
<td>2889 (9.20)</td>
<td>907 (6.56)</td>
</tr>
<tr>
<td>Health system user with health conditions</td>
<td>21,156 (67.40)</td>
<td>7271 (52.62)</td>
</tr>
</tbody>
</table>

<sup>a</sup>No school grade dose data are available for these records as these are youths who did receive a human papillomavirus dose from public health.

<sup>b</sup>GP: general practitioner.

<sup>c</sup>CIHI: Canadian Institute for Health Information.
Logistic Regression

A logistic regression model was fitted to estimate the aORs of not being fully HPV immunized, compared to being fully immunized according to study characteristics (Table 2). The odds of those who did not see their GP in the 3 years before index not being fully immunized were 1.965 (95% CI 1.855-2.080) times higher than those who did see their GP. The odds of health system users with health conditions and health system nonusers not being fully immunized were 1.092 (95% CI 1.006-1.185) and 1.831 (95% CI 1.678-1.998) times higher than health system users without health conditions. The odds of those who lived in areas with the most material and social deprivation not being fully immunized were 1.287 (95% CI 1.200-1.381) and 1.099 (95% CI 1.029-1.174) times higher than those who live in areas with the least deprivation. The odds of those who lived in rural areas not being fully immunized were 1.428 (95% CI 1.359-1.501) times higher than those who lived in urban areas.

Table 2. Logistic regression model examining the odds of not being fully immunized based on determinants of health.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Odds of not being fully immunized</th>
<th>aOR(^{b}) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP(^{c}) visits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.00 (ref(^{d}))</td>
<td>1.00 (ref)</td>
</tr>
<tr>
<td>No</td>
<td>2.695 (2.565-2.831)</td>
<td>1.965 (1.855-2.080)</td>
</tr>
<tr>
<td>CIHI(^{e}) grouper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health system user without health conditions</td>
<td>1.00 (ref)</td>
<td>1.00 (ref)</td>
</tr>
<tr>
<td>Health system user with health conditions</td>
<td>1.095 (1.011-1.185)</td>
<td>1.092 (1.006-1.185)</td>
</tr>
<tr>
<td>Health system nonuser</td>
<td>2.447 (2.254-2.657)</td>
<td>1.831 (1.678-1.998)</td>
</tr>
<tr>
<td>Material deprivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (least deprived)</td>
<td>1.00 (ref)</td>
<td>1.00 (ref)</td>
</tr>
<tr>
<td>2</td>
<td>1.082 (1.007-1.163)</td>
<td>1.054 (0.978-1.135)</td>
</tr>
<tr>
<td>3</td>
<td>1.172 (1.091-1.258)</td>
<td>1.110 (1.031-1.195)</td>
</tr>
<tr>
<td>4</td>
<td>1.187 (1.108-1.272)</td>
<td>1.071 (0.996-1.151)</td>
</tr>
<tr>
<td>5 (most deprived)</td>
<td>1.477 (1.381-1.580)</td>
<td>1.287 (1.200-1.381)</td>
</tr>
<tr>
<td>Social deprivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (least deprived)</td>
<td>1.00 (ref)</td>
<td>1.00 (ref)</td>
</tr>
<tr>
<td>2</td>
<td>1.007 (0.941-1.076)</td>
<td>1.029 (0.960-1.102)</td>
</tr>
<tr>
<td>3</td>
<td>1.064 (0.997-1.134)</td>
<td>1.036 (0.969-1.107)</td>
</tr>
<tr>
<td>4</td>
<td>1.127 (1.059-1.199)</td>
<td>1.048 (0.982-1.118)</td>
</tr>
<tr>
<td>5 (most deprived)</td>
<td>1.240 (1.164-1.321)</td>
<td>1.099 (1.029-1.174)</td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>1.00 (ref)</td>
<td>1.00 (ref)</td>
</tr>
<tr>
<td>Rural</td>
<td>1.628 (1.557-1.703)</td>
<td>1.428 (1.359-1.501)</td>
</tr>
</tbody>
</table>

\(^{a}\)OR: odds ratio.
\(^{b}\)aOR: adjusted odds ratio.
\(^{c}\)GP: general practitioner.
\(^{d}\)ref: reference group.
\(^{e}\)CIHI: Canadian Institute for Health Information.

Geospatial Analysis

Spatial autocorrelation of individuals without full HPV immunization at the LGA level was detected using Global Moran’s \(I\). With a Moran’s \(I\) index value of 0.051, a \(z\)-score of 3.74, and \(P<.001\), the null hypothesis was rejected, and it was concluded that there was more than a 99% probability that the distribution of individuals without full HPV immunization formed a clustered pattern by LGA in Alberta, Canada.

Getis-Ord \(Gi^*\) hot-spot analysis was used to determine the location and magnitude of the spatial autocorrelation. Results from Figure 4 indicate that there were 48 statistically significant hot or cold spots of individuals who were not fully immunized with the HPV vaccine. There were 2 clusters of hot spots in the North zone (High Level LGA) and on the eastern side of Alberta.
that cross the North and Central zones (St. Paul, Two Hills County, and Vegreville/Minburn County LGAs). There were also 2 large statistically significant cold spots within the Edmonton and Calgary zones. Nonsignificant data are presented in yellow. Due to small sample sizes (ie, under 20 individuals making up the estimate), data in 16 LGAs were suppressed—these areas are presented in white.

Independent samples t tests were used to determine if determinants of health vary between hot spots of individuals without full HPV immunization and cold spots (Table 3). Individuals living in hot spots were more likely to have a higher mean material deprivation score than those living in cold spots. Individuals living in hot spots had a statistically lower mean number of GP visits compared to those living in cold spots.

**Figure 4.** Areas of statistically significant clusters of local geographic areas with high (ie, hot spots) and low (ie, cold spots) rates of male and female individuals born in 2004 without full human papillomavirus immunization in Alberta, Canada.
Experiencing more material and social deprivation appears to impact vaccine acceptance, uptake, and series completion [33-39]. In school, which has been shown to influence vaccine hesitancy, but further research is needed to dissect the relationship between vaccine hesitancy and immunization behaviors at the community level [48].

**Strengths and Limitations**
This study has several methodological strengths. Large population-based sample sizes have high internal validity, and findings could be generalized to jurisdictions similar to Alberta. In addition, the HPV immunization data were sourced from the Meditech Health Information System, an administrative database, which is significantly more reliable than self-reporting, which is susceptible to recall bias. Furthermore, by solely focusing on doses administered by public health, recommendations for improvement can focus on school-based programs.

In the past, it has taken an average of 17 years for research evidence to move into practice and implementation [49]. To circumvent this issue, it is critical for researchers to partake in integrated knowledge translation, wherein knowledge users (eg, policy makers, practitioners, and administrators) are collaborators throughout the entire research process [50]. By

### Discussion

**Principal Results**
This population-based cohort study analyzed material and social deprivation, health system, and geospatial factors to describe school-based HPV immunization rates in Alberta. Overall, 69% (31,388/45,207) of the cohort was fully immunized with the HPV vaccine, and 31% (13,819/45,207) were not fully immunized. One interesting finding was that while most fully immunized youths in Alberta received their doses in grade 5, the grade 9 catch-up program caught almost an additional 10% (2750/31,388). While the majority of Canadian provinces and territories implemented catch-up programs for girls when the HPV vaccine was initiated in the respective jurisdiction, only 2 other provinces implemented such a program for boys [32]. Alberta continues to rescreen youths in grade 9 to determine whether they are missing one or more doses of their HPV vaccine. Considering the proportion of youths who were immunized by this catch-up program, it is an initiative that other jurisdictions could consider implementing to increase HPV vaccine uptake.

The results indicate that a lack of health system use may be an important indicator of HPV vaccine uptake. Individuals who do not visit their GP are less likely to be recommended the HPV vaccine by a health care provider, which has been demonstrated to impact vaccine acceptance, uptake, and series completion [33-39]. In urban areas, this may be more of a matter of taking advantage of missed opportunities for education when children and their parents or guardians visit their health care provider. In rural areas, it may be more of an issue of access to a health care provider [1,40]. On the other hand, the parents or guardians of health system nonusers may have lower trust in public institutions, which may impact their willingness to consent to the HPV vaccine for their child. For example, direct or intergenerational trauma associated with colonization or previous negative experience in health care, such as racism or discrimination, has discouraged trust in government-provided health care for some Indigenous Peoples of Canada, causing some individuals to avoid or delay seeking care [41]. One way we can improve the current system is by adopting an approach to cultural safety, which involves practitioners self-reflecting on their assumptions, challenging inequalities, and improving health care access by shifting the power from the practitioner to the patient [41,42].

Experiencing more material and social deprivation appears to be associated with not being fully immunized with the HPV vaccine. Other studies in Canada have found that living in areas of high material and social deprivation was associated with lower HPV vaccine refusal [43], those living in areas of high material deprivation were more likely to complete the HPV vaccine series [44], and those living in areas of high social deprivation had lower HPV vaccine coverage [45]. These results indicate that social and material deprivation as it relates to HPV vaccine uptake is complex, and individual-level associations may be difficult to discern using area-level information. Further analysis of individual-level factors should be an area of further research.

The results indicate that living in rural areas may be associated with not being fully immunized with the HPV vaccine. This is reflected in coverage variation by health zone, as a lesser proportion of youths who were not fully immunized were from the Central, North, and South health zones, with clustering occurring in the North and Central zones, which comprise more rural communities than the Calgary and Edmonton zones. With rural Canadian communities experiencing a shortage of family doctors [46], rural youths and their parents or guardians may not be receiving a recommendation to receive the HPV vaccine in school, which has been shown to influence vaccine acceptance, uptake, and series completion [33-39]. There is also an imbalance between the supply of nurses working in rural and remote areas in Alberta (12% in 2015) and the portion of the population residing there (18.5% in 2015) [47]. This discrepancy could lead to rural and remote public health nurses having less capacity to complete vaccine series for school-aged children. Parents and guardians in rural communities may also be declining the HPV vaccine for their child due to increased levels of vaccine hesitancy, but further research is needed to dissect the relationship between vaccine hesitancy and immunization behaviors at the community level [48].

### Table 3. Descriptive statistics and independent samples t test comparisons of material and social deprivation indicators and health system use between hot and cold spots of youths not being fully immunized with the human papillomavirus vaccine in Alberta.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hot spot, mean (SD)</th>
<th>Cold spot, mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material deprivation</td>
<td>4.30 (0.14)</td>
<td>2.91 (0.99)</td>
<td>.008</td>
</tr>
<tr>
<td>Social deprivation</td>
<td>3.15 (1.08)</td>
<td>3.06 (0.90)</td>
<td>.85</td>
</tr>
<tr>
<td>Number of GP visits</td>
<td>2.78 (1.04)</td>
<td>4.84 (1.15)</td>
<td>.001</td>
</tr>
</tbody>
</table>

*GP: general practitioner.*
focusing on questions that are relevant to practice and working with partners who are capable of implementing identified recommendations, integrated knowledge translation can effectively promote the uptake and use of research findings [50-52]. The key strength of this study was the use of a co-design team to enable an integrated knowledge translation approach. By using an integrated knowledge translation approach, the co-design team strives to shorten the typical 17-year research evidence-to-innovation gap, integrating study findings into public health planning to reach the 90% immunization target by 2025. Notably, these results will be supplemented with information on barriers and facilitators of the school-based HPV immunization program from the perspective of parents and guardians, as well as program leaders and front-line providers, through a CPAC-funded study being conducted by members of the co-design team [15].

There are several methodological limitations that should be noted, as they will impact the generalizability of the findings. First, HPV coverage was determined by applying current immunization standards to all follow-up years, despite the policy change that occurred in 2018, and immune status was not considered when determining immunization status. While this may overestimate the number of individuals fully immunized for HPV in the pre-2018 period, the objectives of this study center around determining the characteristics associated with being underimmunized in order to inform improvements to the current school-based immunization program, including who and where to target improvements, rather than solely determining coverage rates.

Second, the cohort denominator is population-based rather than school-based due to a lack of school list availability. Consequently, the AHCIP Provincial Registry needed to be used as a proxy denominator, which could result in underestimated coverage rates.

Third, the outcome variable of not being fully immunized may not necessarily mean that the individual was not fully immunized in certain situations. Immunization data were gathered from the Meditech Health Information System, thus only allowing us to classify those HPV doses administered by Alberta public health. Children may have received immunization at a local pharmacy, family doctor, or out of province, and consequently would have been misclassified as not being fully immunized.

Fourth, the exclusion criterion of removing those who did not have valid Alberta health care coverage over the full study period represented more than 5% (8381/53,588) of the original cohort. These individuals were excluded due to incomplete data, either for covariate data before index or immunization data at index or during follow-up. Due to incomplete data, it cannot be determined if these individuals would have a different propensity to be fully immunized if they had had access to the school-based HPV immunization program throughout the entire study period; thus, these individuals were excluded to avoid bias in the results.

Fifth, attributing area-level information to individuals, as was the case for the Pampalon Index and comparisons of the hot and cold spots, assumes area-level data can be attributed to all individuals in that geographic location, which is an ecological fallacy. Research has also shown that such inferences can attenuate true relationships between the factor and the outcome (ie, immunization status) [53]. These results will have to be interpreted with caution to ensure appropriate conclusions are drawn.

Lastly, LGA was used as the boundary for our geographic mapping. This boundary can help identify areas of the province to prioritize for intervention, but LGAs are based on population patterns of health system access (ie, use of hospital services) and may not capture the similarities in HPV immunization outcomes across these boundaries. Further research could explore the use of other geographic boundaries that might provide additional context into HPV immunization rates, such as school districts. Despite this limitation, LGA is currently the smallest official geographic unit used by Alberta Health Services and Alberta Health [22], provides sufficient sample size to allow for accurate and precise estimates in most LGAs, and has been used in other geospatial studies of health outcomes in Alberta [54-56]. Furthermore, in 16 LGAs, there were fewer than 20 cases, making up the estimates for the geospatial analysis. While conducting hot-spot analysis on small sample sizes is possible, it limits the reliability of the estimates over time. Therefore, we chose to suppress the findings from these 16 LGAs, which limits the information available in those specific areas but was necessary to avoid reporting hot spots based on potentially unstable estimates.

Conclusions
This study highlighted how HPV immunization status varies by individual and geospatial determinants of health. Those who were not fully immunized with the HPV vaccine tended to be health system nonusers, less likely to visit their GP, live in rural areas, and experience higher levels of material and social deprivation. In Alberta, there were 2 statistically significant hot spots of individuals not fully immunized with the HPV vaccine, located on the north and east sides of the province, with individuals living in hot spots being more likely to have a higher mean material deprivation score and less likely to visit a GP than those living in cold spots.

Study results will enable the co-design team to make evidence-informed decisions when tailoring the school-based HPV immunization program for priority populations and regions. Public health professionals in other jurisdictions should consider these determinants of health when designing interventions to increase HPV immunization coverage.

Acknowledgments
We would like to acknowledge the members of the school-based human papillomavirus (HPV) immunization project co-design team for their support throughout this project. We would like to thank Howie Thomson for the creation of Figures 1-3. Data extraction and linkage support were provided by the Alberta Strategy for Patient Oriented Research Data and Research Services.
This research was funded by the Urban Public Health Network in partnership with the Canadian Partnership Against Cancer (CPAC), which is funded by Health Canada, as well as Cancer Prevention and Screening Innovation in partnership with Alberta Health. The provision of funding by Alberta Health does not signify that this project represents the policies or views of Alberta Health. CPAC and Alberta Health had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest
None declared.

References
13. About AHS. Alberta Health Services. URL: https://www.albertahealthservices.ca/about/about.aspx [accessed 2021-09-07]


Abbreviations

- AHCIP: Alberta Health Care Insurance Plan
- aOR: adjusted odds ratio
- CIHI: Canadian Institute for Health Information
- CPAC: Canadian Partnership Against Cancer
- DAD: Discharge Abstract Database
- GP: general practitioner
- HPV: human papillomavirus
- LGA: local geographic area
- NACRS: National Ambulatory Care Reporting System
- OR: odds ratio

https://publichealth.jmir.org/2024/1/e45508
Timely Pulmonary Tuberculosis Diagnosis Based on the Epidemiological Disease Spectrum: Population-Based Prospective Cohort Study in the Republic of Korea

Yousang Ko, MD, PhD; Jae Seuk Park, MD, PhD; Jinsoo Min, MPH, MD; Hyung Woo Kim, MD; Hyeon-Kyoung Koo, MD; Jeeyoun Oh, MD; Yun-Jeong Jeong, MD; Eunhye Lee, MD; Bumhee Yang, MD; Ju Sang Kim, MD; Sung-Soon Lee, MD, PhD; Yunhyung Kwon, PhD; Jiyeon Yang, RN, MSN; Jieyoun Han, PhD; You Jin Jang, MS; Jinseob Kim, MD, MPH

Abstract

Background: Timely pulmonary tuberculosis (PTB) diagnosis is a global health priority for interrupting transmission and optimizing treatment outcomes. The traditional dichotomous time-divided approach for addressing time delays in diagnosis has limited clinical application because the time delay significantly varies depending on each community in question.

Objective: We aimed to reevaluate the diagnosis time delay based on the PTB disease spectrum using a novel scoring system that was applied at the national level in the Republic of Korea.

Methods: The Pulmonary Tuberculosis Spectrum Score (PTBSS) was developed based on previously published proposals related to the disease spectrum, and its validity was assessed by examining both all-cause and PTB-related mortality. In our analysis, we integrated the PTBSS into the Korea Tuberculosis Cohort Registry. We evaluated various time delays, including patient, health care, and overall delays, and their system-associated variables in line with each PTBSS. Furthermore, we reclassified...
the scores into distinct categories of mild (PTBSS=0-1), moderate (PTBSS=2-3), and severe (PTBSS=4-6) using a multivariate regression approach.

**Results:** Among the 14,031 Korean patients with active PTB whose data were analyzed from 2018 to 2020, 37% (n=5191), 38% (n=5328), and 25% (n=3512) were classified as having a mild, moderate, and severe disease status, respectively, according to the PTBSS. This classification can therefore reflect the disease spectrum of PTB by considering the correlation of the score with mortality. The time delay patterns differed according to the PTBSS. In health care delays according to the PTBSS, greater PTB disease progression was associated with a shorter diagnosis period, since the condition is microbiologically easy to diagnose. However, with respect to patient delays, the change in elapsed time showed a U-shaped pattern as PTB progressed. This means that a remarkable patient delay in the real-world setting might occur at both apical ends of the spectrum (ie, in both mild and severe cases of PTB). Independent risk factors for a severe PTB pattern were age (adjusted odds ratio 1.014) and male sex (adjusted odds ratio 1.422), whereas no significant risk factor was found for mild PTB.

**Conclusions:** Timely PTB diagnosis should be accomplished. This can be improved with use of the PTBSS, a simple and intuitive scoring system, which can be more helpful in clinical and public health applications compared to the traditional dichotomous time-only approach.

*JMIR Public Health Surveill 2024;10:e47422* doi:10.2196/47422

**KEYWORDS**
pulmonary tuberculosis; disease spectrum; timely diagnosis; patient delay; health care delay; risk factor; epidemiological disease; tuberculosis; treatment; TB; PTB disease spectrum; mortality; early diagnosis

### Introduction

In 2020, an estimated 10 million cases of tuberculosis (TB) were reported worldwide [1]. In the Republic of Korea (ROK), the number of notified TB cases had long remained stable without a decrease [2,3]; however, the number of notified TB cases has decreased significantly in the last decade following continuous nationwide efforts [4,5]. Specifically, in 2021, the overall notification rate for incident TB cases (n=18,335) was 35.7/100,000 persons, which constituted a 53.6% decrease in the number of cases compared with those notified in 2011 (39,557 incident TB cases; notification rate 78.9/100,000 persons) [3]. As a low-TB-burden country, it is now an appropriate time to take the next step toward the goal of TB elimination in the ROK.

Timely TB diagnosis plays a crucial role in TB elimination by reducing the disease burden and preventing further community-based infections [6,7]. Numerous studies in various countries have focused on estimating the time delay for early TB diagnosis and treatment while identifying associated risk factors [7-19]. Previously, we also conducted research at the national level in the ROK, following a similar approach and drawing on references from previous studies [20]. However, we concluded that the binary time approach used in previous studies from other countries did not offer significant assistance in the context of the ROK. This is because the absolute value of the diagnostic delay time in the ROK was shorter than that observed in other countries. Therefore, we considered the clinical usefulness of this conventional dichotomous time-divided approach as limited for our context and attempted to approach the delayed diagnosis issue from a new perspective by focusing on the pulmonary TB (PTB) disease spectrum.

The “iceberg concept” has been applied to describe TB infection [21], illustrating that TB infection exists in varying quantities within the population. Furthermore, this concept helps to elucidate the disease spectrum, ranging from latent TB to clinical disease. According to the iceberg concept, two directions of effort are necessary to eliminate TB: the first involves moving from the bottom to the top of the iceberg in preventing progression from latent TB to active TB infection, whereas the second involves moving from the top to the bottom of the iceberg, reflecting the fact that early diagnosis and prompt treatment of active disease are crucial in preventing further community-based transmission of the infection.

In the ROK, active TB screening, including the progression from latent TB infection (LTBI) to clinical TB infection, especially in schools and military services, helped to significantly decrease the TB incidence among younger individuals in their 10s to 20s [4]. Active TB contact investigation and highly recommended treatment for LTBI are needed. Therefore, the remaining areas of early diagnosis of active TB at a national level are needed to assess and help establish TB control policies, while informing where efforts should be focused.

In this study, we investigated the diagnostic delay in timely PTB diagnosis; quantified the time-consuming processes contributing to patient, health care, and overall delays in access to TB treatment in the ROK based on the PTB disease spectrum; and compared demographic characteristics between cases diagnosed with mild versus severe disease. These data could reveal the epidemiological characteristics of PTB in the ROK, thereby informing the development of a real-world approach to PTB diagnosis that could provide important insights into the nature of the process wherein a patient develops symptoms and is finally diagnosed with PTB.

### Methods

**Korea TB Cohort and Recruitment**

The Korea TB Cohort (KTBC) is a nationwide, prospective, and observational cohort comprising active TB cases from 172 public-private mix (PPM)–participating hospitals in 21 districts (>70% of all patients with TB in the ROK were treated in these
PPM hospitals) since July 2018 [22]. Each patient with TB was notified, treated, and followed up every month until the completion of anti-TB treatment based on the national TB program. An investigation of the detailed data of the characteristics of active TB cases was planned to enable the establishment of a long-term plan in the future that shared the aim of an advanced National TB Elimination Project, which is operated by the Korean Academy of Tuberculosis and Respiratory Diseases under supervision of the Korea Disease Control and Prevention Agency (KDCA). The inclusion criteria for the KTBC include notified TB cases in all participating hospitals by the Korea National TB Surveillance System.

After enrollment of the KTBC, specialist TB nurses from each hospital conducted detailed interviews with patients with TB and completed standardized case-level forms. This process includes comprehensive investigations into patient information, including comorbidities, height, body weight, economic status, employment status, social status, education level, and symptoms. Additionally, data related to the program were gathered, including details about treatment initiation, discontinuation, termination, and adverse effects, along with mortality. The collected data were checked by regional and central data managers. Following a regional and central audit, a central statistical team analyzed and organized the data every quarter.

Data collected from July 1, 2018, to December 31, 2020, were obtained from the KTBC. We included all PTB cases and excluded those with only extrapulmonary TB (EPTB) because the disease spectrum of EPTB could not be clearly determined. Furthermore, we excluded patients with rifampicin-resistant PTB to reduce heterogeneity when validating the developed Pulmonary Tuberculosis Spectrum Score (PTBSS) in comparing the disease spectrum of TB with outcomes (Figure 1).

**Figure 1.** Flow diagram of the study design. EPTB: extrapulmonary tuberculosis; KTBC: Korea Tuberculosis Cohort; PTB: pulmonary tuberculosis.

---

**Definitions of Time Delays in the Diagnostic Pathway**

We divided the PTB time interval into patient, health care, and overall delays [7]. A patient delay was defined as the duration between the onset of PTB-related symptoms and the first hospital visit. A health care delay was defined as the duration between the first hospital visit and initiation of anti-PTB treatment after a confirmed diagnosis of PTB. The overall delay was defined as the sum of the patient and health care delays.

**Determination of the PTBSS Based on the PTB Disease Spectrum**

The PTBSS was designed based on previously published proposals for the PTB disease spectrum (Figure 2) [23-28]. Due to the lack of a true reference value that can accurately reflect the PTB disease spectrum, we identified factors that are commonly used in the diagnosis of PTB and can assess disease severity based on the epidemiological concept of PTB. We calculated the PTBSS according to six important variables: presence of symptoms, positive sputum in TB-polymerase chain reaction (PCR), positive sputum in an acid-fast bacilli (AFB) smear, sputum culture-positive *Mycobacterium tuberculosis*, cavitation detected on chest x-ray, and bilateral lung involvement of PTB on chest x-ray. However, since these six factors we identified and set were not weighted through statistical analysis, it may be challenging to view them as equally important. In addition, asymmetry can occur as the frequency of the manifestation of each factor can differ depending on the stage in the PTB disease spectrum. Therefore, for analysis, we proceeded by grouping in the following manner. Depending on the score assigned from 0 to 6 points, scores of 0-1, 2-3, and 4-6 were classified as mild, moderate, and severe disease, respectively.
Figure 2. Determination of the Pulmonary Tuberculosis Spectrum Score (PTBSS). The PTBSS was calculated and distributed from 0 to 6 according to the presence of six clinical variables: symptoms, positive sputum in tuberculosis polymerase chain reaction (TBPCR), positive sputum in the acid-fast bacilli (AFB) smear test, sputum culture-positive Mycobacterium tuberculosis (MTB), cavitation on chest x-ray, and bilateral lung involvement of pulmonary tuberculosis (PTB) on chest x-ray scanning.

Validation of the PTBSS

While the PTBSS was designed based on a consensus around the disease spectrum, as supported by existing scientific evidence, it is necessary to validate the PTBSS within appropriate clinical and public health contexts. However, effective and specific biomarkers that are applicable in real-world settings have not been used, and these biomarkers cannot be implemented within the framework of the KTBC, which consists of clinical practice data from a real-world setting. Therefore, we sought to validate our results by comparing mortality rates according to the PTBSS. In accordance with the iceberg concept, disease progression could ultimately result in death. Hence, higher mortality could be associated with a longer disease duration. If the group with a high PTB score exhibits a high mortality rate, this group may have a longer disease period, which can be interpreted as the time elapsed from disease onset to diagnosis.

Statistical Analysis

Categorical data are described as numbers and percentages, which were compared using the \( \chi^2 \) or Fisher exact test. Continuous variables are expressed as mean (SD) or median (IQR) for normal or skewed distributions, respectively, and were compared using the \( t \) test or Mann-Whitney \( U \) test. The Kaplan-Meier method was used to estimate the cumulative survival rates of active PTB according to the PTBSS during anti-TB treatment. To determine intergroup differences, survival curves were compared by means of log-rank tests, and hazard ratios with 95% CIs were estimated using Cox regression analysis. A multivariate logistic regression analysis by binary classification was used to identify independent risk factors for each mild and severe disease condition at the time of diagnosis, based on the PTBSS, as measured by the estimated odds ratio with 95% CI, including variables with \( P < 0.20 \) on univariate analysis [29]. All analyses were two-sided and statistical significance was set at \( P < 0.05 \). Statistical analyses were performed using R version 4.2.0 (R Foundation for Statistical Computing) and GraphPad Prism 9.4.

Ethical Considerations

The Institutional Review Board of Hallym University Kangdong Sacred Heart Hospital approved the study protocol (approval number 2022-08-003-001) and waived the requirement for written informed consent from the participants because of the purely observational, noninterventional study design and analysis of anonymized patient data. The data are stored by the KDCA with authority to use as surveillance data for public health and research purposes.

Results

Participants

The flowchart in Figure 1 shows the recruitment of the study population after the exclusion of patients according to the above-described criteria. The distributions of these 14,031 patients with PTB according to the PTBSS are shown in Table 1.
The largest proportion of active cases of PTB were diagnosed as grades II to III (22.2% and 21.1%, respectively, according to a PTBSS of 1-2) and the smallest proportion of cases were diagnosed as grade VII (3.0% according to a PTBSS of 6). After the reclassification of severity based on the PTBSS, 37.0%, 38.0%, and 25.0% of patients with PTB were diagnosed with mild, moderate, and severe forms, respectively (Table 2). The detailed clinical characteristics of the enrolled 14,031 patients with PTB according to the PTBSS and reclassification of severity are presented in Table S1 in Multimedia Appendix 1.

Table 1. The number and proportions of patients with pulmonary tuberculosis (PTB) in the Korea Tuberculosis Cohort classified according to the PTB Spectrum Score (PTBSS).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Grade I</th>
<th>Grade II</th>
<th>Grade III</th>
<th>Grade IV</th>
<th>Grade V</th>
<th>Grade VI</th>
<th>Grade VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients, n (%) (N=14,031)</td>
<td>419 (100.0)</td>
<td>1129 (96.7)</td>
<td>1744 (90.6)</td>
<td>1919 (81.0)</td>
<td>2136 (72.2)</td>
<td>1637 (52.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Symptoms present, n (%)</td>
<td>419 (100.0)</td>
<td>1114 (95.5)</td>
<td>1732 (89.9)</td>
<td>1445 (61.0)</td>
<td>756 (25.6)</td>
<td>173 (5.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Sputum TB PCR&lt;sup&gt;a&lt;/sup&gt; positivity, n (%)</td>
<td>0 (0.0)</td>
<td>173 (5.5)</td>
<td>756 (25.6)</td>
<td>1445 (61.0)</td>
<td>1732 (89.9)</td>
<td>1114 (95.5)</td>
<td>419 (100.0)</td>
</tr>
<tr>
<td>Sputum AFB&lt;sup&gt;b&lt;/sup&gt; smear-positive, n (%)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>201 (6.8)</td>
<td>555 (23.4)</td>
<td>1301 (56.6)</td>
<td>1094 (43.7)</td>
<td>419 (100.0)</td>
</tr>
<tr>
<td>Sputum MTB&lt;sup&gt;c&lt;/sup&gt; culture-positive, n (%)</td>
<td>0 (0.0)</td>
<td>743 (23.8)</td>
<td>1608 (54.4)</td>
<td>1818 (76.7)</td>
<td>1793 (73.1)</td>
<td>1137 (46.6)</td>
<td>419 (100.0)</td>
</tr>
<tr>
<td>Cavity in chest x-ray, n (%)</td>
<td>419 (100.0)</td>
<td>180 (5.8)</td>
<td>345 (11.7)</td>
<td>414 (17.5)</td>
<td>374 (15.4)</td>
<td>499 (19.8)</td>
<td>419 (100.0)</td>
</tr>
<tr>
<td>Bilateral disease in chest x-ray, n (%)</td>
<td>419 (100.0)</td>
<td>386 (12.4)</td>
<td>870 (29.4)</td>
<td>959 (40.5)</td>
<td>760 (30.5)</td>
<td>862 (33.9)</td>
<td>419 (100.0)</td>
</tr>
</tbody>
</table>

<sup>a</sup>TB PCR: tuberculosis polymerase chain reaction.
<sup>b</sup>AFB: acid-fast bacilli.
<sup>c</sup>MTB: Mycobacterium tuberculosis.

Table 2. Patients with pulmonary tuberculosis (PTB) classified according to disease severity (N=14,031).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTB Spectrum Score</td>
<td>0 to 1</td>
<td>2 to 3</td>
<td>4 to 6</td>
</tr>
<tr>
<td>Patients, n (%) (N=14,031)</td>
<td>5191 (37.0)</td>
<td>5328 (38.0)</td>
<td>3512 (25.0)</td>
</tr>
<tr>
<td>Symptom present, n (%)</td>
<td>1637 (31.5)</td>
<td>4055 (76.1)</td>
<td>3292 (93.7)</td>
</tr>
<tr>
<td>Sputum TB PCR&lt;sup&gt;a&lt;/sup&gt; positivity, n (%)</td>
<td>173 (3.3)</td>
<td>2202 (41.3)</td>
<td>3265 (93.0)</td>
</tr>
<tr>
<td>Sputum AFB&lt;sup&gt;b&lt;/sup&gt; smear-positive, n (%)</td>
<td>0 (0.0)</td>
<td>756 (14.2)</td>
<td>2814 (80.1)</td>
</tr>
<tr>
<td>Sputum MTB&lt;sup&gt;c&lt;/sup&gt; culture-positive, n (%)</td>
<td>743 (14.3)</td>
<td>3426 (64.3)</td>
<td>3349 (95.4)</td>
</tr>
<tr>
<td>Cavity in chest x-ray, n (%)</td>
<td>180 (3.5)</td>
<td>759 (14.2)</td>
<td>1292 (36.8)</td>
</tr>
<tr>
<td>Bilateral disease in chest x-ray, n (%)</td>
<td>386 (7.4)</td>
<td>1829 (34.3)</td>
<td>2041 (58.1)</td>
</tr>
</tbody>
</table>

<sup>a</sup>TB PCR: tuberculosis polymerase chain reaction.
<sup>b</sup>AFB: acid-fast bacilli.
<sup>c</sup>MTB: Mycobacterium tuberculosis.

Validation of the PTBSS Based on Survival Analysis

We evaluated the cumulative survival rate of KTBC patients with respect to both all-cause and PTB-related mortality, according to both the PTBSS itself and reclassified disease severity based on the PTBSS. With regard to all-cause mortality (Figure 3, Table S2 in Multimedia Appendix 1), survival rates were higher in patients with a low PTBSS and mild disease than in those with high scores and moderate-to-severe disease (Kaplan-Meier analysis, log-rank test P<.001). The overall cumulative mortality rates during anti-TB treatment for active PTB were 4.9% (grade I, score of 0), 8.0% (grade II, score of 1), 11.0% (grade III, score of 2), 14.3% (grade IV, score of 3), 14.6% (grade V, score of 4), 15.6% (grade VI, score of 5), and 14.6% (grade VII, score of 6). The overall cumulative mortality rates during anti-TB treatment for active PTB were 6.8% (mild disease), 12.4% (moderate disease), and 14.9% (severe disease).

With respect to PTB-related mortality (Figure 4, Table S3 in Multimedia Appendix 1), survival rates were higher in patients with a low PTBSS and mild disease compared with those in the high score and moderate-to-severe disease groups (P<.001). The PTB-related cumulative mortality rates during anti-TB
treatment for active PTB were 0.4% (grade I), 0.8% (grade I),
1.5% (grade III), 2.7% (grade IV), 4.2% (grade V), 7.3% (grade
VI), and 8.1% (grade VII). The overall cumulative mortality
rates during anti-TB treatment for active PTB were 0.7% (mild
disease), 2.0% (moderate disease), and 5.7% (severe disease).
Considering the correlation between the PTBSS and mortality,
the PTBSS can be considered to reflect the natural course of
PTB.

**Figure 3.** Kaplan-Meier survival probability curves of all-cause mortality for patients with active pulmonary tuberculosis (PTB) according to the Pulmonary Tuberculosis Spectrum Score (PTBSS; left curve, seven groups) and disease severity classified based on the PTBSS (right curve, three groups).

**Figure 4.** Kaplan-Meier survival probability curves of pulmonary tuberculosis (PTB)-related mortality according to the Pulmonary Tuberculosis Spectrum Score (PTBSS; left curve, seven groups) and disease severity classified based on the PTBSS (right curve, three groups).

**Time Delays According to the PTBSS in the Diagnostic Pathway**

After evaluating whether the PTBSS reflects the natural course of PTB, we checked all possible types of time delays according to the scores (Figure 5). As the PTBSS increased, the health care delay gradually decreased. However, the patient delay decreased with an increase in the PTBSS from 1 to 3 and then increased again from a score of 4 and above. The overall delay, as the sum of the health care and patient delays, gradually increased as the PTBSS increased. When including time delays in the diagnostic pathway and classifying them on a spectrum from mild to severe disease, health care and overall delays increased with increased disease severity. However, the patient delay exhibited a U-shaped pattern (Figure 5). This means that different patient-related time delays in the real-world setting, ranging from symptom onset of PTB to the first hospital visit, appeared at both extremes (ie, in the mild and severe forms) of PTB. This further means that two different approaches are required to reduce the patient delay.

Furthermore, we conducted comparative analyses to verify the impact of each factor included in the construction of the PTBSS on patient delay, health care delay, and overall delay. The results are presented in Table 3. Patients without symptoms exhibited differences in health care delay, likely indicating that symptomatic patients receive expedited testing in hospitals. Both sputum TB PCR positivity and AFB smear positivity were associated with differences in both patient and health care delays. In contrast, a positive *Mycobacterium tuberculosis* (MTB) culture in sputum did not exhibit any significant difference in either the patient or health care delay. The presence of cavitation on chest x-rays appeared to affect both patient and health care delays similarly. However, while bilateral lung involvement seemed to reduce the health care delay, it did not show a significant impact on the patient delay.
Figure 5. Patient, health care, and overall delays according to the PTBSS (top) and disease severity classified according to the PTBSS. Patient, health care, and overall time delays are presented as the median with IQR. PTB: pulmonary tuberculosis; PTBSS: Pulmonary Tuberculosis Spectrum Score.
<table>
<thead>
<tr>
<th>Scoring factor</th>
<th>Patients, n (%)</th>
<th>Patient delay, median (IQR)</th>
<th>Health care delay, median (IQR)</th>
<th>Overall delay, median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symptoms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without symptoms</td>
<td>5047 (36.0)</td>
<td>N/A</td>
<td>8.0 (2.0-22.0)</td>
<td>8.0 (2.0-22.0)</td>
</tr>
<tr>
<td>With symptoms</td>
<td>8984 (64.0)</td>
<td>16.0 (5.0-46.0)</td>
<td>4.0 (1.0-11.0)</td>
<td>28.0 (10.0-58.0)</td>
</tr>
<tr>
<td>P value</td>
<td>N/A</td>
<td>N/A</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Sputum TB PCR(^b)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>8392 (59.8)</td>
<td>15.0 (4.0-41.5)</td>
<td>8.0 (2.0-23.0)</td>
<td>21.0 (7.0-42.0)</td>
</tr>
<tr>
<td>Positive</td>
<td>5639 (40.2)</td>
<td>19.0 (5.0-50.0)</td>
<td>3.0 (1.0-7.0)</td>
<td>21.0 (8.0-44.0)</td>
</tr>
<tr>
<td>P value</td>
<td>N/A</td>
<td>N/A</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Sputum AFB(^c) smear</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>3570 (25.4)</td>
<td>14.0 (4.0-37.0)</td>
<td>7.0 (2.0-20.0)</td>
<td>19.0 (7.0-40.0)</td>
</tr>
<tr>
<td>Positive</td>
<td>10,461 (74.6)</td>
<td>24.0 (6.0-60.0)</td>
<td>3.0 (1.0-6.0)</td>
<td>28.0 (9.0-62.0)</td>
</tr>
<tr>
<td>P value</td>
<td>N/A</td>
<td>N/A</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Sputum MTB(^d) culture</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>5307 (37.8)</td>
<td>16.0 (4.0-56.0)</td>
<td>6.0 (1.0-14.0)</td>
<td>14.0 (6.0-36.0)</td>
</tr>
<tr>
<td>Positive</td>
<td>8724 (62.2)</td>
<td>16.0 (5.0-39.0)</td>
<td>5.0 (2.0-14.0)</td>
<td>24.0 (9.0-48.0)</td>
</tr>
<tr>
<td>P value</td>
<td>N/A</td>
<td>.62</td>
<td>.59</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Cavity in chest x-ray</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>11,800 (84.1)</td>
<td>15.0 (4.0-39.0)</td>
<td>6.0 (2.0-16.0)</td>
<td>19.0 (7.0-40.0)</td>
</tr>
<tr>
<td>Positive</td>
<td>2231 (15.9)</td>
<td>26.0 (7.0-61.0)</td>
<td>3.0 (1.0-8.0)</td>
<td>30.0 (11.0-62.0)</td>
</tr>
<tr>
<td>P value</td>
<td>N/A</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Extent of PTB(^e)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unilateral</td>
<td>9775 (69.7)</td>
<td>16.0 (5.0-41.0)</td>
<td>6.0 (2.0-16.0)</td>
<td>19.0 (7.0-40.0)</td>
</tr>
<tr>
<td>Bilateral</td>
<td>4256 (30.3)</td>
<td>17.0 (4.0-55.0)</td>
<td>4.0 (1.0-10.0)</td>
<td>27.0 (9.8-59.0)</td>
</tr>
<tr>
<td>P value</td>
<td>N/A</td>
<td>.12</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.
\(^b\)TB PCR: tuberculosis polymerase chain reaction.
\(^c\)AFB: acid-fast bacilli.
\(^d\)MTB: *Mycobacterium tuberculosis*.
\(^e\)PTB: pulmonary tuberculosis.

**Risk Factors for a Diagnosis of Mild or Severe Disease Based on the PTBSS**

The independent risk factors associated with mild or severe disease at the time of diagnosis were investigated (Table 4). In the multivariate analysis adjusted for potential confounding factors, age and male sex were independently associated with severe PTB in patient delay. The significant difference in symptoms such as cough and/or sputum, dyspnea, and hemoptysis may be attributed to the progression of PTB itself and may be unrelated to the late patient delay of severe PTB. However, no factor was significantly associated with patient delay for mild PTB.
Table 4. Univariable and multivariable logistic regression models for the probability of diagnosing severe or mild disease according to the Pulmonary Tuberculosis Spectrum Score.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariable logistic regression</th>
<th>Multivariable logistic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>Severe PTB&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.006 (1.004-1.008)</td>
<td>.001</td>
</tr>
<tr>
<td>Age ≥65 years</td>
<td>1.067 (0.988-1.151)</td>
<td>.10</td>
</tr>
<tr>
<td>Sex (male vs female)</td>
<td>1.313 (1.211-1.424)</td>
<td>.001</td>
</tr>
<tr>
<td>Poor economic status</td>
<td>1.364 (1.111-1.675)</td>
<td>.003</td>
</tr>
<tr>
<td>BMI</td>
<td>0.903 (0.893-0.914)</td>
<td>.001</td>
</tr>
<tr>
<td>Comorbidity of chronic kidney disease</td>
<td>0.796 (0.630-1.005)</td>
<td>.05</td>
</tr>
<tr>
<td>Symptom of cough and/or sputum</td>
<td>4.167 (3.842-4.519)</td>
<td>.001</td>
</tr>
<tr>
<td>Symptom of dyspnea</td>
<td>2.050 (1.862-2.258)</td>
<td>.001</td>
</tr>
<tr>
<td>Symptom of hemoptysis</td>
<td>1.725 (1.469-2.027)</td>
<td>.001</td>
</tr>
<tr>
<td>Symptom of fever</td>
<td>1.962 (1.766-2.181)</td>
<td>.001</td>
</tr>
<tr>
<td>Symptom of malaise</td>
<td>2.973 (2.573-3.435)</td>
<td>.001</td>
</tr>
<tr>
<td>Symptom of weight loss</td>
<td>4.009 (3.545-4.532)</td>
<td>.001</td>
</tr>
<tr>
<td>History of TB&lt;sup&gt;b&lt;/sup&gt; infection</td>
<td>0.951 (0.856-1.057)</td>
<td>.35</td>
</tr>
<tr>
<td>With EPTB&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.738 (0.640-0.851)</td>
<td>.001</td>
</tr>
<tr>
<td>Mild PTB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.985 (0.983-0.987)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age ≥65 years</td>
<td>0.654 (0.611-0.701)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BMI</td>
<td>1.076 (1.065-1.087)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Comorbidity of chronic heart disease</td>
<td>0.609 (0.518-0.717)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>PTB: pulmonary tuberculosis.
<sup>b</sup>TB: tuberculosis.
<sup>c</sup>EPTB: extrapulmonary tuberculosis.

Discussion

Principal Results

The time elapsed from the onset of PTB to diagnosis could be heterogeneous depending on personal, cultural, and health system situations. For effective TB control, the best solution needs to be identified according to each situation. We have previously recognized that a time-only approach is insufficient for the timely diagnosis of active PTB. In this study, we developed a new approach based on the disease spectrum of PTB, deviating from the perspective that has been studied based on the time delay itself.

The main findings of this study are as follows. First, the PTBSS could reflect the disease spectrum of PTB by considering the correlation of the score with mortality. Second, the pattern of time delays differed according to the PTBSS. In health care delays according to the PTBSS, greater PTB progression was associated with a shorter diagnosis period, because the condition is microbiologically easy to diagnose. However, with respect to patient delays, the change in elapsed time showed a U-shaped pattern as PTB progressed. This means that a remarkable patient delay in the real-world setting might occur at both apical ends of the spectrum (ie, in mild and severe PTB). This further means that two different approaches are required to reduce the patient delay. Third, the independent risk factors of a late visit to a medical institution as patient delay factors were age and male sex in the severe form of PTB. In contrast, there were no significant risk factors for mild PTB. Considering the natural course of PTB infection within the population, our approach could be helpful for a diagnostic strategy of active PTB involving passive case finding for the severe form and active case finding for mild forms [26] (Figure 5).

Comparison With Prior Work

A diagnostic approach for the timely diagnosis and subsequent treatment of PTB is essential to reduce ongoing transmission in the community and PTB-related morbidity and mortality [30,31]. This approach is composed of two different types of time delays: patient and health care delays [7]. Both patient and health care delays have different medical and public health implications. Patient delay is the major determining time period for the total duration of the PTB diagnostic pathway, and is associated with infectiousness due to long-term exposure to...
others in the community and poor outcomes due to disease progression [11,20]. In contrast, the health care delay is the period immediately preceding the diagnosis of PTB, which is characterized by the highest infectious state in the disease course of PTB and is associated with in-hospital transmission to health care workers or other patients [32]. To date, studies have investigated timely diagnosis and related risk factors according to the time dichotomy [8,9,11,12,14,20,33]. However, this approach has inherent limitations for general application because the time delay and risk factors can vary greatly depending on the environments within each country [7,9,10,12,13,15,19,33]. Among the types of time delays, the patient delay is greatly influenced by the culture of each country, such as the perception of disease, personal circumstances, and medical policy. Health care delay will be affected by the specialty of the health care provider and the medical resources of each country. Therefore, it was difficult to identify consistent risk factors in previous studies.

PTB can be presented as a dynamic spectrum along with pathophysiology resulting from bacterial progression and associated changes in the host response, which is distinct from the binary simple classification of active and latent TB infection [25,27]. To reach the goal of TB elimination, a promising approach that possesses the above-mentioned characteristics is essential for a timely diagnosis of PTB and includes recognizing the development of subclinical PTB from an LTBI state, early detection of active PTB from subclinical PTB, and timely differential diagnosis of severe PTB from mild PTB [24,27,28,34]. The insights and perspectives on PTB as a spectrum in a disease state are becoming increasingly accepted, leading to new diagnostic approaches for different stages of the disease spectrum. However, in clinical practice, in a real-world setting, this is considerably more complicated than what is usually expected because of the limited evidence on predictive biomarkers for disease progression within active PTB [24]. Therefore, a simple and accessible approach is needed from both clinical and public health perspectives. Thus, we attempted to design the PTBSS for clinical and public use, which is expected to be easily applied in clinical and public health practice, as the PTBSS consists of variables that are widely used in real-world settings.

We aimed to validate this proposed and devised scoring system by examining the correlation of the PTBSS and the mortality rate during the treatment of PTB. The results revealed that for all-cause mortality, there was no significant difference in the Kaplan-Meier curve and hazard ratio between high-grade scores (eg, grade IV to VII), while clear distinctions were observed when grouping into mild, moderate, and severe disease categories. Conversely, for PTB-related mortality, significant differences were observed either when considering each PTBSS grade or when classifying patients into three distinct groups. We believe that the PTBSS, as an operational definition to predict the disease spectrum of PTB, reflects PTB-related mortality relatively better than all-cause mortality, which indirectly might demonstrate its excellence in reflecting the disease spectrum of PTB. Among 1538 deaths (10.9%) recorded after a confirmed diagnosis and treatment initiation in the KTBC, 342 (2.4%) were related to PTB, while 1196 (8.5%) were due to other causes. Most PTB-related deaths were a result of respiratory failure (n=324, 94.7%), with secondary ischemic heart disease accounting for 20 deaths (5.4%) and death from massive hemoptysis for 5 cases (1.5%).

As shown in Table 3, we conducted comparative analyses to assess the impact of various factors on each delay in PTBSS construction. Asymptomatic patients experienced health care delays, suggesting that hospitals may expedite testing for those with symptoms. A positive sputum TB PCR result was linked to both patient and health care delays; health care providers might quickly decide on treatment for TB PCR–positive patients, thereby reducing health care delays. However, TB PCR positivity may also lead to longer patient delays due to late hospital visits after symptom onset. AFB smear positivity of sputum also affected both delays, likely for similar reasons. Conversely, positive MTB cultures in sputum did not influence either delay, possibly because treatment typically starts when PTB is suspected clinically rather than waiting for culture results. Cavitation on chest x-rays was associated with delays in a similar way to TB PCR and AFB smear positivity. Bilateral lung involvement seemed to shorten the health care delay but not patient delay, as health care providers might act quickly in such cases, while the patient’s perception of the disease onset—whether unilateral or bilateral—might not affect their decision to seek care. With this novel approach, we determined at what level of severity, the degree of patient delay, and the degree of health care delay that cases of PTB are diagnosed in the ROK based on the severity of PTB. This new perspective will help to change the classic perspective of active PTB and establish a more specific and realistic PTB elimination policy. We confirmed that the patient delay of PTB in the ROK was not linearly correlated, such as an increase in time with PTB progression.

The definition of patient delay, which refers to the time taken by a patient to visit a health care facility after the onset of symptoms, can lead to discrepancies in the measured period and the associated risk factors due to the inherent inaccuracy of a patient’s subjective symptom onset. Because of these limitations, numerous studies have shown that while the health care delay remains somewhat consistent, the patient delay significantly varies depending on each country in question. Factors such as each society’s medical infrastructure, economic conditions, perspectives on the disease, and language barriers are complexly intertwined, suggesting that an integrated approach that a society can employ is necessary. From this standpoint, we believe that using objective factors, rather than just relying on a patient’s subjective symptom onset, and employing the PTBSS, which can reflect the natural progression of PTB, could be helpful in understanding the factors causing TB diagnostic delays in each society.

According to our study, older individuals with PTB were more likely to be diagnosed with advanced disease, possibly because they were unaware of TB-related symptoms because of their preexisting comorbidities or because they considered the possibility of diseases other than PTB. Thus, older adults were probably more likely to visit medical facilities late after symptom onset. Moreover, even if older people visit a medical institution, the diagnosis of PTB can be delayed because of the
possibility of other diseases. Furthermore, male patients are likely to be diagnosed with advanced disease. This is consistent or inconsistent with the results of previous observational studies conducted in different countries [8,11,12,14]. Thus, more research is needed.

**Limitations**

This study has some limitations. First, this study was performed in the ROK, a country with a high level of medical resources, the highest rate of health care use among the Organisation for Economic Co-operation and Development (OECD) countries, an aging population, and a low prevalence of HIV. Therefore, our analysis may have overestimated or underestimated the prevalence of PTB. Second, the patient delay could be affected by recall bias because this factor is determined by the patient’s own symptom description. Third, we only evaluated the disease spectrum using clinical variables, without specific immunological or bacteriological data, due to a lack of laboratory-based information. Fourth, the timing data for patient delay may vary between passive and active case-finding. Active case-finding might result in a shorter patient delay compared to passive case-finding. Nevertheless, we believe that a patient delay would primarily occur in symptomatic patients in the ROK, where a nationwide health insurance system enables citizens to access medical services despite economic constraints. Koreans are more likely to seek medical care at facilities compared to citizens from other OECD countries due to the ease of accessibility to medical services. However, we acknowledge that despite these favorable circumstances, there could be instances where some symptomatic patients might not seek medical attention due to personal reasons. To gain a deeper understanding of these diagnostic situations, the KTBC made the decision, in March 2022 during a meeting with the KDCA, to investigate the reasons for hospital visits. As a result, we plan to consider these factors and incorporate relevant analyses in our future follow-up studies.

**Conclusions**

A diagnostic approach for the timely diagnosis of PTB would be improved if based on the disease spectrum rather than the traditional dichotomous time-only approach. The PTBSS, as a simple and intuitive scoring system, could facilitate clinical and public approaches for TB detection and elimination specific to the context of each country.

**Acknowledgments**

We would like to thank all the participants of this study. The nationwide Korea Tuberculosis Cohort was supported by the National Health Promotion Fund, funded by the Korea Disease Control and Prevention Agency, Republic of Korea.

**Data Availability**

The data that support the findings of this study are available upon request from the Korea Disease Control and Prevention Agency (KDCA). These data are not publicly available because of privacy or ethical restrictions. The Korea Tuberculosis Cohort data are currently being organized and there is a plan to make these data available to interested researchers in 2024, provided that prior permission is obtained from the KDCA.

**Authors' Contributions**

Y Ko, JM, JSK, and JSP conceptualized the study. Y Ko, JM, HWK, HKK, JYO, JSK, YJJ, EL, and BY curated the data. Y Ko, JM, HWK, HKK, JYO, JSK, YJJ, EL, BY, and JSP conducted the formal analysis. Y Ko, JM, and JSK designed the methodology. Y Kwon, JY, JPH, YJJ, JYK, SSL, JSP, and JSK supervised the study. Y Ko and JSP wrote the original draft. Y Ko, JM, HWK, HKK, JYO, JSK, YJJ, JSP, Y Kwon, JY, JPH, YJJ, EL, and BY reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Characteristics and time delays of participants according to Pulmonary Tuberculosis Spectrum Score (Table S1); Cox proportional hazard model for all-cause mortality (Table S2); Cox proportional hazard model for pulmonary tuberculosis–related mortality (Table S3).

[DOCX File, 50 KB - publichealth_v10i1e47422_app1.docx ]

**References**


Abbreviations

- AFB: acid-fast bacilli
- EPTB: extrapulmonary tuberculosis
- KDCA: Korea Disease Control and Prevention Agency
- KTBC: Korean Tuberculosis Cohort
- LTBI: latent tuberculosis infection
- MTB: Mycobacterium tuberculosis
- OECD: Organisation for Economic Co-operation and Development
- PCR: polymerase chain reaction
- PPM: public-private mix
- PTB: pulmonary tuberculosis
- PTBSS: Pulmonary Tuberculosis Spectrum Score
- ROK: Republic of Korea
- TB: tuberculosis
Preeclampsia Onset, Days to Delivery, and Autism Spectrum Disorders in Offspring: Clinical Birth Cohort Study

Sarah Carter¹, MA, PhD; Jane C Lin¹, MSc; Ting Chow¹, MPH; Mayra P Martinez¹, MPH; Chunyuan Qiu², MD; R Klara Feldman³, MD; Rob McConnell⁴, MD; Anny H Xiang⁴, PhD

¹Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States
²Department of Anesthesiology and Perioperative Medicine, Baldwin Park Medical Center, Kaiser Permanente Southern California, Baldwin Park, CA, United States
³Department of Obstetrics and Gynecology, Baldwin Park Medical Center, Kaiser Permanente Southern California, Baldwin Park, CA, United States
⁴Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

Abstract

Background: Maternal preeclampsia is associated with a risk of autism spectrum disorders (ASD) in offspring. However, it is unknown whether the increased ASD risk associated with preeclampsia is due to preeclampsia onset or clinical management of preeclampsia after onset, as clinical expectant management of preeclampsia allows pregnant women with this complication to remain pregnant for potentially weeks depending on the onset and severity. Identifying the risk associated with preeclampsia onset and exposure provides evidence to support the care of high-risk pregnancies and reduce adverse effects on offspring.

Objective: This study aimed to fill the knowledge gap by assessing the ASD risk in children associated with the gestational age of preeclampsia onset and the number of days from preeclampsia onset to delivery.

Methods: This retrospective population-based clinical cohort study included 364,588 mother-child pairs of singleton births between 2001 and 2014 in a large integrated health care system in Southern California. Maternal social demographic and pregnancy health data, as well as ASD diagnosis in children by the age of 5 years, were extracted from electronic medical records. Cox regression models were used to assess hazard ratios (HRs) of ASD risk in children associated with gestational age of the first occurrence of preeclampsia and the number of days from preeclampsia onset to delivery.

Results: Preeclampsia occurred in 16,205 (4.4%) out of 364,588 pregnancies; among the 16,205 pregnancies, 2727 (16.8%) first occurred at <34 weeks gestation, 4466 (27.6%) first occurred between 34 and 37 weeks, and 9012 (55.6%) first occurred at ≥37 weeks. Median days from preeclampsia onset to delivery were 4 (IQR 2,16) days, 1 (IQR 1,3) day, and 1 (IQR 0,1) day for those first occurring at <34, 34-37, and ≥37 weeks, respectively. Early preeclampsia onset was associated with greater ASD risk (P=.003); HRs were 1.62 (95% CI 1.33-1.98), 1.43 (95% CI 1.20-1.69), and 1.23 (95% CI 1.08-1.41), respectively, for onset at <34, 34-37, and ≥37 weeks, relative to the unexposed group. Within the preeclampsia group, the number of days from preeclampsia onset to delivery was not associated with ASD risk in children; the HR was 0.995 (95% CI 0.986-1.004) after adjusting for gestational age of preeclampsia onset.

Conclusions: Preeclampsia during pregnancy was associated with ASD risk in children, and the risk was greater with earlier onset. However, the number of days from first preeclampsia onset to delivery was not associated with ASD risk in children. Our study suggests that ASD risk in children associated with preeclampsia is not increased by expectant management of preeclampsia in standard clinical practice. Our results emphasize the need to identify effective approaches to preventing the onset of preeclampsia, especially during early pregnancy. Further research is needed to confirm if this finding applies across different populations and clinical settings.
autism spectrum disorders; autism; clinical management; diagnosis; expectant management; fetal exposure; fetal; management; preeclampsia; pregnancy; pregnant women; risk

**Introduction**

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by repetitive behaviors and difficulties in communication, social behavior, and sensory processing [1]. ASD is typically diagnosed in early childhood [2], with prevalence rising [3]. In 2020, it was estimated that 2.8% of children in the United States had an ASD diagnosis at 8 years old, an increase from 0.7% in 2000 and 2.3% in 2018, with rates varying by sex, race, and ethnicity [3]. ASD is characterized by a range of symptoms, behaviors, and severity; children with autistic disorders, Asperger syndrome, or pervasive developmental disorder not otherwise specified can be diagnosed with ASD [4]. The etiology of ASD is multifactorial, and genetic factors only explain a small proportion [5]. Early life exposures, including adverse maternal health conditions during pregnancy, such as obesity and diabetes, and environmental pollutants, such as near-roadway air pollution and particulate matter 2.5, are all shown to be associated with the development of ASD [5-9].

One common pregnancy complication is preeclampsia, a hypertensive condition defined by placental vascular deterioration and maternal liver or kidney dysfunction [10]. As preeclampsia is an inflammatory condition complicating 2% to 8% of pregnancies worldwide each year [11], it is important to understand its effect on offspring’s development and long-term health. Previous studies have reported associations between maternal hypertensive disorders and an increased risk of ASD in children [12-15]. Hypertensive disorders of pregnancy, which include preeclampsia, were shown to be associated with adverse childhood neurodevelopment [16,17] and a 2-fold increase in ASD risk in childhood [15]. A meta-analysis reported a higher risk of ASD in the offspring of mothers with preeclampsia than those with gestational hypertension [18]. Previous studies also reported that preeclampsia is associated with an increased risk of ASD [6,19], as well as with preterm birth, which is a risk factor for ASD [20]. However, it is unknown whether the increased ASD risk associated with preeclampsia is due to preeclampsia onset or clinical management of preeclampsia after the onset. Identifying the risk associated with preeclampsia onset and exposure is important to provide evidence supporting the care of high-risk pregnancies and to reduce adverse effects on offspring.

Evidence suggests that preeclampsia results from impaired vascularization of the placenta, with inadequate diastolic uterine arteries causing insufficient blood flow to the placenta and increased systemic maternal inflammation [21]. The primary treatment for preeclampsia is delivering the baby or managing the condition until the best time to deliver the baby [22-24]. Depending on the severity, health care providers will closely monitor symptoms, including blood pressure, platelet count, and fetal growth, as well as renal and hepatic function, and hypertensive medications can be added to control blood pressure [24]. Current preeclampsia treatment guidelines advise expectant management for preeclampsia symptoms until 34 or 37 weeks gestation [24], depending on the severity of symptoms, such that preeclampsia occurring early could be expected to continue for several weeks before delivery.

The purpose of this study is to fill the knowledge gap by examining whether the number of days from preeclampsia onset to delivery in clinical standard practice is associated with an increased risk of ASD in offspring, considering the gestational age of preeclampsia onset. This study will provide important information concerning clinical guidelines for the management of preeclampsia and the effort to minimize the impact on offspring. Data are derived from a large, representative clinical birth cohort with comprehensive electronic medical records (EMR).

**Methods**

**Study Population**

This population-based retrospective birth cohort study used mother-child pairs of singleton births at Kaiser Permanente Southern California (KPSC) hospitals between January 1, 2001, and December 31, 2014. Of 414,463 mother-child pairs, 49,875 were excluded for lack of Kaiser Permanente membership by the age of 1 year or death of the child by the age of 1 year. The final cohort for this study was 364,588 mother-child pairs. Children were followed through KPSC EMR from birth until age 5 years. The KPSC health care system includes a diverse population of 4.5 million members throughout Southern California, and member demographic data reflect that of local census tracts [25]. Maternal social demographic and pregnancy health data, as well as the child’s ASD diagnosis, were extracted from KPSC’s integrated EMR.

**Outcome: ASD Diagnosis in Children at or Before the Age of 5 Years**

The outcome of this study was whether a child had an ASD diagnosis before or at the age of 5 years and the age of the initial diagnosis in KPSC EMR. We chose the follow-up of children up to the age of 5 years because the majority of ASD cases were diagnosed by the age of 5 years, and prenatal exposure is likely to manifest its adverse effect on early developmental disorders. The diagnosis of ASD is recorded in EMR using the International Classification of Diseases (ICD) codes developed by the World Health Organization. KPSC transitioned from ICD-9 codes to ICD-10 on October 1, 2015. In EMR records dated before October 1, 2015, ASD diagnoses were identified by ICD-9 codes 299.0, 299.1, 299.8, and 299.9. After that date, ASD diagnosis was identified by ICD-10 codes F84.0, F84.3, F84.5, F84.8, and F84.9. The diagnostic codes included autistic disorders, Asperger syndrome, and pervasive developmental disorder not otherwise specified but excluded Rett syndrome or childhood disintegrative disorder. ASD diagnosis was
determined if codes were present in the EMR at 2 or more separate health visits. This method was previously validated by an expert chart review with a positive predictive value of 88% [8,26-28].

**Exposure: Maternal Preeclampsia During Pregnancy**

Maternal preeclampsia during pregnancy was identified by ICD-9 codes 642.50-642.54, 642.60-642.64, and 642.40-642.44 and included eclampsia and hemolysis, elevated liver enzymes, low platelet count (HELLP) syndrome, a serious form of preeclampsia characterized by a low platelet count and elevated liver enzymes [29]. Preeclampsia is diagnosed when blood pressure measurements are >140 systolic or 90 diastolic and when a urine test reports >0.3 g of protein in 24 hours [24,30]. Severe cases involve blood pressure >160 systolic or 110 diastolic and 1 or more of the following: persistent headache, vision changes, thrombocytopenia (platelet count <100,000/mcL), impaired liver function, progressive renal insufficiency (serum creatinine >1.1 mg/dL or doubling of serum creatinine not explained by other known renal disease), and pulmonary edema [24,30]. The date of first occurrence (ie, onset defined for this study) of preeclampsia during pregnancy was extracted, and the corresponding gestational week of preeclampsia onset was calculated by subtracting the date of last menstrual period (LMP) from the date of first occurrence recorded in the EMR. Preeclampsia onset was categorized into the following three groups: (1) <34 weeks (diagnosis between 24 weeks 0 days and 33 weeks 6 days); (2) 34-37 weeks (34 weeks 0 days to 36 weeks 6 days); and (3) ≥37 weeks (≥37 weeks 0 days). This categorization reflected known gestational ages for fetal viability at birth and clinical management guidance for preeclampsia. The duration of 24 weeks has been reported to be the lower boundary for infant survival [31,32], and expectant management guidelines aim to maintain pregnancies until 34 or 37 weeks, depending on the timing of preeclampsia onset and the severity of symptoms [30]. Expectant management duration was calculated as the number of days between the date of preeclampsia onset and the date of delivery.

**Covariates**

Covariates selected to adjust for potential confounding were maternal age, self-reported race, ethnicity, and education; prepregnancy obesity, diabetes, and smoking during pregnancy; history of comorbidity (≥1 diagnosis of heart, lung, kidney, liver disease, or cancer); offspring sex; and census tract–level household income at child’s first birthday. These covariates were potential risk factors associated with child’s ASD risk, as shown in previous studies [21,24]. The birth year was also included as a covariate to account for trends of increasing ASD prevalence over the study period [8]. Maternal obesity was defined as a prepregnancy BMI ≥30 kg/m². Maternal prepregnancy BMI was calculated using maternal height and weight recorded in EMR from the date closest to LMP, with a window of 6 months before and 3 months after LMP [33]. Diabetes during pregnancy included preexisting type 1 or type 2 diabetes and gestational diabetes mellitus diagnosed before 26 weeks, as these were previously associated with ASD risk in this study sample [26,27].

**Statistical Analyses**

Outcome variables were child’s ASD diagnosis by the age of 5 years and the age of ASD diagnosis. Exposure variables of interest were the gestational age of preeclampsia onset and days from preeclampsia onset to delivery. The gestational age of preeclampsia onset was analyzed as a continuous variable as well as a categorical variable categorized as <34 weeks, 34-37 weeks, and ≥37 weeks. Duration from preeclampsia onset to delivery was analyzed both as a continuous variable and a categorical variable, using the median number of days between onset and delivery for each onset group as the cutoff. Maternal and child characteristics by preeclampsia exposure status were reported as median and IQR for continuous variables and total number (n) and proportion (%) for categorical variables. Wilcoxon rank-sum tests and chi-square tests were used to assess differences in maternal and child characteristics between preeclampsia exposures.

Associations between preeclampsia exposure and child’s ASD risk were assessed using Cox regression models. Robust standard errors were used to correct for potential correlation between siblings born to the same mothers. Associations were quantified as hazard ratios (HRs) with 95% CI. We first assessed the risk of ASD associated with preeclampsia exposure (yes vs no), followed by assessing ASD risk associated with gestational age of onset within the preeclampsia group and comparing the risk of ASD in the <34 weeks, 34-37 weeks, and ≥37 weeks onset groups relative to the unexposed. We then examined ASD risk associated with duration from onset to delivery among those with preeclampsia exposure, adjusting for gestational weeks of the first occurrence of preeclampsia. All models adjusted for birth year, maternal age, self-reported race and ethnicity, educational qualifications, prepregnancy obesity, diabetes, smoking during pregnancy, census tract–level household income at child’s first birthday, and child’s sex. Birth year was modeled as a penalized spline to account for the nonlinear relationship between birth year and outcomes. We also assessed the roles of gestational age at delivery and birth weight as pathways to risk associated with early onset of preeclampsia and duration from onset to delivery by further adjusting these 2 variables.

Statistical significance was set at P<.05. All statistical analyses were performed in R software (version 3.6; R Foundation for Statistical Computing).

**Ethical Considerations**

This study was approved by KPSC Institutional Review Boards (review #12075), with individual participant consent waived. All data analyzed were deidentified. No compensation was offered to individual participants. There was no community involvement in the study.

**Results**

Table 1 presents cohort characteristics by maternal preeclampsia status. Of the 364,588 children included in this study, 16,205 (4.4%) were exposed to preeclampsia in utero (Table 1). The preeclampsia group had more nulliparous women than the non-preeclampsia group (8238/16,205, 50.8% vs 124,405/348,383, 35.7%). Maternal age at delivery, race,
ethnicity, and educational qualifications, as well as census-tract household income and smoking behavior during pregnancy, did not differ between the 2 groups. Larger proportions of women with preeclampsia had diabetes preexisting or diagnosed at ≤26 weeks gestation (2194/16,204, 13.5% vs 20,147/348,383, 5.8%), obesity (4242/16,205, 26% vs 53,210/348,383, 15%), and histories of comorbidities (2694/16,205, 16.6% vs 45,291/348,383, 13%) than women who did not have preeclampsia. Sex was comparable among children of mothers with and without preeclampsia. Both median gestational age at delivery (38 weeks vs 39 weeks) and median birth weight (3005 g vs 3390 g) were lower in children of mothers with preeclampsia than in children of mothers without preeclampsia.

Table 1. Cohort characteristics by preeclampsia exposure. Tests for differences in each characteristic between the 2 groups were statistically significant at *P* <.001.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No preeclampsia (n=348,383)</th>
<th>Preeclampsia (n=16,205)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maternal characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at delivery (years), median (IQR)</td>
<td>30.1 (26.0-34.1)</td>
<td>30.3 (25.6-34.8)</td>
</tr>
<tr>
<td>Parity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>124,405 (35.7)</td>
<td>8238 (50.8)</td>
</tr>
<tr>
<td>1</td>
<td>113,130 (32.5)</td>
<td>3602 (22.2)</td>
</tr>
<tr>
<td>&gt;2</td>
<td>90,515 (26)</td>
<td>3069 (18.9)</td>
</tr>
<tr>
<td>Unknown</td>
<td>20,333 (5.8)</td>
<td>1296 (8)</td>
</tr>
<tr>
<td><strong>Race and ethnicity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>88,014 (25.3)</td>
<td>3591 (22.2)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>177,877 (51.1)</td>
<td>8491 (52.4)</td>
</tr>
<tr>
<td>Other</td>
<td>82,492 (23.7)</td>
<td>4123 (25.4)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or unknown</td>
<td>129,909 (37.3)</td>
<td>6057 (37.4)</td>
</tr>
<tr>
<td>Some college</td>
<td>101,839 (29.2)</td>
<td>5240 (32.3)</td>
</tr>
<tr>
<td>College and postgraduate study</td>
<td>116,635 (33.5)</td>
<td>4908 (30.3)</td>
</tr>
<tr>
<td>Census-tract household annual income (US $), median (IQR)</td>
<td>55,700 (41,500-74,000)</td>
<td>53,500 (39,800-70,300)</td>
</tr>
<tr>
<td>Diabetes, n (%)</td>
<td>20,147 (5.8)</td>
<td>2194 (13.5)</td>
</tr>
<tr>
<td>Obesity (BMI ≥30 kg/m²), n (%)</td>
<td>53,210 (15)</td>
<td>4242 (26)</td>
</tr>
<tr>
<td>History of comorbidity, n (%)</td>
<td>45,291 (13)</td>
<td>2694 (16.6)</td>
</tr>
<tr>
<td><strong>Child characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>178,388 (51.2)</td>
<td>8510 (52.5)</td>
</tr>
<tr>
<td>Gestational age (weeks), median (IQR)</td>
<td>39 (38-40)</td>
<td>38 (36-39)</td>
</tr>
<tr>
<td>Birth weight (g), median (IQR)</td>
<td>3390 (3080-3710)</td>
<td>3005 (2450-3462)</td>
</tr>
</tbody>
</table>

*a* Maternal diabetes includes preexisting type 1 diabetes (T1D) and type 2 diabetes (T2D) and gestational diabetes mellitus (GDM) diagnosed at ≤26 weeks (no preeclampsia—T1D: 574, 0.2%; T2D: 9150, 2.6%; GDM ≤26 weeks: 10,423, 3%; preeclampsia—T1D: 137, 0.8%; T2D: 1128, 7%; GDM ≤26 weeks: 929, 5.7%)

*b* Height and weight at each clinical visit were not recorded in Kaiser Permanente Southern California electronic medical records until late 2006; therefore, maternal obesity data were missing for children born between 2001 and 2006. Maternal prepregnancy BMI was categorized as obese (BMI ≥30 kg/m²), nonobese (BMI <30 kg/m²), and unknown (including mothers with unavailable BMI information). Maternal obesity proportion excludes women with missing BMI data (no preeclampsia: missing BMI sample size=135,139; preeclampsia=6229)

*c* Maternal comorbidity was defined as ≥1 diagnosis of heart, lung, kidney, liver disease, or cancer.

Of the 16,205 pregnancies with preeclampsia, 2727 (16.8%) first occurred at <34 weeks, 4466 (27.6%) first occurred between 34 and 37 weeks, and 9012 (55.6%) first occurred at ≥37 weeks. When assessing days from first occurrence to delivery, expectedly, mothers with onset at <34 weeks had the greatest number of days from first occurrence to delivery, with 31.6% (863/2727) having ≥10 days between onset to delivery in this group. For first onset at ≥37 weeks, 78.8% (7104/9012) delivered in ≤1 day. Median days from preeclampsia onset to delivery were 4 (IQR 2,16) days, 1 (IQR 1.3) day, and 1 (IQR 0,1) day for onset at <34, 34-37, and ≥37 weeks, respectively. A total of 7194 (2.1%) of the 348,383 children were diagnosed with ASD by the age of 5 years: 2.9% (465/16,205) in the preeclampsia exposed and 1.9% (6729/348,383) in the
unexposed groups, with an HR of 1.36 (95% CI 1.23-1.49) of ASD risk for exposed versus unexposed after adjusting for birth year, maternal age, race, ethnicity, education, history of comorbidity, prepregnancy obesity, diabetes, smoking during pregnancy, offspring sex, and census-tract household income (Table 2). When considering preeclampsia exposure by gestational week of first onset, earlier preeclampsia onset was associated with greater ASD risk ($P=0.003$ for testing ASD association with gestational age of preeclampsia onset among the preeclampsia group). The HRs were 1.62 (95% CI 1.33-1.98), 1.43 (95% CI 1.20-1.69), and 1.23 (95% CI 1.08-1.41) for onset at <34, 34-37 weeks, and ≥37 weeks, respectively, relative to the unexposed group (Table 2).

### Table 2. Hazard ratios (95% CI) of child’s ASD risk associated with preeclampsia exposure and diagnosis at different gestational weeks, relative to unexposed. Adjusted for birth year, maternal age, race, ethnicity, education, comorbidities, prepregnancy obesity, diabetes, smoking during pregnancy, offspring sex, and census-tract household income (per US $10,000).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard ratio (95% CI)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preeclampsia</td>
<td>1.36 (1.23-1.49)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Gestational age at diagnosis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;34 weeks</td>
<td>1.62 (1.33-1.98)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>34-37 weeks</td>
<td>1.43 (1.20-1.69)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≥37 weeks</td>
<td>1.23 (1.08-1.41)</td>
<td>.002</td>
</tr>
</tbody>
</table>

Among those exposed to preeclampsia, the number of days from first onset to delivery was not associated with ASD risk. When analyzed as a continuous variable adjusting for preeclampsia onset, the HR of ASD risk associated with days from onset to delivery was 0.995 (95% CI 0.986-1.004; Table 3). When analyzed as a categorical variable cut at median days stratified by gestational week of preeclampsia onset, the HR associated with days above the median relative to at or below the median days for each onset group were 1.16 (95% CI 0.74-1.82) for onset at <34 weeks; 1.11 (95% CI 0.74-1.65) for onset between 34 and 37 weeks; and 1.23 (95% CI 0.85-1.78) for onset at ≥37 weeks (Table 3).

### Table 3. Hazard ratios (HRs; 95% CI) of child’s ASD risk associated with the number of days from preeclampsia diagnosis to delivery among those exposed to preeclampsia. Adjusted for birth year, maternal age, race, ethnicity, education, comorbidities, prepregnancy obesity, diabetes, smoking during pregnancy, offspring sex, and census-tract household income (per US $10,000).

<table>
<thead>
<tr>
<th>Variable</th>
<th>HR (95% CI)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time from diagnosis to delivery as a continuous variable, in days</strong></td>
<td>0.995 (0.986-1.004)</td>
<td>.28</td>
</tr>
<tr>
<td><strong>Time from diagnosis to delivery as a categorical variable stratified by gestational week of preeclampsia diagnosis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;34 weeks</td>
<td>1.05 (0.70-1.57)</td>
<td>.81</td>
</tr>
<tr>
<td>34-37 weeks</td>
<td>1.08 (0.77-1.51)</td>
<td>.65</td>
</tr>
<tr>
<td>≥37 weeks</td>
<td>1.26 (0.93-1.71)</td>
<td>.13</td>
</tr>
</tbody>
</table>

a Adjusted for gestational age at preeclampsia diagnosis in weeks, as a continuous variable.

b The hazard ratio represents the risk associated with days from diagnosis to delivery above the median relative to equal or below the median distribution within each group (<34 weeks: median 4, IQR 2.16 days); (34-37 weeks: median 1, IQR 1.3 day); (≥37 weeks: median 1, IQR 0.1 day).

Early preeclampsia diagnosis was positively correlated with early gestational age at delivery ($r=0.90$) and lower birth weight ($r=-0.70$), but the number of days from onset to delivery was not correlated with gestational age at delivery ($r=-0.01$) or birth weight ($r=-0.005$). Further adjusting for gestational age at delivery and birth weight reduced HRs to 1.19 (95% CI 0.97-1.48) for onset at <34 weeks and 1.27 (1.07-1.50) for onset between 34 and 37 weeks but did not change the HR for preeclampsia onset ≥37 weeks (1.24, 95% CI 1.08-1.41). The number of days from onset to delivery remained unassociated with ASD risk in offspring after adjustment for gestational age at delivery and birth weight.

### Discussion

#### Overview

In this large sample and multiethnic clinical cohort study, children exposed to preeclampsia in utero were at higher risk of ASD than children who were not exposed, with a greater risk for children exposed to preeclampsia earlier in pregnancy. However, among children exposed to preeclampsia, there were no significant associations of risk of ASD with the number of days from first onset to delivery. This was analyzed both as a continuous variable adjusting for gestational age of preeclampsia first onset and as a categorical variable cut at the median number of days and stratified by gestational age of preeclampsia first occurrence. These results suggest that a child’s risk of ASD associated with preeclampsia is not increased by expectant management of preeclampsia in standard clinical practice.

https://publichealth.jmir.org/2024/1/e47396
To our knowledge, this study is the first to assess whether the reported ASD risk associated with maternal preeclampsia was due to preeclampsia onset or clinical management of preeclampsia during pregnancy. Current clinical management of preeclampsia advises maintenance of pregnancies until 34 or 37 weeks gestation, depending on the week of diagnosis and preeclampsia severity [24]. Expectant management of preeclampsia requires balancing maternal medical needs with intrapartum developmental requirements and reduction of infant comorbidities associated with preterm birth [34,35]. Planned preterm delivery of pregnancies complicated by preeclampsia improves maternal outcomes but increases the risk of neonatal admissions for prematurity [36]. Expectant management of moderate preeclampsia before 37 weeks gestation has been reported to extend pregnancies without increasing neonatal comorbidities; however, studies have reported the risk of stillbirth [37] and neonatal mortality [38] associated with expectant management for severe preeclampsia diagnosed before 34 weeks gestation. The results presented in our study demonstrate that early onset of preeclampsia is associated with a greater risk of ASD in offspring; however, clinical management for preeclampsia does not add additional risk, potentially eliminating one comorbidity of concern associated with preeclampsia exposure and expectant management.

Prenatal preeclampsia may affect child neurodevelopment by altering the maternal environment during fetal maturation. Preeclampsia has been observed to influence maternal immune activation [39], increasing circulation of maternal proinflammatory cytokines, and contributing to higher levels of oxidative stress, which have been associated with divergent neurodevelopment in children [40]. Studies have reported differences in the etiology of early and late-onset preeclampsia [41,42], as well as varying risks of adverse outcomes associated with the timing of exposure [43,44]. Early-onset preeclampsia may be driven by placental dysfunction [45], which occurs when a placental abnormality restricts blood flow, potentially due to suppression of estrogen-related receptor-gamma leading to vascular abnormalities [46]. Late-onset preeclampsia could be primarily a maternal hypertensive condition [45]. Placental inflammation and vascular dysfunction have been linked to an increased risk of ASD in children [47,48]. Therefore, differences in ASD risk by gestational week of onset may be explained by the influence of preeclampsia on specific stages of fetal neurodevelopment [19].

In our study, earlier onset of preeclampsia had larger HRs than onset in later pregnancy, which is consistent with results from a study of ASD risk and preeclampsia, which reported an increased risk of ASD in children exposed to preeclampsia earlier in pregnancy [19]. Adjustment for gestational age at delivery and birth weight attenuated the risk of ASD after early preeclampsia in our study, suggesting that there may be different mechanisms underlying the risk of ASD by the timing of preeclampsia exposure. Exposure to preeclampsia has been reported to be associated with an increased risk of ASD in offspring [49]. Preeclampsia is also associated with preterm birth, another risk factor for ASD [20]. Our results are consistent with previous results that prematurity and low birth weight are risk factors for ASD [6,20] and extend them to show that early preeclampsia is one of the root causes for prematurity, low birth weight, and associated ASD risk.

Our results concerning the increased risk of ASD associated with maternal preeclampsia during pregnancy are consistent with previous findings. However, the novel finding in our study is that the number of days between preeclampsia’s first occurrence and delivery was not associated with ASD risk in offspring after taking the gestational age of preeclampsia onset into account. Thus, the results of our study suggest that clinical expectant management of preeclampsia after diagnosis does not increase the risk of ASD in offspring. As this is the first study to report these results, future research is needed to determine whether this finding is consistent in other populations and to assess whether these associations vary by type of preeclampsia treatment or severity of illness. Our results provide evidence demonstrating the need to identify effective approaches to prevent the onset of preeclampsia, especially during early pregnancy, to mitigate risk not only to mothers but also to offspring.

**Strengths and Limitations**

A strength of this study is its large, clinical, and longitudinal birth cohort with characteristics reflecting census tract–level social and demographic information of Southern California. Comprehensive EMR data with detailed pregnancy history and dates allowed us to assess preeclampsia by gestational week of diagnosis and duration from diagnosis to delivery, and to adjust for relevant covariates. The continuity of care at KPSC minimizes the risk of ascertainment bias in exposures and outcomes. We think the assessment of expectant management of preeclampsia on ASD risk is novel.

This observational study has some limitations. The results presented here do not establish a causal link between preeclampsia exposure and ASD risk. The severity of preeclampsia was not explicitly considered due to the lack of a clear definition of severity in EMR. However, diagnosis to delivery periods allowed some inference about severity, as severe cases were likely to be delivered more quickly than moderate cases [24]. Genetic information was unavailable; therefore, we were unable to control for genetic contributions to ASD risk. There may be other perinatal risk factors or postnatal environmental exposures not adjusted for in these analyses.

**Conclusion**

In this population-based retrospective clinical birth cohort study, exposure to preeclampsia in utero was associated with an increased risk of ASD in offspring, with a greater risk for children of mothers with preeclampsia occurring earlier during pregnancy. However, among children of mothers with preeclampsia, the number of days between preeclampsia diagnosis and delivery was not associated with increased ASD risk. Our study suggests that clinical management of pregnancies with preeclampsia does not increase the risk of ASD in offspring. Future research into the prevention of preeclampsia is still needed.
Acknowledgments
The authors thank patients of Kaiser Permanente Southern California for helping us improve care using information collected through our integrated electronic health record systems.

This study was supported in part by the National Institutes of Health (R01ES029963) and by Kaiser Permanente Southern California Direct Community Benefit Funds. Funding agencies had no role in the study design, data analysis or interpretation, or manuscript preparation or approval.

Data Availability
The data set analyzed in this study is not publicly available as it was drawn from electronic medical records.

Authors' Contributions
AHX obtained funding, acquired data, was responsible for the study concept and design, analyzed and interpreted data, revised the manuscript for important intellectual content, was a guarantor of this work, and took responsibility for the integrity of the data and the accuracy of the data analysis. SC was responsible for the study concept and design, analyzed and interpreted data, drafted the manuscript, revised the manuscript for important intellectual content, and was a guarantor of this work and, as such, had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. JCL, RKF, CQ, RM, and TC analyzed and interpreted data and revised the manuscript for important intellectual content. MPM acquired, analyzed, and interpreted data, and revised the manuscript for important intellectual content.

All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Conflicts of Interest
None declared.

References


32. Royal College of Obstetricians & Gynaecologists. Perinatal management of pregnant women at the threshold of infant viability (the obstetric perspective). Royal College of Obstetricians & Gynaecologists 2014 [FREE Full text]


Abbreviations

- ASD: autism spectrum disorder
- EMR: electronic medical record
- HELLP: hemolysis, elevated liver enzymes, low platelet count
- HR: hazard ratio
- ICD: International Classification of Diseases
- KPSC: Kaiser Permanente Southern California
- LMP: last menstrual period
Mutual Associations of Exposure to Ambient Air Pollutants in the First 1000 Days of Life With Asthma/Wheezing in Children: Prospective Cohort Study in Guangzhou, China

Fenglin Tian\textsuperscript{1,2*}, MPH; Xinqi Zhong\textsuperscript{3*}, PhD; Yufeng Ye\textsuperscript{4}, MD; Xiaohan Liu\textsuperscript{1,2}, MPH; Guanhao He\textsuperscript{1,2}, PhD; Cuiling Wu\textsuperscript{1,2}, MPH; Zhiqing Chen\textsuperscript{1,2}, MPH; Qijiong Zhu\textsuperscript{1,2}, MPH; Siwen Yu\textsuperscript{1,2}, MPH; Jingjie Fan\textsuperscript{5}, MD; Huan Yao\textsuperscript{1,2}, MPH; Wenjun Ma\textsuperscript{1,2}, PhD; Xiaomei Dong\textsuperscript{1,2}, PhD; Tao Liu\textsuperscript{1,2}, PhD

\textsuperscript{1}Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China
\textsuperscript{2}China Greater Bay Area Research Center of Environmental Health, School of Medicine, Jinan University, Guangzhou, China
\textsuperscript{3}Department of Neonatology, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
\textsuperscript{4}Guangzhou Panyu Central Hospital, Guangzhou, China
\textsuperscript{5}Department of Prevention and Health Care, Shenzhen Maternity and Child Healthcare Hospital, Southern Medical University, Shenzhen, China

* these authors contributed equally

Corresponding Author:
Tao Liu, PhD
Department of Public Health and Preventive Medicine
School of Medicine, Jinan University
Number 601, West Huangpu Avenue
Guangzhou, 510632, Guangdong
China
Phone: 86 18927588462
Email: gztt_2002@163.com

Abstract

Background: The first 1000 days of life, encompassing pregnancy and the first 2 years after birth, represent a critical period for human health development. Despite this significance, there has been limited research into the associations between mixed exposure to air pollutants during this period and the development of asthma/wheezing in children. Furthermore, the finer sensitivity window of exposure during this crucial developmental phase remains unclear.

Objective: This study aims to assess the relationships between prenatal and postnatal exposures to various ambient air pollutants (particulate matter 2.5 [PM\textsubscript{2.5}], carbon monoxide [CO], sulfur dioxide [SO\textsubscript{2}], nitrogen dioxide [NO\textsubscript{2}], and ozone [O\textsubscript{3}]) and the incidence of childhood asthma/wheezing. In addition, we aimed to pinpoint the potential sensitivity window during which air pollution exerts its effects.

Methods: We conducted a prospective birth cohort study wherein pregnant women were recruited during early pregnancy and followed up along with their children. Information regarding maternal and child characteristics was collected through questionnaires during each round of investigation. Diagnosis of asthma/wheezing was obtained from children’s medical records. In addition, maternal and child exposures to air pollutants (PM\textsubscript{2.5}, CO, SO\textsubscript{2}, NO\textsubscript{2}, and O\textsubscript{3}) were evaluated using a spatiotemporal land use regression model. To estimate the mutual associations of exposure to mixed air pollutants with the risk of asthma/wheezing in children, we used the quantile g-computation model.

Results: In our study cohort of 3725 children, 392 (10.52\%) were diagnosed with asthma/wheezing. After the follow-up period, the mean age of the children was 3.2 (SD 0.8) years, and a total of 14,982 person-years were successfully followed up for all study participants. We found that each quartile increase in exposure to mixed air pollutants (PM\textsubscript{2.5}, CO, SO\textsubscript{2}, NO\textsubscript{2}, and O\textsubscript{3}) during the second trimester of pregnancy was associated with an adjusted hazard ratio (HR) of 1.24 (95\% CI 1.04-1.47). Notably, CO made the largest positive contribution (64.28\%) to the mutual effect. After categorizing the exposure according to the embryonic respiratory development stages, we observed that each additional quartile of mixed exposure to air pollutants during the pseudoglandular and canalicular stages was associated with HRs of 1.24 (95\% CI 1.03-1.51) and 1.23 (95\% CI 1.01-1.51), respectively. Moreover, for the first year and first 2 years after birth, each quartile increment of exposure to mixed air pollutants...
was associated with HRs of 1.65 (95% CI 1.30-2.10) and 2.53 (95% CI 2.16-2.97), respectively. Notably, SO2 made the largest positive contribution in both phases, accounting for 50.30% and 74.70% of the association, respectively.

**Conclusions:** Exposure to elevated levels of mixed air pollutants during the first 1000 days of life appears to elevate the risk of childhood asthma/wheezing. Specifically, the second trimester, especially during the pseudoglandular and canalicular stages, and the initial 2 years after birth emerge as crucial susceptibility windows.

**Trial Registration:** Chinese Clinical Trial Registry ChiCTR-ROC-17013496; https://tinyurl.com/2ctufw8n

**KEYWORDS**
pregnancy; air pollution; asthma; wheezing; birth cohort; children

### Introduction

Asthma is one of the major respiratory diseases. It is also the most common noncommunicable disease in children and imposes a huge economic and disease burden worldwide [1]. Asthma is estimated to affect around 14% of children worldwide, with its prevalence reported to be on the rise [2]. It is well-known that childhood asthma increases the burden on society in terms of disruption to children’s lives, reduced physical ability, increased caregiver strain, and various direct medical costs. Wheezing during early childhood is frequently regarded as the primary symptom linked to asthma later in life [3]. Research has shown that the intricate interplay of environmental exposures and genetic susceptibility can contribute to the onset of wheezing and asthma [4].

The first 1000 days of life, encompassing pregnancy and the initial 2 years after birth, represent a crucial period for human health development and interventions [5]. Lung and airway development commences between the 4th and 7th weeks of pregnancy, reaching the alveolar stage by 36 weeks. During pregnancy, fetal cells exhibit more rapid replication and differentiation compared with mature cells, making them highly responsive to external signals [6]. Consequently, they are particularly susceptible to external exposure events. Air pollution, for instance, can disrupt alveolarization, leading to compromised lung development and function postnatally [7,8].

The lungs undergo growth from conception through early adulthood, with the prenatal and early postnatal phases being particularly critical [9]. There is evidence suggesting that exposure to air pollution during early life can trigger the onset of asthma or wheezing and exacerbate preexisting conditions [5,10,11]. A review, which has consolidated the link between exposure to air pollutants throughout the first 1000 days of life and the onset of childhood asthma/wheezing, was associated with HRs of 1.65 (95% CI 1.30-2.10) and 2.53 (95% CI 2.16-2.97), respectively. Notably, SO2 made the largest positive contribution in both phases, accounting for 50.30% and 74.70% of the association, respectively.

**Conclusions:** Exposure to elevated levels of mixed air pollutants during the first 1000 days of life appears to elevate the risk of childhood asthma/wheezing. Specifically, the second trimester, especially during the pseudoglandular and canalicular stages, and the initial 2 years after birth emerge as crucial susceptibility windows.

**Trial Registration:** Chinese Clinical Trial Registry ChiCTR-ROC-17013496; https://tinyurl.com/2ctufw8n

**KEYWORDS**
pregnancy; air pollution; asthma; wheezing; birth cohort; children

### Methods

**Study Settings and Participants**
All participant data were sourced from the Prenatal Environments and Offspring Health (PEOH) cohort study.
conducted at Panyu Central Hospital in Guangzhou, China. Detailed descriptions of this cohort study have been provided in previous publications [15,17,18]. Further information regarding the population within the hospital’s catchment area, along with some basic demographic details concerning the broader populations of Guangzhou and China, can be found in Multimedia Appendix 1. Pregnant women were enrolled in the study at the antenatal care unit, adhering to the following inclusion criteria: (1) gestational age <13 weeks; (2) between 18 and 50 years of age; and (3) absence of significant medical conditions, including hyperthyroidism, hypertension, chronic kidney disease, tuberculosis, and mental illness. All eligible women underwent face-to-face interviews to collect baseline information, and they were subsequently followed up during hospital delivery. Furthermore, we administered follow-up surveys on the children during their hospital visits. Throughout the follow-up period, we excluded cases involving multiple pregnancies and study participants with missing key variables. Recruitment for our study commenced in January 2016 and continued until July 31, 2020, when the follow-up concluded.

**Baseline Investigation**

The baseline survey for the study was initiated in January 2016 and concluded in December 2017. A total of 4928 pregnant women were successfully recruited and their personal profiles were established. The collected information encompassed their demographic characteristics, lifestyle behaviors, changes in home address, living environment at home, work environment, activity patterns, medical history, diet during pregnancy, and antenatal care records.

**Follow-Up Investigation**

As illustrated in Figure 1, of the 4928 pregnant women enrolled in the baseline survey, 4279 were effectively followed up during their hospitalization for delivery. Throughout this period, maternal information was gathered via a follow-up questionnaire, while neonatal birth records were extracted from maternal medical records. Following the exclusion of cases involving multiple births (n=79) and those with missing key variables (n=10), a total of 4190 mother-child pairs were included in the follow-up. All children who attended child health clinics, pediatric outpatient clinics, or emergency departments were tracked for follow-up. Clinical symptoms, diagnostic disorders, and anthropometric data were extracted from their medical records, and questionnaires were administered to their parents or accompanying individuals. Throughout the follow-up process, a total of 465 infants were not successfully followed up. The final follow-up survey was concluded by July 31, 2020, with a total of 3725 children enrolled in the study (Figure 1).

**Figure 1.** Flow chart of the progress of the research topic.

Major Health Outcomes

The health outcomes in this study were delineated in accordance with the 10th edition of the International Classification of Diseases (ICD-10), encompassing asthma (J45) and wheezing (R06). Asthma was characterized by recurrent and variable clinical manifestations, such as wheezing, chest tightness, persistent cough, shortness of breath, and restricted expiratory airflow. Our health outcomes also incorporated wheezing because of the challenges in diagnosing early childhood asthma, which led to a relatively low number of children diagnosed with asthma (n=17). Conversely, early childhood wheezing is frequently regarded as the primary symptom linked to asthma in later life and holds significant prognostic value for the early detection of asthma. Consequently, we ultimately delineated the health outcome as the first admission to the hospital for wheezing.
asthma or wheezing, using the time of initial occurrence of asthma/wheezing as the onset time. For cases where asthma or wheezing did not manifest, the follow-up continued until July 31, 2020. Eventually, to enhance statistical power, we combined asthma and wheezing (n=392) as a unified outcome.

**Ethical Considerations**

The PEOH study received approval from the Ethics Committee of the Guangdong Center for Disease Control and Prevention under the reference number W96027E-2015001. Furthermore, it was registered in the Chinese Clinical Trial Registry with the registration number ChiCTR-ROC-17013496. This analysis was approved by the Ethics Committee of the Faculty of Medicine of Jinan University under the approval number JNUKY-2023-0075. During the establishment of the original cohort, comprehensive details about the study were provided to the recruited women, and informed consent was obtained from them through signed documents under conditions of autonomy. When accessing the data set for this research, it was entirely deidentified, and there was no feasible way to link the data to the women, either through a key in the coding system or otherwise. Stringent measures have been implemented to safeguard the personal information of the women.

**Exposure Assessment of Air Pollution**

In this study, we incorporated 5 major ambient air pollutants (PM$_{2.5}$, SO$_2$, NO$_2$, CO, and O$_3$). To evaluate each individual’s weekly exposure to these air pollutants, we used a spatiotemporal land use regression (ST-LUR) model, which has been extensively described in our team’s previous publications [18-20]. In particular, we collected spatiotemporal data from air quality monitoring stations located throughout Guangdong Province to establish the ST-LUR model. We used an inverse distance weighting method to extract the weekly visibility of each address. This method relied on the latitude and longitude information to calculate the proximity between each participant’s home and work addresses and the air quality monitoring stations. Population density, road length, and land use–type data were obtained for each air quality monitoring station address within a radius of 1300 m. Predictor variables for both residential and work addresses of each participant were incorporated into the model. Subsequently, we used these variables to predict the weekly average concentrations of air pollutants at each address throughout the pregnancy period. To evaluate mothers’ exposure to air pollutants during pregnancy, we forecasted weekly pollutant concentrations at their individual residences and workplaces. Subsequently, we multiplied these concentrations at the home address for children aged 0-2 years. To identify sensitive exposure time windows during childhood, considering the considerable changes in children’s activity patterns as they grow, we calculated air pollutant concentrations for 2 distinct periods: from birth to 12 months (the first year of life) and from birth to 24 months (the first 2 years of life) [25].

**Covariates**

Upon reviewing the literature, the following variables were identified as covariates: ambient temperature, maternal age, parity, gravidity, maternal occupation, annual per capita income, passive smoking, history of asthma, preterm birth, gestational diabetes, season of conception, and maternal diet (Multimedia Appendix 2). We used variance inflation factors (VIFs) to detect collinearity among covariates both between individual pollutants and among covariates themselves. A maximum VIF exceeding 10 often suggests that multiple correlations significantly influence least squares estimates. The findings revealed that when all covariates, including temperature, were incorporated, the VIF for PM$_{2.5}$ exceeded 10 in all 3 trimesters, as well as during the saccular and alveolar stages. However, when temperature was excluded from the covariates, all VIFs for PM$_{2.5}$ were below 10 except for the third trimester (Multimedia Appendices 3-5).

**Statistical Analysis**

We conducted a comparison of characteristics between the case and control groups using a chi-square test. Pearson correlation was applied to examine correlations between air pollutants and temperature (Multimedia Appendix 4). The mutual effects of air pollution mixtures on asthma/wheezing in children were analyzed using the quantile g-computation (QG-comp) model. QG-comp is used for the impact analysis of exposure mixtures, where it evaluates the exposure as a whole rather than as individual components in its calculations. The QG-comp model combines the simplicity of weighted quantile sum regression inference with the flexibility of the g-calculation method for causal effect estimation. Unlike traditional methods, QG-comp does not necessitate that the direction of the effect is consistent between the exposure variable and the outcome. Moreover, it can capture the effects of all pollutant exposures, 1 quartile increase at a time [26]. The model used the Cox proportional hazards model as the base model and established quartiles of...
simultaneous increases in all air pollutant concentrations. Hazard ratios (HRs) along with 95% CIs were reported. In addition, the positive or negative weight of each pollutant was documented in each QG-comp model.

In the sensitivity analyses, we narrowed our focus to children with clear information on the feeding method. Furthermore, we adjusted for children’s feeding methods in both prenatal and postnatal QG-comp models for this subgroup to mitigate the potential confounding effect of the feeding method on childhood asthma/wheezing. To assess whether the COVID-19 outbreak in early 2020 and subsequent public health interventions targeting COVID-19 might introduce confounding factors, we conducted sensitivity analyses by limiting the data to the period before the COVID-19 outbreak. In addition, in our sensitivity analyses, we adjusted for air pollutants in the preceding periods in the model for that specific period. This adjustment was made to account for the cumulative effects of early developmental exposure.

All analyses were performed using R 4.2.2 (R Development Core Team 2019 [27]). All statistical tests were conducted using a 2-sided approach.

Results

Characteristics of Study Participants

Of the 3725 children included in this study, 392 (10.52%) were diagnosed with asthma/wheezing, with a mean age of onset of 1.48 (SD 0.89) years. At the conclusion of the follow-up period, the mean age of the children was 3.2 (SD 0.8) years, and a total of 14,982 person-years were successfully followed up for all study participants. These included 1984 (53.26%) boys and 1741 (46.74%) girls, with a prevalence of asthma/wheezing of 13.16% (261/1984) in boys compared with 7.52% (131/1741) in girls. The majority of mothers were engaged in commercial activities (2008/3725, 53.91%). Demographic information, the number of asthma cases, and prevalence rates stratified by covariates are presented in Table 1.
Table 1. General characteristics of the study children.

<table>
<thead>
<tr>
<th>Characteristics (variables)</th>
<th>Total (n=3725)</th>
<th>Asthma/wheezing</th>
<th>Chi-square/t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yes (n=392)</td>
<td>No (n=3333)</td>
<td></td>
</tr>
<tr>
<td>Sex of children, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>1984 (53.26)</td>
<td>261 (13.16)</td>
<td>1723 (86.84)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Girl</td>
<td>1741 (46.74)</td>
<td>131 (7.52)</td>
<td>1610 (92.48)</td>
<td></td>
</tr>
<tr>
<td>Maternal age (years), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>193 (5.18)</td>
<td>22 (11.40)</td>
<td>171 (88.60)</td>
<td>.07</td>
</tr>
<tr>
<td>26-30</td>
<td>1253 (33.64)</td>
<td>124 (9.90)</td>
<td>1129 (90.10)</td>
<td></td>
</tr>
<tr>
<td>31-35</td>
<td>1304 (35.01)</td>
<td>159 (12.19)</td>
<td>1145 (87.81)</td>
<td></td>
</tr>
<tr>
<td>&gt;35</td>
<td>975 (26.17)</td>
<td>87 (8.92)</td>
<td>888 (91.08)</td>
<td></td>
</tr>
<tr>
<td>Maternal occupation, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Manual worker</td>
<td>183 (4.91)</td>
<td>22 (12.02)</td>
<td>161 (87.98)</td>
<td></td>
</tr>
<tr>
<td>Government official and clerk</td>
<td>78 (2.09)</td>
<td>12 (15.38)</td>
<td>66 (88.62)</td>
<td></td>
</tr>
<tr>
<td>Housewife</td>
<td>372 (9.99)</td>
<td>43 (11.56)</td>
<td>329 (88.44)</td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>278 (7.46)</td>
<td>43 (15.47)</td>
<td>235 (84.53)</td>
<td></td>
</tr>
<tr>
<td>Technician</td>
<td>689 (18.50)</td>
<td>65 (9.43)</td>
<td>624 (90.57)</td>
<td></td>
</tr>
<tr>
<td>Business (commercial activities)</td>
<td>2008 (53.91)</td>
<td>193 (9.61)</td>
<td>1815 (90.39)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>117 (3.14)</td>
<td>14 (11.97)</td>
<td>103 (88.03)</td>
<td></td>
</tr>
<tr>
<td>Parity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.28</td>
</tr>
<tr>
<td>1</td>
<td>1132 (30.39)</td>
<td>96 (8.48)</td>
<td>1036 (91.52)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2220 (59.60)</td>
<td>249 (11.22)</td>
<td>1971 (88.78)</td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>373 (10.01)</td>
<td>47 (12.60)</td>
<td>326 (87.40)</td>
<td></td>
</tr>
<tr>
<td>Gravidity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.43c</td>
</tr>
<tr>
<td>1</td>
<td>1104 (29.64)</td>
<td>101 (9.15)</td>
<td>1003 (90.85)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1432 (38.44)</td>
<td>163 (11.38)</td>
<td>1269 (88.62)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>792 (21.26)</td>
<td>82 (10.35)</td>
<td>710 (89.65)</td>
<td></td>
</tr>
<tr>
<td>≥4</td>
<td>397 (10.66)</td>
<td>46 (11.59)</td>
<td>351 (88.41)</td>
<td></td>
</tr>
<tr>
<td>Yearly income per capita (×1000 Yuana), n (%)</td>
<td>N/Ab</td>
<td></td>
<td></td>
<td>.43c</td>
</tr>
<tr>
<td>&lt;30</td>
<td>193 (5.18)</td>
<td>22 (11.40)</td>
<td>171 (88.60)</td>
<td></td>
</tr>
<tr>
<td>30~</td>
<td>2228 (59.81)</td>
<td>248 (11.13)</td>
<td>1980 (88.87)</td>
<td></td>
</tr>
<tr>
<td>100~</td>
<td>1088 (29.21)</td>
<td>100 (9.19)</td>
<td>988 (90.81)</td>
<td></td>
</tr>
<tr>
<td>≥200~</td>
<td>178 (4.78)</td>
<td>17 (9.55)</td>
<td>161 (90.45)</td>
<td></td>
</tr>
<tr>
<td>Refused to answer/missing</td>
<td>38 (1.02)</td>
<td>5 (13.16)</td>
<td>33 (86.84)</td>
<td></td>
</tr>
<tr>
<td>Feeding method, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Artificial feeding</td>
<td>400 (10.74)</td>
<td>60 (15.00)</td>
<td>340 (85.00)</td>
<td></td>
</tr>
<tr>
<td>Breastfeeding</td>
<td>449 (12.05)</td>
<td>45 (10.02)</td>
<td>404 (89.98)</td>
<td></td>
</tr>
<tr>
<td>Mixed feeding</td>
<td>431 (11.57)</td>
<td>32 (7.42)</td>
<td>399 (92.58)</td>
<td></td>
</tr>
<tr>
<td>Refused to answer/missing</td>
<td>2445 (65.64)</td>
<td>255 (10.43)</td>
<td>2190 (89.57)</td>
<td></td>
</tr>
<tr>
<td>Passive smoking, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.49</td>
</tr>
<tr>
<td>Yes</td>
<td>2599 (69.77)</td>
<td>267 (10.27)</td>
<td>2332 (89.73)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1126 (30.23)</td>
<td>125 (11.10)</td>
<td>1001 (88.90)</td>
<td></td>
</tr>
<tr>
<td>Vegetable consumption (time/week), n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.29</td>
</tr>
</tbody>
</table>
## Associations Between Maternal Ambient Pollutant Exposure and Childhood Wheezing Risk

Table 2 displays the risk of asthma/wheezing in children exposed to mixed air pollutants during pregnancy. In the adjustment model, each quartile increase in exposure to mixed air pollutants during the second trimester was linked to an adjusted HR of 1.24 (95% CI 1.04-1.47), with CO contributing the most positively (64.28%) to the joint effect (Figure 2). Following staging by lung development, each additional quartile of exposure to mixed air pollutants during the pseudoglandular and canalicular stages was associated with HRs of 1.24 (95% CI 1.03-1.51) and 1.23 (95% CI 1.01-1.51), respectively. During the pseudoglandular stage, PM$_{2.5}$ (50.98%) and CO (48.92%) made the greatest positive contributions, while CO made the largest positive contribution during the canalicular stage (65.50%; Figure 2).

### Table 2: Risk of Childhood Wheezing/Immunological Atopy by Maternal Ambient Air Pollution Exposure (n=3,725)

<table>
<thead>
<tr>
<th>Characteristics (variables)</th>
<th>Total (n=3725)</th>
<th>Asthma/wheezing (n=392)</th>
<th>Chi-square/t test (d/f)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n=392)</td>
<td>No (n=3333)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categorical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age of children at end of follow-up (years), mean (SD)</td>
<td>39.75 (2.13)</td>
<td>39.65 (1.61)</td>
<td>39.74 (2.08)</td>
<td>1.12d (3273)</td>
</tr>
<tr>
<td>Gestation age of children at birth (ie, gestational weeks), mean (SD)</td>
<td>36.5 (1.61)</td>
<td>39.75 (2.13)</td>
<td>39.74 (2.08)</td>
<td>1.12d (3273)</td>
</tr>
<tr>
<td>Conception season, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warm (May to October)</td>
<td>1775 (47.65)</td>
<td>170 (9.58)</td>
<td>1605 (90.42)</td>
<td>.08</td>
</tr>
<tr>
<td>Cold (November to April)</td>
<td>1950 (52.35)</td>
<td>222 (11.38)</td>
<td>1728 (88.62)</td>
<td></td>
</tr>
<tr>
<td>History of maternal asthma, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3546 (95.19)</td>
<td>371 (10.46)</td>
<td>3175 (89.54)</td>
<td>.68</td>
</tr>
<tr>
<td>Yes</td>
<td>179 (4.81)</td>
<td>21 (11.73)</td>
<td>158 (88.27)</td>
<td></td>
</tr>
<tr>
<td>History of gestational diabetes mellitus, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.59c</td>
</tr>
<tr>
<td>No</td>
<td>3516 (94.39)</td>
<td>368 (10.47)</td>
<td>3148 (89.53)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>197 (5.29)</td>
<td>22 (11.17)</td>
<td>175 (88.83)</td>
<td></td>
</tr>
<tr>
<td>Refused to answer/missing</td>
<td>12 (0.32)</td>
<td>2 (16.67)</td>
<td>10 (83.33)</td>
<td></td>
</tr>
<tr>
<td>Fruit consumption (time/week), n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5</td>
<td>271 (7.28)</td>
<td>24 (8.86)</td>
<td>247 (91.14)</td>
<td></td>
</tr>
<tr>
<td>6–7</td>
<td>3041 (81.64)</td>
<td>334 (10.98)</td>
<td>2707 (89.02)</td>
<td></td>
</tr>
<tr>
<td>≥8</td>
<td>343 (9.21)</td>
<td>26 (7.58)</td>
<td>317 (92.42)</td>
<td></td>
</tr>
<tr>
<td>Refused to answer/missing</td>
<td>70 (1.88)</td>
<td>8 (11.43)</td>
<td>62 (88.57)</td>
<td></td>
</tr>
<tr>
<td>Premature, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.17</td>
</tr>
<tr>
<td>No</td>
<td>3520 (94.50)</td>
<td>364 (10.34)</td>
<td>3156 (89.66)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>205 (5.50)</td>
<td>28 (13.66)</td>
<td>177 (86.34)</td>
<td></td>
</tr>
<tr>
<td>Refused to answer/missing</td>
<td>12 (0.32)</td>
<td>2 (16.67)</td>
<td>10 (83.33)</td>
<td></td>
</tr>
<tr>
<td>Refused to answer/missing</td>
<td>12 (0.32)</td>
<td>2 (16.67)</td>
<td>10 (83.33)</td>
<td></td>
</tr>
<tr>
<td>Chi-square/t test (d/f)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*1 yuan=US $0.14.

bN/A: not applicable.

Fisher exact test.

t-test for 2 independent samples (paired, 2-tailed).
Table 2. Associations (HR\textsuperscript{a} and 95% CI) of per quartile increase in a mixture of air pollutants during the pregnancy period with the risk of childhood asthma/wheezing.

<table>
<thead>
<tr>
<th>Association</th>
<th>Model 1\textsuperscript{b}, crude HR (95% CI)</th>
<th>Model 2\textsuperscript{c}, adjusted HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole pregnancy</td>
<td>1.05 (0.92-1.21)</td>
<td>1.11 (0.94-1.30)</td>
</tr>
<tr>
<td><strong>Grouped by clinical staging</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First trimester</td>
<td>1.19 (0.99-1.42)</td>
<td>1.22 (1.00-1.48)\textsuperscript{d}</td>
</tr>
<tr>
<td>Second trimester</td>
<td>1.16 (0.99-1.36)</td>
<td>1.24 (1.04-1.47)\textsuperscript{d}</td>
</tr>
<tr>
<td>Third trimester</td>
<td>1.10 (0.92-1.28)</td>
<td>1.10 (0.89-1.36)\textsuperscript{d}</td>
</tr>
<tr>
<td><strong>Grouped by lung development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The embryonic stage</td>
<td>1.03 (0.87-1.22)</td>
<td>1.02 (0.85-1.24)</td>
</tr>
<tr>
<td>The pseudoglandular stage</td>
<td>1.15 (0.97-1.37)</td>
<td>1.24 (1.03-1.51)</td>
</tr>
<tr>
<td>The canalicular stage</td>
<td>1.14 (0.97-1.34)</td>
<td>1.23 (1.01-1.50)</td>
</tr>
<tr>
<td>The saccular stage</td>
<td>0.99 (0.84-1.16)</td>
<td>1.08 (0.89-1.32)\textsuperscript{d}</td>
</tr>
<tr>
<td>The alveolar stage</td>
<td>1.04 (0.88-1.22)</td>
<td>0.98 (0.82-1.18)\textsuperscript{d}</td>
</tr>
</tbody>
</table>

\textsuperscript{a}HR: hazard ratio.
\textsuperscript{b}Crude model.
\textsuperscript{c}Adjusted for temperature, maternal age, gravidity, parity, maternal occupation, yearly income per capita, passive smoking, maternal diet, history of asthma, gestational diabetes mellitus, preterm birth, and season of conception.
\textsuperscript{d}There is a potential collinearity between temperature and air pollutants. The adjusted covariates do not include temperature.

**Figure 2.** Weights of the proportion of the positive or negative partial effect of each air pollutant. CO: carbon monoxide; NO\textsubscript{2}: nitrogen dioxide; O\textsubscript{3}: ozone; PM\textsubscript{2.5}: particulate matter 2.5; SO\textsubscript{2}: sulfur dioxide.

**Associations Between Child Ambient Pollutants Exposure and Childhood Wheezing Risk**

The adjusted HRs for childhood asthma and wheezing were 1.24 (95% CI 1.30-2.10) and 1.23 (95% CI 2.16-2.97), respectively, for each quartile increase in air pollutants during the first year and first 2 years of childhood. Notably, SO\textsubscript{2} made the largest positive contribution in both periods, accounting for 50.30% and 74.70%, respectively. Another significant contributor is NO\textsubscript{2}, accounting for 36.8% and 25.3% in the first year and first 2 years of childhood, respectively (Figures 2 and 3).
Figure 3. Associations (HR, 95%CI) of per quartile increase in the mixture of air pollutants during the childhood with the risk of childhood asthma/wheezing. HR: hazard ratio.

Sensitivity Analyses
We conducted sensitivity analyses in a subgroup of children with clear information on the feeding method and restricted the data to the period before the COVID-19 outbreak. Adjusting for the feeding method and the COVID-19 outbreak did not substantially alter the associations of prenatal and postnatal exposure to a mixture of air pollutants with the risk of childhood wheezing (Table 3). Similarly, the relationship between air pollutants and childhood asthma/wheezing did not change significantly after adjusting for air pollutants in the preceding periods in the model for that specific period (Table 4).

Table 3. Sensitivity analyses for subgroups of children with clear information on the feeding method and adjusted for cut-off date.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Whole pregnancy, HR(^a) (95% CI)</th>
<th>The first 2 years, HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1(^b)</td>
<td>1.32 (1.00-1.77)</td>
<td>2.39 (1.83-3.12)</td>
</tr>
<tr>
<td>Model 2(^c)</td>
<td>1.34 (1.01-1.78)</td>
<td>2.42 (1.85-3.15)</td>
</tr>
<tr>
<td>Model 3(^d)</td>
<td>1.11 (0.94-1.30)</td>
<td>2.53 (2.16-2.96)</td>
</tr>
<tr>
<td>Model 4(^e)</td>
<td>1.16 (0.98-1.37)</td>
<td>2.68 (2.27-3.18)</td>
</tr>
</tbody>
</table>

\(^a\)HR: hazard ratio.
\(^b\)Adjusted for the feeding mode, body temperature, maternal age, maternity, maternal occupation, per capita annual income, passive smoking, maternal diet, history of asthma, gestational diabetes mellitus, preterm birth, and season of conception among those with clear information on the feeding method.
\(^c\)Adjusted for confounders in model 1 except for the feeding method.
\(^d\)The cohort’s follow-up cut-off date was set at July 31, 2020.
\(^e\)The cohort’s follow-up cut-off date was set at December 31, 2019.
Several epidemiological studies have also reported associations between exposure to mixed air pollutants and the incidence of asthma/wheezing in children. The study was conducted in 3 northern Chinese cities (Urumqi, Beijing, and Taiyuan) and 4 southern cities (Nanjing, Shanghai, Chongqing, and Changsha). However, the 3 northern cities typically experience lower temperatures compared with Guangzhou. Nevertheless, both studies reached similar conclusions to our study, despite significant differences in terms of the number of years studied, air temperature, and air pollutant concentrations between the 2 studies and this study.

Table 4. Association between air pollutants and childhood asthma/wheezing after adjusting for preceding cyclic pollutants.

<table>
<thead>
<tr>
<th>Association</th>
<th>Model 1, HR (95% CI)</th>
<th>Model 2, adjusted HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grouped by clinical staging</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First trimester</td>
<td>1.22 (1.00-1.48)</td>
<td>N/A</td>
</tr>
<tr>
<td>Second trimester</td>
<td>1.24 (1.04-1.47)</td>
<td>1.41 (1.13-1.77)</td>
</tr>
<tr>
<td>Third trimester</td>
<td>1.10 (0.89-1.36)</td>
<td>1.04 (0.83-1.31)</td>
</tr>
<tr>
<td><strong>Grouped by lung development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The embryonic stage</td>
<td>1.02 (0.85-1.24)</td>
<td>N/A</td>
</tr>
<tr>
<td>The pseudoglandular stage</td>
<td>1.24 (1.03-1.51)</td>
<td>1.45 (1.13-1.86)</td>
</tr>
<tr>
<td>The canalicular stage</td>
<td>1.23 (1.01-1.50)</td>
<td>1.30 (1.03-1.64)</td>
</tr>
<tr>
<td>The saccular stage</td>
<td>1.08 (0.89-1.32)</td>
<td>1.03 (0.83-1.28)</td>
</tr>
<tr>
<td>The alveolar stage</td>
<td>0.98 (0.82-1.18)</td>
<td>0.94 (0.77-1.15)</td>
</tr>
<tr>
<td><strong>Postnatal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First year</td>
<td>1.65 (1.30-2.10)</td>
<td>1.77 (1.36-2.29)</td>
</tr>
<tr>
<td>First 2 years</td>
<td>2.53 (2.16-2.96)</td>
<td>3.66 (3.05-4.38)</td>
</tr>
</tbody>
</table>

HR: hazard ratio.

We adjusted the pollutants for the period preceding each period based on model 1. For example, air pollutants from the first trimester were included as covariates in the model for the second trimester, and air pollutants from both the first and second trimesters were included as covariates in the model for the third trimester. Air pollutants from the embryonic stage were added to the covariates in the model for the pseudoglandular stage, while contaminants from both the embryonic and pseudoglandular stages were added to the covariates in the model for the canalicular stage to adjust the model. Contaminants from the previous 3 periods were added to the covariates in the saccular stage model to adjust the model, and contaminants from the previous 4 periods were added to the covariates in the model for the alveolar stage to adjust the model. Contaminants for the period preceding each period were added to the covariates in the model for adjustment. The adjusted covariates do not include temperature.


dN/A: not applicable.

eAdjusted for temperature, maternal age, gravidity, parity, maternal occupation, yearly income per capita, passive smoking, maternal diet, history of asthma, gestational diabetes mellitus, premature, and season of conception.

Discussion

Principal Findings

In this prospective cohort study, we assessed the effects of exposure to 5 mixed air pollutants, including PM$_{2.5}$, SO$_2$, NO$_2$, CO, and O$_3$, during the first 1000 days of life on the incidence of asthma/wheezing in children. We observed positive associations between exposure to mixed air pollutants and the incidence of asthma/wheezing. Furthermore, exposure to mixed air pollutants during the second trimester, particularly during the pseudoglandular and canalicular stages, as well as in the first 2 years of life, may have a greater impact on the risk of childhood asthma/wheezing. These findings suggest potential critical exposure windows for air pollution. They extend our understanding of the adverse effects of air pollution exposure during early life on children’s respiratory health and provide suggestive information for implementing protective measures to mitigate the adverse effects of air pollution on children’s health.

Several epidemiological studies have also reported associations between maternal exposure to single or multiple air pollutants during pregnancy and the risk of childhood asthma, which aligns with our findings. For instance, a Canadian birth cohort study investigated the impact of perinatal air pollution exposure on asthma exacerbation in children, revealing that early-life exposure to higher concentrations of CO, NO$_2$, PM$_{10}$, and SO$_2$ elevates the risk of asthma in preschool children [28]. However, the aforementioned study took place earlier than this study and involved a city with lower temperatures and air pollutant concentrations compared with Guangzhou. Another cohort study from China demonstrated that each interquartile range increment of maternal exposure to NO$_2$ during pregnancy was positively associated with childhood asthma [29]. The study was conducted in 3 northern Chinese cities (Urumqi, Beijing, and Taiyuan) and 4 southern cities (Nanjing, Shanghai, Chongqing, and Changsha). However, the 3 northern cities typically experience lower temperatures compared with Guangzhou. Nevertheless, both studies reached similar conclusions to our study, despite significant differences in terms of the number of years studied, air temperature, and air pollutant concentrations between the 2 studies and this study.

Subgroup analyses indicated that the second trimester, specifically the pseudoglandular stage and the canalicular stage, may represent critical exposure windows for mixed air pollution–induced asthma in children. This finding is consistent with previous studies [21,30,31]. The Asthma Coalition for...
Community, Environmental, and Social Stress (ACCESS) project in the United States discovered that maternal exposure to higher PM\textsubscript{2.5} concentrations can elevate the risk of asthma attacks in offspring at 6 years of age, with the gestational period between 16 and 25 weeks potentially holding particular significance [21]. Another cohort study from Canada similarly demonstrated a significant association between maternal midgestational exposure to PM\textsubscript{2.5} and NO\textsubscript{2} and the incidence of asthma in children [30].

During the pseudoglandular stage, which occurs during GWs 6-16, all potentially conducting airways are formed, and the emergence of alveolar contours occurs. This stage is characterized by the significant development of fine bronchial structures. By GW 17, the major components of the lung are mostly formed, with the exception of the gas exchange component. During the canalicular stage, which spans GWs 16-24, early development of the lung parenchyma occurs. This stage is characterized by the expansion of the lumen of the bronchi and terminal fine bronchioles, as well as the initiation of thin-walled terminal vesicle formation at the ends of the respiratory fine bronchioles [22,32]. By the end of the canalicular stage, all conducting airways have emerged, allowing respiration and gas exchange to commence [33]. This indicates that during the pseudoglandular and canalicular stages, the developing lungs may be more susceptible to inhaled air pollutants, potentially leading to the development of asthma in childhood.

We also observed that exposure to mixed air pollutants during the first 2 years of life significantly elevated the risk of asthma/wheezing, suggesting that this period may represent another important window. This finding aligns with previous studies, such as the one [10] that investigated the association between air pollution and asthma development in all children born between 1999 and 2000 in southwestern British Columbia, Canada. The study assessed air pollution exposure in quartiles and discovered that children exhibited an increased odds ratio for asthma when exposed to higher quartiles of CO, NO, NO\textsubscript{2}, PM\textsubscript{10}, and SO\textsubscript{2} in the first year of life [10]. Furthermore, a prospective cohort study from Copenhagen identified air pollution as being associated with wheezing symptoms in children. They found a significant effect of NO\textsubscript{2} on asthma symptoms in infants 3 years after birth and also noted a significant correlation between PM\textsubscript{10}, CO concentrations, and wheezing symptoms in infants (0-1-year olds) [34]. A study conducted in Changsha, China, similarly demonstrated that postnatal exposure to outdoor industrial and traffic air pollutants, including PM\textsubscript{10}, SO\textsubscript{2}, and NO\textsubscript{2}, was significantly associated with an increased risk of asthma in children [35].

Studies have indicated that infants are particularly vulnerable to the effects of air pollutants due to their incomplete development of respiratory and immune systems [36]. Furthermore, alveolarization primarily occurs after birth, with approximately 85% of alveoli developing postnatally. Lung volume doubles by 6 months of age and triples by 1 year of age [37]. Collectively, these findings suggest that children’s respiratory systems continue to rapidly develop after birth, potentially rendering their lungs more sensitive to inhaled air pollutants.

In addition, our weighted analyses revealed that air pollutants at various stages may exert distinct impacts on the development of asthma in children. During the second trimester and the canalicular period, CO emerged as the largest contributor to the development of asthma/wheezing in children. Equally significant contributors during the pseudoglandular stage were CO and PM\textsubscript{2.5}. In childhood, the air pollutants contributing the most were SO\textsubscript{2} and NO\textsubscript{2}.

Many previous studies investigating the relationship between air pollutants and respiratory diseases have placed particular emphasis on PM\textsubscript{2.5} [38,39]. These studies suggest that the pathogenic mechanism of PM\textsubscript{2.5} is primarily associated with oxidative stress. This stress can either cross the placental barrier and directly affect the fetus or induce systemic inflammation in the mother, leading to a reduction in the supply of nutrients and oxygen to the fetus, thereby affecting fetal lung function [16]. However, our study proposes that CO may be a more significant contributor to air pollutants affecting lung development in pregnant children than PM\textsubscript{2.5}.

CO is a highly soluble, nonirritant gas that easily enters the bloodstream through the alveoli, capillaries, and placental barrier. Once in the bloodstream, it competes with oxygen and binds to hemoglobin, resulting in tissue hypoxia [40]. Hypoxia in maternal tissues can decrease the amount of gases carried by the fetus, potentially affecting the development of fetal lung function and possibly leading to the development of childhood asthma. However, there has been relatively less focus on its association with the development of asthma in children in developing countries, and the results of the few studies conducted have been inconsistent [41]. Therefore, it is recommended that future research on the relationship between CO and childhood asthma should be intensified to further clarify the association and the pathogenesis of the disease.

SO\textsubscript{2} is primarily released into the atmosphere through the combustion of fossil fuels and is highly soluble in water. Its effects are primarily harmful to the upper respiratory tract. SO\textsubscript{2} can cause severe airway constriction, which may be related to its pathogenic mechanism [41]. Andersson et al [42] reported an HR of 5.8 (95% CI 2.6-13) for asthma among individuals reporting inhalation of SO\textsubscript{2} gas compared with unexposed individuals following a survey of sulfite mill workers. In addition, several studies have demonstrated a connection between SO\textsubscript{2} and the development of asthma in children [43,44]. NO\textsubscript{2} is among the main components of TRAP. Similarly, the association between NO\textsubscript{2} and asthma has been affirmed by several studies [33,45,46]. Health Canada has also proposed a causal relationship between exposure to current levels of ambient NO\textsubscript{2} and increased asthma-related morbidity [47]. NO\textsubscript{2} is insoluble in water; it dissolves in the most distal airways and directly affect the fetus or induce systemic inflammation. This stress can either cross the placental barrier and directly affect the fetus or induce systemic inflammation in the mother, leading to a reduction in the supply of nutrients and oxygen to the fetus, thereby affecting fetal lung function [16]. However, our study proposes that CO may be a more significant contributor to air pollutants affecting lung development in pregnant children than PM\textsubscript{2.5}.

The association and the pathogenesis of the disease.
Limitations

There are several limitations that need to be acknowledged. First, due to the challenges associated with early diagnosis of asthma, there may be cases that do not present to the hospital after the onset of the disease, leading to missed diagnoses. In addition, some of the wheezing cases observed may not have been asthma, potentially resulting in misclassification of outcome measures. By contrast, the small number of diagnosed cases of asthma led us to combine asthma and wheezing into a single outcome. This approach prevented us from accurately estimating the effects of outdoor air pollution exposure on asthma incidence. Third, our study participants were sourced from only 1 hospital, and the administrative area covered by the hospital’s services (Panyu District, Guangzhou) differs somewhat from the demographic composition of the entire city of Guangzhou and of China (Multimedia Appendix 1). This aspect also imposes limitations on the generalizability of our results. Finally, we conducted a passive survey of children during their visits to a pediatric clinic, pediatric outpatient department, or emergency department. Children who did not attend the hospital were lost to follow-up, potentially leading to an underestimation of the effect of air pollutants.

Conclusion

This prospective cohort study offers new evidence suggesting a potential association between exposure to ambient mixed air pollutants during the first 1000 days of life and childhood asthma/wheezing. The second trimester, particularly the pseudoglandular stage and the canalicular stage, as well as the first 2 years after birth, may represent critical susceptibility periods for this exposure window. Our study is a cohort study, and therefore, firm causal inferences cannot be made. However, by incorporating findings from previous studies, our research contributes to a broader understanding of the adverse respiratory effects of outdoor air pollution exposure. This knowledge can empower clinicians and pregnant women to take proactive measures to reduce air pollution exposure, particularly during vulnerable windows of susceptibility. Such preventive measures are crucial for the prevention of childhood asthma.

Acknowledgments

The authors thank all participants who volunteered for the study. This work was supported by The National Natural Science Foundation of China (42175181, 81874276, 81773497, 81903294, and 42375180), the Guangzhou Basic and Applied Basic Research Foundation (202102020120), and the Foreign Expert Program of the Ministry of Science and Technology (G2022199006L).

Data Sharing Statement

The data that support the findings of this study will be available upon reasonable request to the corresponding author of the study.

Authors’ Contributions

FT conducted data analysis, wrote the original draft, and contributed to the literature review. XZ focused on the literature review. YY performed a formal analysis. XL conducted the investigation. GH participated in writing and editing. CW contributed to the investigation. ZC and QZ curated the data. SY assisted with the literature review. JF conducted a formal analysis. HY contributed to the literature review. WM participated in writing and editing. XD and TL supervised the project, conceptualized the study, and contributed to writing and editing. XD and TL are co-corresponding authors and have contributed equally to this article. All authors have reviewed and approved the submitted version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

General demographic information.
[DOCX File, 15 KB - publichealth_v10i1e52456_app1.docx]

Multimedia Appendix 2

Definition of covariates.
[DOCX File, 14 KB - publichealth_v10i1e52456_app2.docx]

Multimedia Appendix 3

VIF was used to check the collinearity of each component in the model. VIF: variance inflation factor.
[DOCX File, 25 KB - publichealth_v10i1e52456_app3.docx]

Multimedia Appendix 4

Correlation between pollutants and temperature.
References


27. R Project. R Foundation. URL: https://www.r-project.org [accessed 2024-03-26]


Abbreviations

ACCESS: Asthma Coalition for Community, Environmental, and Social Stress
CO: carbon monoxide
GW: gestational week
HR: hazard ratio
ICD-10: 10th edition of the International Classification of Diseases
NO2: nitrogen dioxide
O3: ozone
PEOH: Prenatal Environments and Offspring Health
PM2.5: particulate matter 2.5
QG-comp: quantile g-computation
SO2: sulfur dioxide
ST-LUR: spatiotemporal land use regression
TRAP: traffic-related air mixture pollutant
VIF: variance inflation factor

©Fenglin Tian, Xinqi Zhong, Yufeng Ye, Xiaohan Liu, Guanhao He, Cuiling Wu, Zhijing Chen, Qijiong Zhu, Siwen Yu, Jingjie Fan, Huan Yao, Wenjun Ma, Xiaomei Dong, Tao Liu. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Defining the Subtypes of Long COVID and Risk Factors for Prolonged Disease: Population-Based Case-Crossover Study

Skyler Resendez¹, DPhil; Steven H Brown², MD; Hugo Sebastian Ruiz Ayala¹, MSc; Prahalad Rangan¹, DPhil; Jonathan Nebeker², MPH, MD; Diane Montella², MD; Peter L Elkin¹,², MD

¹Department of Biomedical Informatics, University at Buffalo, State University of New York, Buffalo, NY, United States
²Office of Health Informatics, Department of Veterans Affairs, Washington, DC, United States

Abstract

Background: There have been over 772 million confirmed cases of COVID-19 worldwide. A significant portion of these infections will lead to long COVID (post–COVID-19 condition) and its attendant morbidities and costs. Numerous life-altering complications have already been associated with the development of long COVID, including chronic fatigue, brain fog, and dangerous heart rhythms.

Objective: We aim to derive an actionable long COVID case definition consisting of significantly increased signs, symptoms, and diagnoses to support pandemic-related clinical, public health, research, and policy initiatives.

Methods: This research employs a case-crossover population-based study using International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) data generated at Veterans Affairs medical centers nationwide between January 1, 2020, and August 18, 2022. In total, 367,148 individuals with ICD-10-CM data both before and after a positive COVID-19 test were selected for analysis. We compared ICD-10-CM codes assigned 1 to 7 months following each patient’s positive test with those assigned up to 6 months prior. Further, 350,315 patients had novel codes assigned during this window of time. We defined signs, symptoms, and diagnoses as being associated with long COVID if they had a novel case frequency of ≥1:1000, and they significantly increased in our entire cohort after a positive test. We present odds ratios with CIs for long COVID signs, symptoms, and diagnoses, organized by ICD-10-CM functional groups and medical specialty. We used our definition to assess long COVID risk based on a patient’s demographics, Elixhauser score, vaccination status, and COVID-19 disease severity.

Results: We developed a long COVID definition consisting of 323 ICD-10-CM diagnosis codes grouped into 143 ICD-10-CM functional groups that were significantly increased in our 367,148 patient post–COVID-19 population. We defined 17 medical-specialty long COVID subtypes such as cardiology long COVID. Patients who were COVID-19–positive developed signs, symptoms, or diagnoses included in our long COVID definition at a proportion of at least 59.7% (268,320/449,450, based on a denominator of all patients who were COVID-19–positive). The long COVID cohort was 8 years older with more comorbidities (2-year Elixhauser score 7.97 in the patients with long COVID vs 4.21 in the patients with non–long COVID). Patients who had a more severe bout of COVID-19, as judged by their minimum oxygen saturation level, were also more likely to develop long COVID.

Conclusions: An actionable, data-driven definition of long COVID can help clinicians screen for and diagnose long COVID, allowing identified patients to be admitted into appropriate monitoring and treatment programs. This long COVID definition can also support public health, research, and policy initiatives. Patients with COVID-19 who are older or have low oxygen saturation levels during their bout of COVID-19, or those who have multiple comorbidities should be preferentially watched for the development of long COVID.

(JMIR Public Health Surveill 2024;10:e49841) doi:10.2196/49841
KEYWORDS
long COVID; PASC; postacute sequelae of COVID-19; public health; policy initiatives; pandemic; diagnosis; COVID-19 treatment; long COVID cause; health care support; public safety; COVID-19; Veterans Affairs; United States; COVID-19 testing; clinician; mobile phone

Introduction
Numerous symptoms are cited as long-term sequelae of COVID-19. “The symptoms may affect a number of organ systems, occur in diverse patterns, and frequently get worse after physical or mental activity” [1]. Early studies found that the most common long-term symptoms were fatigue, dyspnea, joint pain, and chest pain [2]. Others reported gastrointestinal tract disorders correlated with gut microbiome shifts after COVID-19 infection [3,4]. Cognitive dysfunction, often referred to as brain fog, is another commonly reported long-term symptom [5]. Cognitive dysfunction is particularly concerning given evidence that COVID-19 can alter brain structure [6]. The most common self-reported symptoms documented via a smartphone app were fatigue, headache, dyspnea, and anosmia [7,8].

More recent studies have added to the knowledge base concerning symptoms of long COVID (post–COVID-19 condition). For example, a study observing cohorts in 4 Chinese cities showed that fatigue, cough, sore throat, difficulty in concentrating, feeling of anxiety, myalgia, and arthralgia were common severe long COVID symptoms. While there is considerable overlap, there is still value in new studies as they help validate previous studies and add new insights, such as identifying a previously underappreciated increase in anxiety among those who had COVID-19 [9].

Recent reviews have analyzed and integrated long COVID research to date [10]. Reviews like these contribute knowledge (eg, that over 200 symptoms have been identified, affecting multiple organ systems while proposing potential mechanisms). They also point out where our knowledge base is lacking. For example, Davis et al [10] observed a study stating that postural tachycardia syndrome can be a potential complication [11] but research has recently shown that long COVID can also greatly increase the likelihood of complications such as atrial fibrillation [12]. Such contributions are why continued research is critically important to combat the detrimental effects of long COVID.

Concerningly high long COVID frequencies have been reported since near the start of the pandemic. A cohort study from the Netherlands found that approximately 1 in 8 patients with COVID-19 developed long-term somatic symptoms [13]. Another study showed that approximately 30% of their cohort reported persistent symptoms, with many experiencing worse health-related quality of life compared with baseline and negative impacts on at least one activity of daily living [14]. More recent studies confirm these alarming statistics, indicating that 1 in 7 adults in the United States have reported symptoms of long COVID [15]. Furthermore, long COVID’s impacts extend beyond individual morbidity to include the health care system and economic consequences. Cutler [16] noted long COVID resulted in reduced workforce participation (eg, 44% out of the workforce), direct earning losses, and worker shortages in service jobs. This is likely directly related to the fatigue associated with long COVID, which has now been linked to muscular abnormalities and overall dysfunction of mitochondria within these tissues [17].

The widespread occurrence of lingering ailments and their impacts on individuals and society make clear the need for a long COVID definition. US public health officials note that we must balance our need for an accurate long COVID definition that includes all afflicted individuals against our need for interim long COVID definitions to expedite immediate action and mobilization [18]. In particular, a working definition of long COVID based on routinely collected coded data could support the identification of at-risk or undiagnosed patients for monitoring, referral, or therapeutic interventions. In this study, we empirically derive an actionable broad-based long COVID definition to support current clinical, public health, research, and policy initiatives related to the pandemic.

Methods
Overview
We selected veterans who had laboratory-confirmed positive COVID-19 tests. We examined the veterans’ electronic health records for novel International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) codes between 1 and 7 months after a positive COVID-19 test. We grouped codes with a novel frequency of 1/1000 or greater by diagnosis type creating ICD-10-CM functional groups and performed χ² testing with Bonferroni correction to compare diagnosis frequencies before and after a positive COVID-19 test. We defined ICD-10-CM functional groups that significantly increased in frequency as “upregulated” (see Figure 1). We then manually aggregated upregulated ICD-10-CM functional groups into medical specialties to organize our empiric definition of long COVID.
Population Definition and Data Extraction

We selected patients with laboratory-confirmed positive COVID-19 studies and followed them for 13 months (6 months before the COVID-19 test result and 7 months after COVID-19) to create a long COVID definition. We used the electronic health records of all the patients who tested positive for COVID-19 at Veterans Affairs (VA) medical facilities nationally between January 1, 2020, and August 18, 2022. In total, 2,377,720 patients were tested for COVID-19 during this time period.

We applied SQL queries to VA Informatics and Computing Infrastructure, Corporate Data Warehouse data tables [19] to generate 2 diagnosis files for analysis. The first file (“before”) contains a row of retrospectively collected information for each patient and each ICD-10-CM diagnosis assigned to them in the 6-month control window before their COVID-19 test. The row includes the ICD-10-CM code and its description, a unique patient identifier, the COVID-19 test date, and the calculated number of months between ICD-10-CM code entry and COVID-19 testing. This “before” file contained 14,980,288 observations across 426,970 patients.

We followed the patients for 7 months. The second file (“after”) was created 7 months after the last patient was included. The after file contained ICD-10-CM codes assigned during the 7 months following COVID-19 testing and similar related information as the “before” file. This “after” file contained 15,493,587 observations across 389,677 patients.

We limited the analysis to the 367,148 patients that appeared in both the “before” and “after” files to ensure that we had a diagnostic history for each patient and eliminated acute findings by removing all ICD-10-CM codes documented less than a month after the positive COVID-19 test (Figure 1). We used the date of the first positive COVID-19 test for patients with multiple positive tests. Multiple repeating ICD-10-CM codes for a single patient were counted once. We wrote R (R Development Core Team) and Python (Python Software Foundation) programs to remove all data concerning patients who tested negative for COVID-19, ICD-10-CM codes that were documented less than a month after the positive COVID-19 test, and patients who were not present in both the “before” and “after” files. The methodology used to generate the patient cohort is depicted in Figure 2.

We collected additional data to examine the association of demographics, comorbidities, vaccination status, and COVID-19 case severity with the incidence of long COVID. Demographic data collected included age, sex, race, and ethnicity. Comorbidities were evaluated using 2-year Elixhauser Comorbidity Indices Scores. Patients who were vaccinated were defined as having at least one COVID-19 vaccine dose recorded for at least two weeks and no more than 9 months before their positive COVID-19 test. We defined 2 classes of severe COVID-19 based on the minimum recorded oxygen saturation. The first class, severe COVID-19, was defined by a minimum oxygen saturation of <94% [20]. The second class, severe COVID-19 with severe desaturation, was defined by a minimum oxygen saturation of <88% [21].
Figure 2. Patient selection flowchart for this study. The first step excludes all patients that test negative for COVID-19. The second step excludes all patients that are not in both the “before” and “after” files, indicating either a lack of diagnosis history or a lack of follow-up, respectively. The final step excludes any patient that does not have a novel diagnosis 1 to 7 months after their positive test. A diagnosis is novel if it was not observed in the “before” file but was observed during this time. The “before” file consists of data concerning every ICD-10-CM code assigned to each patient who was COVID-19 positive in our cohort for up to 6-months before their test. The “after” file consists of data concerning every ICD-10-CM code assigned to each COVID-19 positive patient in our cohort for up to 7-months after their test. ICD-10-CM: International Classification of Diseases, 10th Revision, Clinical Modification.

Ethical Considerations
An institutional review board protocol was developed and approved for this research by the Department of Veterans Affairs (number 1580090). It uses electronic health care records data for patients of VA hospitals across the country. These data are protected health care data and are only used outside of the VA system in a summarized, deidentified format. The original informed consent from patients allows these analyses without additional consent. No additional compensation was given. A HIPAA (Health Insurance Portability and Accountability Act) waiver was approved (00004461).

Data Analysis
We chose 6 months “before” and “after” COVID-19 test control and case windows to allow patients to serve as their own controls. The 6 month “after” case window began 1 month after the positive COVID-19 test. We defined “novel” ICD-10-CM codes as those that appeared in a patient’s “after” file but not in the “before” file. We calculated the frequency of each novel ICD-10-CM code as the percentage of this study’s cohort assigned with the code. We excluded novel codes with a frequency of <1:1000 from further analysis. We defined codes as upregulated when the frequency of that code in the after file was significantly increased as compared to the frequency in the before file, if it had a $\chi^2$ with a Bonferroni corrected $P<.00006$. We limited the long COVID ICD-10-CM code functional groups to those that were significantly increased in frequency. We used “before” and “after” frequencies to calculate odds ratios and CIs for each novel ICD-10-CM functional group. We calculated odds ratios from frequency data.

ICD-10-CM Functional and Medical Specialty Groupings
We grouped ICD-10-CM codes in 3 steps. We first combined ICD-10-CM codes that had the same initial 3 characters. We then grouped ICD-10-CM codes with different initial characters if the diagnoses were functionally similar to create our ICD-10-CM functional groups. For example, we grouped I47.1 (supraventricular tachycardia), I47.2 (ventricular tachycardia), and R00.0 (tachycardia, unspecified) as tachycardia. Finally,
we manually curated each of these *ICD-10-CM* functional groups into medical specialties for organizational purposes.

**Long COVID Definition**

We included in our long COVID definition each *ICD-10-CM* code with an incidence over 6 months ($T_0 + 1M - T_0 + 7M$) >1:1000 (M: months; $T_0$: date of positive COVID-19 test) and a significant overall frequency increase. Patients with long COVID were defined as having any of the 323 upregulated *ICD-10-CM* codes between 2 and 7 months after their positive COVID-19 diagnosis, but not in their pre–COVID-19 diagnoses.

**Risk Factors for Long COVID**

The multivariate regression models were done for each risk factor one including age, gender, race, ethnicity, and 2-year Elixhauser score; a second with age, gender, race, ethnicity, 2-year Elixhauser score, and $O_2$ saturation <94%; and a third with age, gender, race, ethnicity, 2-year Elixhauser score, and COVID-19 vaccination status. We present the univariate rates as well as the results of the regression analysis using R (version 4.1.2) and RStudio.

**Results**

**Long COVID Definition**

We extracted *ICD-10-CM* diagnosis codes assigned to 367,148 patients who underwent a positive COVID-19 test at VA. A total of 268,320 patients had one or more novel COVID-19–related diagnoses. The remaining 98,828 patients had no novel long COVID *ICD-10-CM* diagnoses in their post–COVID-19 period when compared to their pre–COVID-19 period. Table 1 contains the demographic characteristics of this study’s cohort. Men were significantly older than women on average, 60.29 years (95% CI 60.24-60.35) versus 47.85 years (95% CI 47.73-47.97), respectively.

We developed a definition of long COVID consisting of 323 *ICD-10-CM* diagnosis codes grouped into 143 *ICD-10-CM* functional groups that were significantly increased in our 367,148 patient post–COVID-19 population. We define 17 medical specialty long COVID subtypes including cardiology long COVID, neurology long COVID, and pulmonary long COVID. Multimedia Appendix 1 shows the *ICD-10-CM* functional groups and medical specialties. Within each field, the *ICD-10-CM* code groups are sorted in descending order by their odds ratios. Combined odds ratios were calculated for each medical specialty category in Table S2 in Multimedia Appendix 2. Additional information about specific codes can be found in Table S3 in Multimedia Appendix 3.

Figures 3-6 show the signs, symptoms, and diagnoses with significantly increased relative risks in the post–COVID-19 period with their respective 95% CIs sorted by medical specialty. The data used to create these figures can be found in Table S4 in Multimedia Appendix 4.

Case counts were greatest for the specialties of cardiology (196,632), neurology (159,358), ophthalmology (149,817), and pulmonary (138,470). The lowest case counts were for oncology (7256), rheumatology (10,543), and dermatology (13,233; see Table 2 for more details).

Patients who were COVID-19 positive were assigned novel signs, symptoms, or diagnoses included in our definition of long COVID at a rate of between 59.7% (268,320/449,450, the percentage is based on patients who were COVID-19 positive and tested at the VA) and 76.6% (268,320/350,315, the percentage is based on all patients who were COVID-19 positive with a diagnostic history and follow-up diagnoses 1 to 7 months after test).

Most patients with long COVID were documented with at least one *ICD-10-CM* code found in our long COVID definition within 3 months of their positive COVID-19 test (168,194/268,320, 62.7%). The percentage of patients documented with their first long COVID *ICD-10-CM* code decreased with each subsequent month.
Table 1. Demographic data and 2-year Elixhauser scores for patients with long COVID and patients with non–long COVID.

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Patients with non–long COVID (N=98,828)</th>
<th>Patient with long COVID (N=268,320)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (95% CI)</td>
<td>52.14 (52.03-52.24)</td>
<td>60.85 (60.79-60.91)</td>
<td>(_a)</td>
</tr>
<tr>
<td>Elixhauser score, mean (95% CI)</td>
<td>3.03 (2.99-3.07)</td>
<td>7.05 (7.01-7.09)</td>
<td>(_)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>75,418 (76.31)</td>
<td>234,720 (87.48)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Women</td>
<td>16,854 (17.05)</td>
<td>31,651 (11.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not listed</td>
<td>6556 (6.63)</td>
<td>1949 (0.73)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>10,147 (10.27)</td>
<td>26,171 (9.75)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>71,595 (72.44)</td>
<td>227,404 (84.75)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not listed</td>
<td>17,086 (17.29)</td>
<td>14,745 (5.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>743 (0.75)</td>
<td>2243 (0.84)</td>
<td>.011</td>
</tr>
<tr>
<td>Asian</td>
<td>1420 (1.44)</td>
<td>2973 (1.11)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Black or African American</td>
<td>20,779 (21.03)</td>
<td>65,218 (24.31)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>905 (0.92)</td>
<td>2522 (0.94)</td>
<td>=.50</td>
</tr>
<tr>
<td>White</td>
<td>55,359 (56.02)</td>
<td>173,169 (64.54)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not listed</td>
<td>19,622 (19.85)</td>
<td>22,195 (8.27)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\(\_a\)Not available.

Figure 3. Odds ratios <3 for long COVID ICD-10-CM functional groups by medical specialty subtype: cardiology, dentistry, dermatology, endocrinology, gastroenterology, and otolaryngology. ICD-10-CM: International Classification of Diseases, 10th Revision, Clinical Modification.
Figure 4. Odds ratios $<3$ for long COVID ICD-10-CM functional groups by medical specialty subtype: general internal medicine, hematology, infectious disease, neurology, and oncology. ICD-10-CM: International Classification of Diseases, 10th Revision, Clinical Modification. MGUS: monoclonal gammopathy of undetermined significance; MRSA: methicillin-resistant *Staphylococcus aureus*; MSSA: meticillin-sensitive *Staphylococcus aureus*.

Figure 5. Odds ratios $<3$ for long COVID ICD-10-CM functional groups by medical specialty subtype: nephrology, ophthalmology, psychiatry or psychology, pulmonary, rheumatology, and urology. ICD-10-CM: International Classification of Diseases, 10th Revision, Clinical Modification.
Risk Factors for Long COVID

We presented in Table 1 a comparison of demographic characteristics and Elixhauser comorbidity scores of patients with long COVID and patients with non–long COVID. The long COVID cohort was older with more comorbidities. The long COVID cohort also had higher percentages of White and Black individuals and non-Hispanic and non-Latino ethnicities. The patients with a 2-year Elixhauser score of greater than 21 had a much higher proportion to develop long COVID ($P<.001$, Pearson $\chi^2$; see Table 3).

Our data did not indicate that vaccination was protective against the development of long COVID. However, vaccination resulted in significantly lower rates of novel acute respiratory distress syndrome in the post–COVID-19 period (13.2%, 95% CI

Table 2. Case counts by medical specialty.

<table>
<thead>
<tr>
<th>Subspecialty</th>
<th>Diagnoses</th>
<th>Cases, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiology</td>
<td>38</td>
<td>196,632</td>
</tr>
<tr>
<td>Neurology</td>
<td>38</td>
<td>159,358</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>20</td>
<td>149,817</td>
</tr>
<tr>
<td>Pulmonary</td>
<td>42</td>
<td>138,470</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>23</td>
<td>97,884</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>27</td>
<td>87,302</td>
</tr>
<tr>
<td>Nephrology</td>
<td>33</td>
<td>86,582</td>
</tr>
<tr>
<td>Psychiatry or psychology</td>
<td>9</td>
<td>75,292</td>
</tr>
<tr>
<td>Hematology</td>
<td>20</td>
<td>46,372</td>
</tr>
<tr>
<td>Urology</td>
<td>9</td>
<td>39,336</td>
</tr>
<tr>
<td>General internal medicine</td>
<td>9</td>
<td>33,130</td>
</tr>
<tr>
<td>Infectious diseases</td>
<td>22</td>
<td>30,998</td>
</tr>
<tr>
<td>Dentistry</td>
<td>13</td>
<td>30,998</td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>4</td>
<td>23,583</td>
</tr>
<tr>
<td>Dermatology</td>
<td>7</td>
<td>13,223</td>
</tr>
<tr>
<td>Rheumatology</td>
<td>2</td>
<td>10,543</td>
</tr>
<tr>
<td>Oncology</td>
<td>4</td>
<td>7256</td>
</tr>
<tr>
<td>Totals</td>
<td>320</td>
<td>1,225,980</td>
</tr>
</tbody>
</table>

Figure 6. Odds ratios >3 for long COVID ICD-10-CM functional groups by medical specialty subtype. ICD-10-CM: International Classification of Diseases, 10th Revision, Clinical Modification.
10.4%-16.9%) as compared with the unvaccinated population (19.6%, 95% CI 18.1%-21.2%; \(P<.001\)). Patients with minimum \(O_2\) saturations constituting severe COVID-19 and severe COVID-19 with severe desaturation were significantly more likely to develop long COVID (both had \(P<.001\), Pearson \(\chi^2\); see Table 4).

The multivariate regression models all confirmed that patients with COVID-19 during the Omicron variant predominant period were at a slightly higher risk of developing long COVID at \(P<.001\).

### Table 3. Proportion of patients that developed long COVID comparing different 2-year Elixhauser score ranges.

<table>
<thead>
<tr>
<th>2 year Elixhauser score</th>
<th>Non–long COVID count</th>
<th>Long COVID count</th>
<th>Percent long COVID(^a) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-21</td>
<td>96,055</td>
<td>242,108</td>
<td>71.60 (71.44-71.75)</td>
</tr>
<tr>
<td>22-42</td>
<td>2454</td>
<td>22,716</td>
<td>90.25 (89.88-90.61)</td>
</tr>
<tr>
<td>43-63</td>
<td>308</td>
<td>3317</td>
<td>91.50 (90.55-92.36)</td>
</tr>
<tr>
<td>64-84</td>
<td>11</td>
<td>179</td>
<td>94.21 (89.93-96.77)</td>
</tr>
</tbody>
</table>

\(^a\)The percentages come from the numbers to the left of each percentage. It is the long COVID count divided by the sum of both categories.

### Table 4. Low oxygen saturations and the proportion of patients that developed long COVID.

<table>
<thead>
<tr>
<th>Severe COVID-19</th>
<th>Non–long COVID count</th>
<th>Long COVID count</th>
<th>Percent long COVID(^a) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (O_2) (NIH(^b) definition(^c))</td>
<td>3903</td>
<td>29,411</td>
<td>88.28 (87.93-88.63)</td>
</tr>
<tr>
<td>No low (O_2) (NIH definition(^c))</td>
<td>94,925</td>
<td>238,909</td>
<td>71.57 (71.41-71.72)</td>
</tr>
<tr>
<td>Low (O_2) (severe desaturation(^d))</td>
<td>637</td>
<td>5566</td>
<td>89.73 (88.95-90.46)</td>
</tr>
<tr>
<td>No low (O_2) (severe desaturation(^d))</td>
<td>98,191</td>
<td>262,754</td>
<td>72.80 (72.65-72.94)</td>
</tr>
</tbody>
</table>

\(^a\)The percentages come from the numbers to the left of each percentage. It is the long COVID count divided by the sum of both categories.

\(^b\)NIH: National Institutes of Health.

\(^c\)Minimum (\(O_2\) saturation) <94%.

\(^d\)Minimum (\(O_2\) saturation) <88%.

### Discussion

#### Conclusions

Numerous reports document specialty-specific signs, symptoms, and diagnoses correlated with long COVID. We present a novel analysis based on a large national data set and the full multispecialty breadth of ICD-10-CM diagnosis codes to create a holistic long COVID definition that confirms and extends previous reports.

We allowed patients to be their own controls and used the entire cohort before and after COVID-19 infection to determine the relative risk of signs, symptoms, and disorders. This ensured that the signal was both novel and upregulated. We found patients who were COVID-19 positive developed signs, symptoms, or diagnoses included in our long COVID definition at a proportion of between 59.7% (268,320/449,450, the percentage is based on a denominator of all patients who were COVID-19 positive and tested at the VA) and 76.6% (268,320/350,315, the percentage is based on a denominator of all patients who were COVID-19 positive with a diagnostic history and follow-up diagnoses 1 to 7 months after test). More than three-fourths of patients with long COVID met our long COVID definition within 4 months of their positive COVID-19 test.

We found long COVID frequency differences based on race and ethnicity. These differences may be related to socioeconomic status, which is directly correlated with the presence of comorbidities [22-24]. The long COVID cohort was 8 years older with more comorbidities (2-year Elixhauser score 7.97 in the patients with long COVID vs 4.21 in the patients with non–long COVID). In our cohort, the men were significantly older than the women on average, 60.29 years (95% CI 60.24-60.35) versus 47.85 years (95% CI 47.73-47.97), respectively. We found that long COVID frequency was increased in patients who were more severely ill before infection and patients who had a more severe bout of COVID-19 as judged by their minimum oxygen saturation.

We found 143 upregulated diagnostic groups, with odds ratios as high as 23. We also found 17 upregulated medical specialty groupings containing between 3 and 21 signs, symptoms, or diagnoses. This provides strong evidence for a broad definition of long COVID.

Carfi et al [2] found that the most common long-term symptoms were fatigue, dyspnea, joint pain, and chest pain. Each except joint pain is represented in our long COVID definition. However, joint pain may be related to findings in our definition such as difficulty walking and an overall decrease in mobility. COVID-19 is known to cause lung abnormalities, especially in

https://publichealth.jmir.org/2024/1/e49841
cases with pneumonia [25]. We found that the likelihood of developing pneumonia after COVID-19 infection is significantly upregulated, potentially interconnected with the numerous findings in our pulmonary long COVID definition. Autopsy evaluation of COVID-19 victims’ lung tissue demonstrated diffuse alveolar damage with perivascular T-cell infiltration and severe endothelial injury [26]. Patients with long COVID have been found to have abnormal 133Xe magnetic resonance imaging gas exchange and computed tomography vascular density measurements, which we postulate could be related to the pulmonary fibrosis (J84.10) or emphysema (J43.9) diagnoses identified in our definition [27].

Our definition shows that the long-term effects of COVID-19 are associated with damage to numerous body systems including the kidneys, heart, eyes, and nervous system. Our results are corroborated by other studies. Cognitive dysfunction (brain fog) is often associated with long COVID and can be difficult to diagnose and treat [5]. COVID-19 infection is far more likely to cause cardiac complications than vaccination [28]. The gastrointestinal codes we observed reflect previous literature [29] and may relate to reported alterations to the gastrointestinal tract after COVID-19 [3,4]. Finally, previous studies have noted that COVID-19 can alter ocular physiology, supporting our ophthalmology-related findings [30].

Patients with more severe cases of COVID-19, as manifested by low oxygen saturations, should be watched carefully for the development of long COVID as they are significantly more likely to develop long COVID. Sicker patients with higher 2-year Elixhauser scores were significantly more likely to develop long COVID. Patients with multiple comorbidities should be made aware of this risk and participate in active surveillance for the development of signs and symptoms of long COVID.

The American Medical Association notes there are 3 categories of patients with long COVID: those who do not recover completely and have ongoing symptoms, those with symptoms related to chronic hospitalization, and those who develop new symptoms after recovery [31]. In our study, we did not differentiate by these subtypes and instead leave that to future research. It is possible that some of these signs and symptoms may have occurred during the first month and may be the persistent subtype. It is possible that some of the upregulated codes may be found with other serious illnesses, though only 9.1% (33,314/367,148) of our cohort had severe COVID-19 based on oxygen saturation <94%. We are not able to distinguish conditions that represent an acceleration of pre-existing disease from those that represent de novo COVID-19–related conditions. For example, is the increased incidence of nonsore throat elevation myocardial infarction (I21.4) related to the general stress of acute illness impacting pre-existing coronary artery disease or to an underlying de novo long COVID–related condition? A better understanding will require additional research. In any event, whether causal or associative, de novo disease or exacerbation of chronic disease and new or persistent clinical problems require assessment, treatment, and monitoring.

Limitations include that the cohort study population is 84% men, reflective of the overall patient population of VA which is between 87% and 95% men (depending upon data source and whether gender has been self-reported) [32,33]. Additionally, the male veteran population who use the VA health care system is older than the population of female veterans who use the VA. Our study did not include home testing for COVID-19 that went unreported to the VA health care system. Patients who tested positive during the omicron dominant time period were slightly more likely to develop long COVID when compared to the earlier strains (66,643/87,522, 76%; 201,677/279,626, 72%; P<.001). The reality of emerging viral variants emphasizes the need for a well-defined and well-maintained definition of long COVID over time and with variant-specific derivation. This study was not powered to show the independence of the individual risk factors for long COVID.

We hope that our empirically defined long COVID definition will lead to more consistent identification of long COVID and its medical specialty subtypes and support of a variety of COVID-19–related initiatives. Our definition is actionable as individuals who have multiple comorbidities and more severe bouts of COVID-19 should be followed more closely for the development of long COVID signs or symptoms. Our definition can also inform screening questions for high-risk patients. For example, helping clinicians identify patients with enhanced long COVID risk who may benefit from monitoring programs or patients with previously undiagnosed long COVID for whom it may be appropriate to create a referral to a long COVID clinic. We also anticipate that our long COVID definition will support the standardization of future subspecialty-specific long COVID research.

Future research should look at health outcomes for each long COVID-19 medical specialty subtype to identify those at greatest risk of developing severe morbidity. Predictive analytics should be used to help refer these individuals earlier to monitoring and treatment programs.

As of December 17, 2023, there have been over 772 million confirmed cases of COVID-19 worldwide [34]. Case counts are ever-increasing. As Levine [18] notes, immediately useful long COVID definitions are needed as are ultimately more fully inclusive definitions. We offer our long COVID definition as a public health contribution to our pandemic response.

Acknowledgments

This work has been supported in part by grants from the National Institutes of Health (National Library of Medicine T15LM012495 and R25LM014213, National Institute on Alcohol Abuse and Alcoholism R21AA026954 and R33AA0226954, and National Center for Advancing Translational Sciences UL1TR001412). This study was funded in part by the Department of Veterans Affairs.
Data Availability

All data generated or analyzed during this study are either protected patient data that require a Veterans Affairs research appointment and institutional review board protocol to access or summarized and included in this published paper (and its supplementary information files).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Long COVID definition results by International Classification of Diseases, 10th Revision, Clinical Modification functional group description and medical specialty.

[DOCX File, 24 KB - publichealth_v10i1e49841_app1.docx ]

Multimedia Appendix 2

Combined odds ratios calculated for each medical specialty category.

[XLSX File (Microsoft Excel File), 12 KB - publichealth_v10i1e49841_app2.xlsx]

Multimedia Appendix 3

Additional information about specific codes.

[XLSX File (Microsoft Excel File), 24 KB - publichealth_v10i1e49841_app3.xlsx]

Multimedia Appendix 4

Data used to create these figures.

[XLSX File (Microsoft Excel File), 27 KB - publichealth_v10i1e49841_app4.xlsx]

References


©Skyler Resende, Steven H Brown, Hugo Sebastian Ruiz Ayala, Prahalad Rangan, Jonathan Nebeker, Diane Montella, Peter L Elkin. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 30.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Association of the Type of Public Pension With Mental Health Among South Korean Older Adults: Longitudinal Observational Study

Seung Hoon Kim, MD, JD, PhD; Hyunkyu Kim, MD, PhD; Sung Hoon Jeong, PhD; Eun-Cheol Park, MD, PhD

1Department of Ophthalmology, Soonchunhyang University Hospital Cheonan, Soonchunhyang University College of Medicine, Chenonan, Republic of Korea
2Institute of Health Services Research, Yonsei University, Seoul, Republic of Korea
3Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea
4Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul, Republic of Korea
5National Traffic Injury Rehabilitation Research Institute, National Traffic Injury Rehabilitation Hospital, Yangpyeong, Republic of Korea
6Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Seung Hoon Kim, MD, JD, PhD
Department of Ophthalmology
Soonchunhyang University Hospital Cheonan
Soonchunhyang University College of Medicine
31 Sooncheonhyang 6-gil
Dongnam-gu
Chenonan, 31151
Republic of Korea
Phone: 82 425702114
Email: immergru@gmail.com

Abstract

Background: As income and health are closely related, retirement is considered undesirable for health. Many studies have shown the association between pension and health, but no research has considered the association between contribution-based public pensions or their types and health.

Objective: This study investigates the association between the type of contributory public pension and depressive symptoms among older adults.

Methods: We analyzed the data of 4541 older adults who participated in the South Korea Welfare Panel Study (2014-2020). Depressive symptoms were measured using the 11-item Center for Epidemiologic Studies Depression scale. Public pensions in South Korea are classified into specific corporate pensions and national pensions. For subgroup analyses, pensioners were categorized according to the amount of pension received and the proportion of public pension over gross income. Analyses using generalized estimating equations were conducted for longitudinal data.

Results: Individuals receiving public pension, regardless of the pension type, demonstrated significantly decreased depressive symptoms (national pension: \( \beta = -0.734; P < .001 \); specific corporate pension: \( \beta = -0.775; P = .02 \)). For both pension types, the higher the amount of benefits, the lower were the depression scores. However, this association was absent for those who received the smaller amount among the specific corporate pensioners. In low-income households, the decrease in the depressive symptoms based on the amount of public pension benefits was greater (fourth quartile of national pension: \( \beta = -1.472; P < .001 \); second and third quartiles of specific corporate pension: \( \beta = -3.646; P < .001 \)).

Conclusions: Our study shows that contributory public pension is significantly associated with lower depressive symptoms, and this association is prominent in low-income households. Thus, contributory public pensions may be good income sources for improving the mental health of older adults after retirement.

(JMIR Public Health Surveill 2024;10:e49129) doi:10.2196/49129

https://publichealth.jmir.org/2024/1/e49129

JMIR Public Health Surveill 2024 | vol. 10 | e49129 | p.1617
(page number not for citation purposes)
KEYWORDS
depression; retirement; contributory public pension; low-income household; public health; mental health; data; big data; retirement; longitudinal data; low income

Introduction
Depression in older adults decreases successful aging, adversely affecting not only the health-related quality of life of individuals but also of families and communities [1]. Older adults are more likely to be exposed to depression because they experience a series of adversities such as death, retirement, and health problems [2,3]. In old age, depression is a risk factor for nonsuicidal mortality, and both major and minor depression have been associated with suicide [4].

The prevalence of depression in South Korea (hereafter referred to as Korea) was 2.09% in 2019, which was lower than that reported in Japan (2.10%) and in the United States (4.38%) [5]. However, the prevalence of depression tends to increase with age, reaching 13.9% in those in their 60s and 70s and 18.4% in those in their 80s or older [6]. The proportion of older adults with depression reached 14% in 2018, and Korea is expected to become a superaged society by 2025, when the older adult population reaches 20.3% of the total population in Korea [7]. Therefore, a better understanding of depression in old age is of paramount importance from a clinical and public health perspective.

Income and health are closely related; to some extent, the more money people make, the better is their health [8]. This makes retirement, where a certain income for a person ceases, undesirable for health. In other words, old age income is important for health and other aspects [9]. Low economic status often leads to adjustment disorders accompanied by a depressed mood [10]. In Korea, the disposable income of older adults is less than 70% of the economy-wide average, which is the lowest among the Organization for Economic Co-operation and Development countries [11]. Moreover, Korea has a relative poverty rate of 45% among adults over 65 years, which is the highest among the Organization for Economic Co-operation and Development countries. Relative poverty is defined as an income below half the national median equivalized household income.

As in many countries, Korea’s retirement income security system has a multilayered structure. First, the National Basic Livelihood Security System and the noncontributory basic old age pension are positioned as safety nets for the low-income class. Second, public pension, which provides pension benefits based on the subscribers, plays the most important role in guaranteeing retirement income [12]. The types of public pensions vary depending on the country, but in Korea, they are divided into (1) specific corporate pension for special workers such as public officials, private school faculty members, and military personnel, and (2) national pension for the general public [13,14]. Lastly, individuals can prepare additional retirement income by choosing to subscribe to a private pension operated by a company rather than the government. Among these multilayered income security systems, public pensions play a role in dispersing the risk of poverty that may occur among citizens. Furthermore, they allow the entire society to bear the costs evenly, without the double burden of paying for other people’s retirement living expenses on income-earning people due to those who are guaranteed a minimum living by relying on the public assistance system.

The relationship between health and pension, which guarantees income for the aged population, is debatable [15-18]. Unlike noncontributory pensions, which are paid only if income requirements are met (even if there is no certain contribution), few studies have been conducted on how contribution-based pensions (guaranteed for their contributions) affect the mental health of older adults [19]. When public pensions are divided into national pensions and specific corporate pensions, as in Korea, and there is a large difference in the amount received based on the type of pension (assuming that the contribution is different), the association of depression may differ depending on the type and amount of public pension received.

We hypothesized that a contribution-based pension would help the mental health of the older population and that larger public pension benefits would be more beneficial to mental health. It was also assumed that as the characteristics of the 2 pillars of Korea’s public pension are different, the degree of impact on mental health will also be different. Therefore, this study examines whether public pensions are associated with depressive symptoms, depending on the type and amount of receipt, among older adults over 65 years.

Methods

Study Population and Data
Data from the Korea Welfare Panel Study (KoWePS) conducted by the Korea Institute for Health and Social Affairs and Seoul National University were analyzed in this study. The KoWePS was a large-scale, annual, longitudinal panel survey conducted from 2006 (wave 1) to 2020 (wave 15). We used samples extracted from wave 9 (2014) to wave 15 (2020), as the 11-item Center for Epidemiologic Studies-Depression (CESD-11) scale measures depressive symptoms. The study participants were also limited to those who were 65 years or older and who did not have a history of depression in wave 9 in order to evaluate the association between pension benefits and depressive symptoms. After excluding participants with missing data, the CESD-11 score (n=254), 4541 respondents were included in the final sample of the baseline study year.

Depressive Symptoms
The primary outcome was depressive symptoms, which were evaluated based on the 11-item version of the CESD scale. The CESD-11 is a well-validated self-reported screening tool and is a shorter version of the original 20-item instrument [20]. It is designed to evaluate 11 depressive symptoms for 1 week on a 4-point scale (0-3). The respondents reported symptoms for each year of the study period. The total score was multiplied by a 4-point scale (0-3). The respondents reported symptoms for each year of the study period. The total score was multiplied
analyses were performed to investigate whether the annual variables in contrast to the fixed effects model [24]. Subgroup analyses can predict the coefficients of time-dependent average response for every 1-unit increase in a covariate across study design [23]. The GEE model assesses the change in the association between type of public pension and depressive symptoms, we used the generalized estimating equation (GEE) design was used to select households, including multiple household members, household income, highest level of education, and employment status. Health-related factors included smoking status, alcohol consumption, and chronic diseases. Income-related factors were realized property income, including interest or rental income; public transfer income, including basic old age pension, a noncontributory social pension in Korea; private transfer income, including income from the adult children and private pension; and household debt.

Covariates
This study includes demographics, socioeconomic characteristics, health-related factors, and income-related factors as covariates, which reflected the results obtained for each wave. Demographic variables included sex, age, and region. Socioeconomic variables included marital status, number of household members, household income, highest level of education, and employment status. Health-related factors included smoking status, alcohol consumption, and chronic diseases. Income-related factors were realized property income, including interest or rental income; public transfer income, including basic old age pension, a noncontributory social pension in Korea; private transfer income, including income from the adult children and private pension; and household debt.

Statistical Analyses
The distribution of general characteristics was calculated at baseline. Two-sided t tests, analysis of variance, and univariate linear regression were used to analyze the differences in the CESD-11 scores. The KoWePS data used in this study were hierarchically structured, and a stratified multistage probability design was used to select households, including multiple individuals from the same household [22]. For estimating the association between type of public pension and depressive symptoms, we used the generalized estimating equation (GEE) model, which can explain the time variations and correlations among the repeated measurements observed in the longitudinal study design [23]. The GEE model assesses the change in the average response for every 1-unit increase in a covariate across the population and can predict the coefficients of time-dependent variables in contrast to the fixed effects model [24]. Subgroup analyses were performed to investigate whether the annual amount of pension benefit was associated with depressive symptoms by using the GEE model. Furthermore, this study assessed the relationship between the proportion of public pension income in gross income and depressive symptoms. Finally, stratified analysis was performed by stratifying low- and high-income households by defining less than 60% of the equalized median income as low-income households [25]. All calculated P values were 2-sided; P values <.05 were considered significant. All analyses were performed using SAS software (version 9.4; SAS Institute). This study follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for observational studies [26].

Ethics Approval
This study was approved by the institutional review board of Severance Hospital at Yonsei University College of Medicine (approval 4-2021-0310). Informed consent was waived by the institutional review board, as the data of KoWePS are open to the public and do not contain any personally identifiable information.

Results
Characteristics of the Study Participants
The study participants were more likely to be older than the excluded individuals, live with their family, have higher incomes, and be unemployed (Table S1 of Multimedia Appendix 1). Table 1 describes the participants’ general characteristics in the baseline year (2014). A total of 4541 participants were included, and the mean CESD-11 score was 6.13 (SD 5.41); 3128 (68.9%) participants did not receive public pension, 1225 (26.9%) received national pension, and 188 (4.1%) received
specific corporate pension. Their mean CESD-11 scores were 6.72 (SD 5.55), 5.03 (SD 4.89), and 3.61 (SD 4.42), respectively. Table S2 of Multimedia Appendix 2 demonstrates the demographic characteristics of the study participants in 2014 according to pension type. The rate of not receiving public pensions among women was higher than that among men and among those aged 80 years or older. Furthermore, the lower the level of education and the lower the income, the higher was the rate of not being enrolled in public pensions.

Table 2 presents the results of the GEE model, which shows that the scores of CESD-11 were lower among those who received public pensions. After adjusting all covariates, including income-related factors, individuals receiving public pension, regardless of pension type, tended to have lower CESD-11 scores than those who did not receive public pension (national: \( \beta = -0.734; P < 0.001 \); specific: \( \beta = -0.775; P = 0.02 \)). The results also showed that a higher realized property income was correlated with lower CESD-11 scores (\( \beta = -0.283; P < 0.001 \)). However, public or private transfer income were not associated with depressive symptoms.

Table 3 shows the results of the GEE model for the subgroup analyses of the associations among amount of pension benefit, proportion of public pension income, and depressive symptoms. After adjusting for all the covariates, older adults who received more than the fourth quartile among the specific corporate pension recipients reported the lowest depression score (\( \beta = -1.221; P = 0.001 \)). For both national pension and specific corporate pension recipients, the CESD-11 scores tended to decrease as the amount of pension benefits increased. The CESD-11 score was significantly lowered when the proportion of national pension income over gross income was 75% or less (\( P < 0.001 \)), but no statistical significance was found when it exceeded 75%. Meanwhile, when the proportion of specific corporate pension income over gross income exceeded 25%, the depression score decreased significantly (\( P < 0.05 \)).

Table 4 shows the results of the analyses, which present the associations of public pension with depressive symptoms stratified by household income. In low-income households with less than 60% of equalized gross income, there was a clear relationship among public pension benefits, amount of benefits, and lower CESD-11 scores. Moreover, the CESD-11 score was significantly lowered even if the proportion of national pension income over gross income exceeded 75% (\( \beta = -2.783; P < 0.001 \)).
Table 1. General characteristics of the study participants (2014 baseline year) (N=4541).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values, n (%)</th>
<th>CESD-11$^a$ score, mean (SD)$^b$</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pension type</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>3128 (68.9)</td>
<td>6.72 (5.55)</td>
<td></td>
</tr>
<tr>
<td>National pension</td>
<td>1225 (26.9)</td>
<td>5.03 (4.89)</td>
<td></td>
</tr>
<tr>
<td>Specific corporate pension</td>
<td>188 (4.1)</td>
<td>3.61 (4.42)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>1724 (38)</td>
<td>4.88 (4.97)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2817 (62)</td>
<td>6.90 (5.53)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>65-69</td>
<td>1066 (22.9)</td>
<td>4.92 (4.95)</td>
<td></td>
</tr>
<tr>
<td>70-74</td>
<td>1449 (32.1)</td>
<td>5.77 (5.22)</td>
<td></td>
</tr>
<tr>
<td>75-79</td>
<td>1181 (26.2)</td>
<td>6.47 (5.33)</td>
<td></td>
</tr>
<tr>
<td>≥80</td>
<td>845 (18.7)</td>
<td>7.82 (5.93)</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Metropolitan</td>
<td>1619 (35.7)</td>
<td>5.91 (5.40)</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>2771 (61)</td>
<td>6.32 (5.45)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>151 (3.3)</td>
<td>5.15 (4.83)</td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Married</td>
<td>2668 (58.8)</td>
<td>5.16 (5.06)</td>
<td></td>
</tr>
<tr>
<td>Divorced, widowed, separated, or never married</td>
<td>1873 (41.2)</td>
<td>7.52 (5.60)</td>
<td></td>
</tr>
<tr>
<td>Number of household members</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1</td>
<td>1362 (30)</td>
<td>7.80 (5.59)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2380 (52.4)</td>
<td>5.40 (5.21)</td>
<td></td>
</tr>
<tr>
<td>≥3</td>
<td>799 (17.6)</td>
<td>5.49 (5.09)</td>
<td></td>
</tr>
<tr>
<td>Household income</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High</td>
<td>1138 (25.1)</td>
<td>4.38 (4.55)</td>
<td></td>
</tr>
<tr>
<td>Upper-middle</td>
<td>1131 (24.9)</td>
<td>5.50 (5.26)</td>
<td></td>
</tr>
<tr>
<td>Lower-middle</td>
<td>1133 (25)</td>
<td>6.55 (5.57)</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1139 (25.1)</td>
<td>8.10 (5.51)</td>
<td></td>
</tr>
<tr>
<td>Highest level of education</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>College and higher</td>
<td>3725 (82)</td>
<td>6.50 (5.49)</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>565 (12.4)</td>
<td>4.54 (4.75)</td>
<td></td>
</tr>
<tr>
<td>Middle school or lower</td>
<td>251 (5.5)</td>
<td>4.35 (4.59)</td>
<td></td>
</tr>
<tr>
<td>Employment type</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Wage workers</td>
<td>495 (10.9)</td>
<td>4.87 (4.74)</td>
<td></td>
</tr>
<tr>
<td>Self-employed</td>
<td>816 (18)</td>
<td>4.94 (4.85)</td>
<td></td>
</tr>
<tr>
<td>Not employed</td>
<td>3230 (71.1)</td>
<td>6.63 (5.57)</td>
<td></td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>3374 (74.3)</td>
<td>6.67 (5.57)</td>
<td></td>
</tr>
<tr>
<td>~Once/week</td>
<td>566 (12.5)</td>
<td>4.55 (4.50)</td>
<td></td>
</tr>
<tr>
<td>&gt;Once/week</td>
<td>601 (13.2)</td>
<td>4.63 (4.71)</td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td>.40</td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>4054 (89.3)</td>
<td>6.11 (5.41)</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Values, n (%)</td>
<td>CESD-11&lt;sup&gt;a&lt;/sup&gt; score, mean (SD)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>P value</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Current smoker</td>
<td>487 (10.7)</td>
<td>6.29 (5.47)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chronic diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>447 (9.8)</td>
<td>4.03 (4.51)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4094 (90.2)</td>
<td>6.36 (5.45)</td>
<td></td>
</tr>
<tr>
<td>Realized property income&lt;sup&gt;c&lt;/sup&gt;</td>
<td>4541 (100)</td>
<td>1.59 (4.70)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Public transfer income&lt;sup&gt;d&lt;/sup&gt;</td>
<td>4541 (100)</td>
<td>1.10 (0.71)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Private transfer income&lt;sup&gt;e&lt;/sup&gt;</td>
<td>4541 (100)</td>
<td>4.22 (5.27)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Household debt</td>
<td>4541 (100)</td>
<td>7.10 (29.99)</td>
<td>.26</td>
</tr>
</tbody>
</table>

<sup>a</sup>CESD-11: 11-item Center for Epidemiologic Studies Depression (scale).

<sup>b</sup>Total mean (SD) score of all participants was 6.13 (5.41).

<sup>c</sup>Including interest income and rental income. The mean income was considered in US $1000.

<sup>d</sup>Including basic old age pension. The mean income was considered in US $1000.

<sup>e</sup>Including subsidy from children and personal pension. The mean income was considered in US $1000.
Table 2. Results of the generalized estimating equation analysis of factors associated with depressive symptoms.

<table>
<thead>
<tr>
<th>Variables</th>
<th>CESD-11 score, β (SE)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pension type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>National pension</td>
<td>−.734 (.110)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specific corporate pension</td>
<td>−.775 (.242)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Female</td>
<td>.543 (.134)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-69</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>70-74</td>
<td>.033 (.100)</td>
<td>.74</td>
</tr>
<tr>
<td>75-79</td>
<td>.259 (.122)</td>
<td>.03</td>
</tr>
<tr>
<td>≥80</td>
<td>.779 (.144)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metropolitan</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Urban</td>
<td>.129 (.109)</td>
<td>.24</td>
</tr>
<tr>
<td>Rural</td>
<td>−.422 (.265)</td>
<td>.11</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Divorced, widowed, separated, or never married</td>
<td>.654 (.169)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Number of household members</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>−.415 (.184)</td>
<td>.02</td>
</tr>
<tr>
<td>≥3</td>
<td>−.195 (.216)</td>
<td>.37</td>
</tr>
<tr>
<td><strong>Household income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Upper middle</td>
<td>.656 (.109)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lower middle</td>
<td>1.055 (.127)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Low</td>
<td>1.501 (.163)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Highest level of education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College and higher</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>High school</td>
<td>.099 (.238)</td>
<td>.68</td>
</tr>
<tr>
<td>Middle school or lower</td>
<td>.561 (.227)</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Employment type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage workers</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Self-employed</td>
<td>.294 (.142)</td>
<td>.04</td>
</tr>
<tr>
<td>Not employed</td>
<td>.899 (.110)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Alcohol consumption</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>~Once/week</td>
<td>−.536 (.099)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&gt;Once/week</td>
<td>−.614 (.117)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Variables</td>
<td>CESD-11&lt;sup&gt;a&lt;/sup&gt; score, β (SE)</td>
<td>P value</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-----------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Current smoker</td>
<td>.749 (.161)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Chronic diseases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>1.015 (.110)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Realized property income&lt;sup&gt;b&lt;/sup&gt; per US $10,000/year higher</td>
<td>–.283 (.074)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Public transfer income&lt;sup&gt;c&lt;/sup&gt; (including basic old age pension) per US $10,000/year higher</td>
<td>–.474 (.283)</td>
<td>.09</td>
</tr>
<tr>
<td>Private transfer income&lt;sup&gt;d&lt;/sup&gt; per US $10,000/year higher</td>
<td>.059 (.074)</td>
<td>.42</td>
</tr>
<tr>
<td>Household debt per US $10,000/year higher</td>
<td>.021 (.005)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>CESD-11: 11-item Center for Epidemiologic Studies Depression (scale).

<sup>b</sup>Including interest income and rental income. If the participant’s realized property income increases by US $10,000 per year, the estimated 11-item Center for Epidemiologic Studies Depression (scale) score will decrease by .283 points.

<sup>c</sup>Including basic old age pension.

<sup>d</sup>Including subsidy from children and personal pension.

**Table 3.** Results of the generalized estimating equation model for the subgroup analyzes of the associations among amount of pension benefit, proportion of public pension income, and depressive symptoms.

<table>
<thead>
<tr>
<th>Variables</th>
<th>CESD-11&lt;sup&gt;a&lt;/sup&gt; score, β (SE)</th>
<th>P value&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amount of pension benefit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>National pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>~25% (US $1440/year)</td>
<td>−.676 (.154)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>~75% (US $2790/year)</td>
<td>−.703 (.124)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>~100% (US $29,400/year)</td>
<td>−.979 (.175)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specific corporate pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>~25% (US $16,800/year)</td>
<td>−.548 (.380)</td>
<td>.15</td>
</tr>
<tr>
<td>~75% (US $27,600/year)</td>
<td>−.834 (.291)</td>
<td>.004</td>
</tr>
<tr>
<td>~100% (US $44,880/year)</td>
<td>−1.221 (.375)</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Proportion of pension income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>National pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0&lt;pension income/gross income≤0.25</td>
<td>−.665 (.112)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>0.25&lt;pension income/gross income≤0.75</td>
<td>−1.220 (.208)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>0.75&lt;pension income/gross income≤1.0</td>
<td>−.992 (.881)</td>
<td>.26</td>
</tr>
<tr>
<td>Specific corporate pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0&lt;pension income/gross income≤0.25</td>
<td>−.179 (.795)</td>
<td>.82</td>
</tr>
<tr>
<td>0.25&lt;pension income/gross income≤0.75</td>
<td>−.746 (.266)</td>
<td>.005</td>
</tr>
<tr>
<td>0.75&lt;pension income/gross income≤1.0</td>
<td>−.929 (.305)</td>
<td>.002</td>
</tr>
</tbody>
</table>

<sup>a</sup>CESD-11: 11-item Center for Epidemiologic Studies Depression (scale).

<sup>b</sup>Adjusted for all covariates.
Table 4. Results of generalized estimating equation model for subgroup analyzes of association of amount of pension benefit and proportion of pension in gross income with regard to pension type with depressive symptoms stratified by household income.

<table>
<thead>
<tr>
<th>Variables</th>
<th>CESD-11⁸ score, β (SE)</th>
<th>P value⁹</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amount of pension benefit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High-income households (≥ 60%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>National pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>~25% (US $1440/year)</td>
<td>−.694 (.215)</td>
<td>.001</td>
</tr>
<tr>
<td>~75% (US $2790/year)</td>
<td>−.449 (.165)</td>
<td>.007</td>
</tr>
<tr>
<td>~100% (US $29,400/year)</td>
<td>−.575 (.211)</td>
<td>.006</td>
</tr>
<tr>
<td>Specific corporate pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>~25% (US $16,800/year)</td>
<td>−.426 (.480)</td>
<td>.38</td>
</tr>
<tr>
<td>~75% (US $27,600/year)</td>
<td>−.432 (.312)</td>
<td>.17</td>
</tr>
<tr>
<td>~100% (US $44,880/year)</td>
<td>−.848 (.387)</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Low-income households (&lt;60%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>National pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>~25% (US $1440/year)</td>
<td>−.663 (.195)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>~75% (US $2790/year)</td>
<td>−.861 (.162)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>~100% (US $29,400/year)</td>
<td>−1.472 (.266)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specific corporate pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>~25% (US $16,800/year)</td>
<td>−.782 (.531)</td>
<td>.14</td>
</tr>
<tr>
<td>~75% (US $27,600/year)</td>
<td>−3.646 (.642)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>~100% (US $44,880/year)</td>
<td>Undetectable</td>
<td>N/A⁶</td>
</tr>
<tr>
<td><strong>Proportion of pension income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High-income households (≥ 60%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>National pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0&lt;pension income/gross income≤0.25</td>
<td>−.532 (.149)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>0.25&lt;pension income/gross income≤0.75</td>
<td>−.704 (.389)</td>
<td>.07</td>
</tr>
<tr>
<td>0.75&lt;pension income/gross income≤1.0</td>
<td>.357 (1.429)</td>
<td>.80</td>
</tr>
<tr>
<td>Specific corporate pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0&lt;pension income/gross income≤0.25</td>
<td>−.318 (.897)</td>
<td>.72</td>
</tr>
<tr>
<td>0.25&lt;pension income/gross income≤0.75</td>
<td>−.686 (.294)</td>
<td>.02</td>
</tr>
<tr>
<td>0.75&lt;pension income/gross income≤1.0</td>
<td>−.581 (.353)</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Low-income households (&lt;60%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>National pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0&lt;pension income/gross income≤0.25</td>
<td>−.625 (.150)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>0.25&lt;pension income/gross income≤0.75</td>
<td>−1.254 (.242)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>0.75&lt;pension income/gross income≤1.0</td>
<td>−2.783 (.835)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Specific corporate pension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0&lt;pension income/gross income≤0.25</td>
<td>.188 (1.400)</td>
<td>.89</td>
</tr>
<tr>
<td>0.25&lt;pension income/gross income≤0.75</td>
<td>−.443 (.743)</td>
<td>.55</td>
</tr>
</tbody>
</table>
the association between these circumstances, we evaluated the association between public pensions and mental health, as each country has different forms and types of pensions. Many European countries have studied the association between disability pensions and health outcomes [29]. Research in Europe has shown that receiving a disability pension can have a more negative impact on mental health than not receiving it. It is presumed that this phenomenon is due to the psychosocial stress experienced by beneficiaries in obtaining eligibility for disability pension. Additionally, claims stigma can have a negative impact on mental health [30].

Contrary to the findings in the aforementioned reports, our findings show a positive relationship between public pension as a means of maintaining income in old age and mental health. Recently, many studies in China have found that China’s new rural pension scheme reduces depressive symptoms and lowers the prevalence of depression [9,31,32]. The findings of a Japanese study on the relationship between contribution-based public pensions and happiness were also similar to our study results [19]. Furthermore, the results of a study analyzing the relationship between aged pension schemes and mental health conducted in New Zealand reported that material hardship can cause health inequities, including mental health [33]. These studies suggest that older adults can benefit their mental health by reducing labor supply and spending more leisure time with friends and family through the pension income. Additionally, pensions can reduce depressive symptoms by strengthening the financial capacity of beneficiaries and their families. However, there are limitations in that the existing studies [9,31,32] examined the effects of pension systems that were implemented in rural areas or had cross-sectional research designs. Under these circumstances, we evaluated the association between contributory public pensions and mental health after adjusting for each individual’s pension income type, considering the multilayered Korean pension system.

In this study, among older adults without a history of diagnosed depression, those receiving contribution-based public pensions, regardless of pension type, were less likely to experience depressive symptoms than those who did not receive public pensions. Additionally, in subgroup analyses, we observed that the CESD-11 scores tended to decrease as the amount of pension benefits increased for both national pension and specific corporate pension recipients. These results imply that the policy of guaranteeing old age income through public pensions was associated with a lower depressive score and that the larger the amount of public pension benefits, the better was the mental health.

Several studies have evaluated that pensions could affect mental health through the following pathways: first, through changes in lifestyle habits such as independent living and leisure time; second, by enabling healthy life choices such as with nutrition and medical treatment; and third, by reducing economic stress [9,34,35]. In addition to the 3 pathways mentioned above, the relation between contribution-based public pension receipt and alleviation of depressive symptoms can be explained by the following mechanisms. In the case of a contribution-based public pension, the result of one’s labor is returned later. Therefore, there is a possibility that even a small amount of money may make one feel satisfied, believing that the pension is a reward for one’s effort [36]. In particular, given the hedonic adaptation hypothesis, according to which the pleasure of economic benefits gradually disappears after receiving a noncontributory pension [37], the contribution-based pension has a different mechanism in relieving depressive symptoms in that it reflects the results of previous labor. Another possibility is that public pension recipients were economically active and stable when they were young, since they earned income from work or self-employment for at least 10 years or more, contributing to the pension. This may have lowered their likelihood of experiencing depressive symptoms. In addition, the experience of working long-term at a company or being self-employed to contribute to public pensions may have affected their social networks, which may ultimately influence their mental health [38]. However, no association with lower depressive symptoms was observed when the pension amount among national pension recipients was 75% or more of the gross income, although the CESD-11 score tended to decrease. Considering our research showing that the top 25% of national pension recipients receive only US $2790 per year (US $232.50 per month), there is a possible limitation to improving depressive symptoms merely if older adults receive public pension, requiring them to live with fewer pension benefits. This phenomenon is in line with the results of this study and previous studies, which indicate that receiving a low

### Discussion

#### Principal Results

This longitudinal study investigates whether receiving public pension, which plays a major role in the old age income security policy in Korea, was associated with depressive symptoms in older adults. In particular, we distinguished between national and specific corporate pensions, which are the 2 pillars of public pension in Korea and evaluated whether the 2 types of pensions could improve depressive symptoms. We found that the depression scores decreased as the amount of benefits increased for both public pensions. Regardless of the type, the association between public pension benefits and decreased depressive symptoms was pronounced in low-income households. The results of our study imply that financial resources provided through public pensions can improve the mental health of older people.

Conflicting findings have been reported regarding the relationship between pension income and health [27,28]. However, care must be taken when interpreting the relationship between pensions and health, as each country has different forms and types of pensions. Many European countries have studied the association between disability pensions and health outcomes [29]. Research in Europe has shown that receiving a disability pension can have a more negative impact on mental health than not receiving it. It is presumed that this phenomenon is due to the psychosocial stress experienced by beneficiaries in obtaining eligibility for disability pension. Additionally, claims stigma can have a negative impact on mental health [30].

Contrary to the findings in the aforementioned reports, our findings show a positive relationship between public pension as a means of maintaining income in old age and mental health. Recently, many studies in China have found that China’s new rural pension scheme reduces depressive symptoms and lowers the prevalence of depression [9,31,32]. The findings of a Japanese study on the relationship between contribution-based public pensions and happiness were also similar to our study results [19]. Furthermore, the results of a study analyzing the relationship between aged pension schemes and mental health conducted in New Zealand reported that material hardship can cause health inequities, including mental health [33]. These studies suggest that older adults can benefit their mental health by reducing labor supply and spending more leisure time with friends and family through the pension income. Additionally, pensions can reduce depressive symptoms by strengthening the financial capacity of beneficiaries and their families. However, there are limitations in that the existing studies [9,31,32] examined the effects of pension systems that were implemented in rural areas or had cross-sectional research designs. Under these circumstances, we evaluated the association between contributory public pensions and mental health after adjusting for each individual’s pension income type, considering the multilayered Korean pension system.

In this study, among older adults without a history of diagnosed depression, those receiving contribution-based public pensions, regardless of pension type, were less likely to experience depressive symptoms than those who did not receive public pensions. Additionally, in subgroup analyses, we observed that the CESD-11 scores tended to decrease as the amount of pension benefits increased for both national pension and specific corporate pension recipients. These results imply that the policy of guaranteeing old age income through public pensions was associated with a lower depressive score and that the larger the amount of public pension benefits, the better was the mental health.

Several studies have evaluated that pensions could affect mental health through the following pathways: first, through changes in lifestyle habits such as independent living and leisure time; second, by enabling healthy life choices such as with nutrition and medical treatment; and third, by reducing economic stress [9,34,35]. In addition to the 3 pathways mentioned above, the relation between contribution-based public pension receipt and alleviation of depressive symptoms can be explained by the following mechanisms. In the case of a contribution-based public pension, the result of one’s labor is returned later. Therefore, there is a possibility that even a small amount of money may make one feel satisfied, believing that the pension is a reward for one’s effort [36]. In particular, given the hedonic adaptation hypothesis, according to which the pleasure of economic benefits gradually disappears after receiving a noncontributory pension [37], the contribution-based pension has a different mechanism in relieving depressive symptoms in that it reflects the results of previous labor. Another possibility is that public pension recipients were economically active and stable when they were young, since they earned income from work or self-employment for at least 10 years or more, contributing to the pension. This may have lowered their likelihood of experiencing depressive symptoms. In addition, the experience of working long-term at a company or being self-employed to contribute to public pensions may have affected their social networks, which may ultimately influence their mental health [38]. However, no association with lower depressive symptoms was observed when the pension amount among national pension recipients was 75% or more of the gross income, although the CESD-11 score tended to decrease. Considering our research showing that the top 25% of national pension recipients receive only US $2790 per year (US $232.50 per month), there is a possible limitation to improving depressive symptoms merely if older adults receive public pension, requiring them to live with fewer pension benefits. This phenomenon is in line with the results of this study and previous studies, which indicate that receiving a low

### Table

<table>
<thead>
<tr>
<th>Variables</th>
<th>CESD-11b score, β (SE)</th>
<th>P valueb</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75&lt;pension income/gross income≤1.0</td>
<td>−1.599 (.594)</td>
<td>.007</td>
</tr>
</tbody>
</table>

aCESD-11: 11-item Center for Epidemiologic Studies Depression (scale).
bAdjusted for all covariates.
cN/A: not applicable.
amount of noncontributing basic old age pension was not associated with decreased depressive symptoms [39]. Nevertheless, among low-income people, public pensions were associated with low CESD-11 scores, even for low amounts. These results suggest that for low-income people, even a small amount of national pension can improve their depressive symptoms and that the level of income should be considered when formulating a pension policy.

In our study, even for a specific corporate pensioner whose pension benefit amount was higher than that of the national pensioner, pension receipt was not associated with alleviation of depressive symptoms for individuals whose pension amount was in the bottom 25% or less than 25% of the gross income among specific corporate pensioners. Depressive symptoms may not decrease for those with low specific corporate pension benefits because they must earn income through other means. In Korea, the corporate pension amount is higher than the national pension amount. Thus, specific corporate pension beneficiaries do not have to worry about their postretirement income. Our results suggest that it is necessary to consider the types of public pensions and their implications for beneficiaries in addition to the pension receipt or amount, when studying the effects of public pensions on mental health.

The state’s finances are limited; thus, tax-based noncontributory pensions cannot be increased indefinitely. Moreover, it is almost impossible to expand noncontributory pensions, especially in low-income countries, which are forced to spend a considerable share of the state’s finances on economic development. Therefore, the few financial policies that can be used to improve the mental health of older adults are contribution-based pensions. The objective of the public pension policy is to guarantee a certain amount of benefits to individuals so that they do not experience economic difficulties when they become older. This can be achieved, for example, by increasing contributions or by extending the period of contributions when people are young.

Taken together, our findings may provide several important implications not only for the Korean society but also for countries planning and revising pension policies. First, our findings justify a wide range of policy interventions that promote public health based on retirement income security policies such as public pensions. Second, these findings suggest that mental health can be effectively improved with economic resources. Third, the best way to use economic resources to improve mental health is to ensure an income that is economically sufficient. However, creating this condition without contributions can be difficult. Fourth, for low-income countries that cannot afford to preserve older adults’ income, a contribution-based public pension system may be an important policy in order to maintain the mental health of older adults after retirement.

Limitations
This study has several limitations. First, the exact causal relationship between public pension benefit and depressive symptoms could not be confirmed, although this study used longitudinal data with repeated observations at the individual level over a specific period [40]. Second, residual confounding may exist because of income-related factors. Finally, since the economic and cultural basis that determines the pension policy and the direction of policy in each country may differ, direct application of the results of this study may be limited [41]. Longitudinal studies in more countries and based on larger numbers of public pension recipients are needed to confirm causal relationships.

Conclusions
This study evaluates both the national and specific corporate pensions, which are the 2 pillars of public pensions, and their association with lower prevalence of depressive symptoms in Korea. The depression scores decreased as the amount of benefits increased for both public pensions. However, for beneficiaries of specific corporate pensions with lower benefits, the association with depressive symptoms was absent. Regardless of the pension type, the association between public pension benefits and decreased depressive symptoms was pronounced in low-income households. For low-income countries that cannot afford to preserve older adults’ income, the contribution-based public pension system may be an important policy to maintain the mental health of older adults after retirement.

Acknowledgments
This work was supported by a fund from the research program of the Korea Medical Institute (funding code: 23-042) and the Soonchunhyang University Research Fund.

Data Availability
All the Korea Welfare Panel Study data used in this study are available to the public and can be seen in the official website of the Korea Welfare Panel Study.

Authors’ Contributions
SHK and ECP were responsible for the conception and design of this study. SHK, HK, and SHJ curated the data, and SHK contributed to analysis and interpretation of the data. SHK drafted the manuscript. ECP performed the writing of the review and editing, and all authors read and approved the final version of this manuscript.
Conflicts of Interest

None declared.

References


Ebbinghaus B. Pension reforms and old age inequalities in Europe: From old to new social risks? 2019 Presented at: European Sociological Association Conference; August 21; Manchester.


Abbreviations

CESD-11: 11-item Center for Epidemiologic Studies Depression (scale)
GEE: generalized estimating equation
KoWePS: Korea Welfare Panel Study
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
Risk of Subsequent Primary Cancers Among Adult-Onset 5-Year Cancer Survivors in South Korea: Retrospective Cohort Study

Yoon Young Choi¹*, MD, PhD; Myeongjee Lee²*, PhD; Eun Hwa Kim², MS; Jae Eun Lee³, PhD; Inkyung Jung²*, PhD; Jae-Ho Cheong³, MD, PhD

¹Department of Surgery, Soonchunhyang Bucheon Hospital, Soonchunhyang University College of Medicine, Bucheon si, Republic of Korea
²Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea
³Department of Surgery, Yonsei University Health System, Yonsei University College of Medicine, Seoul, Republic of Korea
*these authors contributed equally

Corresponding Author:
Jae-Ho Cheong, MD, PhD
Department of Surgery
Yonsei University Health System
Yonsei University College of Medicine
Yonsei-ro 50
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 2094
Email: jhcheong@yuhs.ac

Abstract

Background: The number of cancer survivors who develop subsequent primary cancers (SPCs) is expected to increase.

Objective: We evaluated the overall and cancer type–specific risks of SPCs among adult-onset cancer survivors by first primary cancer (FPC) types considering sex and age.

Methods: We conducted a retrospective cohort study using the Health Insurance Review and Assessment database of South Korea including 5-year cancer survivors diagnosed with an FPC in 2009 to 2010 and followed them until December 31, 2019. We measured the SPC incidence per 10,000 person-years and the standardized incidence ratio (SIR) compared with the incidence expected in the general population.

Results: Among 266,241 survivors (mean age at FPC: 55.7 years; 149,352/266,241, 56.1% women), 7348 SPCs occurred during 1,003,008 person-years of follow-up (median 4.3 years), representing a 26% lower risk of developing SPCs (SIR 0.74, 95% CI 0.72-0.76). Overall, men with 14 of the 20 FPC types had a significantly lower risk of developing any SPCs; women with 7 of the 21 FPC types had a significantly lower risk of developing any SPCs. The risk of developing any SPC type differed by age; the risk was 28% higher in young (<40 years) cancer survivors (SIR 1.28, 95% CI 1.16-1.42; incidence: 30 per 10,000 person-years) and 27% lower in middle-aged and older (≥40 years) cancer survivors (SIR 0.73, 95% CI 0.71-0.74; incidence: 80 per 10,000 person-years) compared with the age-corresponding general population. The most common types of FPCs were mainly observed as SPCs in cancer survivors, with lung (21.6%) and prostate (15.2%) cancers in men and breast (18.9%) and lung (12.2%) cancers in women. The risks of brain cancer in colorectal cancer survivors, lung cancer in laryngeal cancer survivors, and both kidney cancer and leukemia in thyroid cancer survivors were significantly higher for both sexes. Other high-risk SPCs varied by FPC type and sex. Strong positive associations among smoking-related cancers, such as laryngeal, head and neck, lung, and esophageal cancers, were observed. Substantial variation existed in the associations between specific types of FPC and specific types of SPC risk, which may be linked to hereditary cancer syndrome: for women, the risks of ovarian cancer for breast cancer survivors and uterus cancers for colorectal cancer survivors, and for men, the risk of pancreas cancer for kidney cancer survivors.

Conclusions: The varying risk for SPCs by age, sex, and FPC types in cancer survivors implies the necessity for tailored prevention and screening programs targeting cancer survivors. Lifestyle modifications, such as smoking cessation, are essential to reduce the risk of SPCs in cancer survivors. In addition, genetic testing, along with proactive cancer screening and prevention strategies, should be implemented for young cancer survivors because of their elevated risk of developing SPCs.

(JMir Public Health Surveill 2024;10:e48380) doi: 10.2196/48380
cancer; survivors; subsequent primary cancer; adult; onset; primary cancer; risk; general population; screening; genetic testing; retrospective; cohort study; health Insurance; survivability; hereditary; FPC; SPC; subsequent cancer

**Introduction**

The number of cancer survivors is on the rise due to population growth, an aging population, and improved survival rates resulting from early diagnosis and advanced cancer treatment strategies [1]. As of 2019, there were approximately 16.9 million and 2 million cancer survivors in the United States and South Korea, respectively [2]. Managing the health of cancer survivors is critical because they face unique physical, psychosocial, and medical conditions, with the more serious being subsequent primary cancers (SPCs) [1,3]. Identifying the associations between the types of SPC and first primary cancer (FPC) is crucial, as cancer survivors continue to live in similar environments to where their FPC developed and the genetic risk persists. Identification of and preemptive intervention for these associations provide opportunities for risk reduction and early detection through tailored prevention and cancer screening programs.

A comprehensive study investigating SPCs among 5-year cancer survivors using the Surveillance, Epidemiology, and End Results (SEER) Program reported a higher risk of SPCs with several types of FPCs [4]. However, the SEER study had a predominantly White population (>80%), and over 40% of male and female participants were prostate and breast cancer survivors, respectively. In Korea, previous studies have mainly focused on specific types of FPCs, rather than comprehensively addressing cancer survivors in general [5-8]. Therefore, given the differences in ethnicity and cancer incidence between the United States and South Korea, a comprehensive epidemiological investigation is necessary to provide new insight on developing SPCs in cancer survivors with diverse socioenvironmental and different ethnic-genetic interactions.

The occurrence of cancers is influenced by environmental and hereditary factors, as well as random mutations [9]. Hereditary factors are linked to young-onset and multiple cancers [10-12]. Consequently, young cancer survivors may be at a higher risk of developing SPCs than their older counterparts and may exhibit unique combinations of FPCs and SPCs that are influenced by genetic susceptibility. Thus, age should be taken into consideration when investigating the association between FPCs and SPCs; however, this aspect has not been assessed in previous studies.

In this study, we assessed the relationships between FPCs and SPCs by analyzing data obtained from the South Korea Health Insurance Review and Assessment (HIRA) database, which covers nearly the entire Korean population.

**Methods**

**Data Source and Study Population**

The study was a nationwide, population-based, cohort study that analyzed data obtained from the HIRA database of South Korea between January 1, 2008, and December 31, 2019, which includes over 50 million Koreans. The study population included patients diagnosed with cancer between January 1, 2009, and December 31, 2010, and newly diagnosed cancer cases during this period were identified by excluding patients diagnosed with cancer in 2008. Cancers were classified into 23 types (20 types for men and 21 types for women) using the 10th revision of the International Classification of Diseases (ICD) code, and the Korean Individual Co-payment Beneficiaries Program (ICBP) data in the HIRA were used to define each cancer type. The ICBP was established for rare and intractable diseases, including cancers, in 2008, and patients in this program pay only 5% to 10% of their medical costs for 5 years from the day of registration. Therefore, the most affected individuals are registered in ICBP, and patients who were diagnosed with cancer before 2008 were retrospectively registered in the program. We followed those cancer patients until December 31, 2019. Cancer survivors were defined as cancer patients who lived longer than 5 years without developing other cancers with no evidence of recurrence of the first diagnosed cancers. SPCs were defined as a newly diagnosed, distinct cancer type, excluding same-type cancers to minimize the bias from misclassification of FPC recurrence. Survivor follow-up began 5 years after their FPC diagnosis and continued until death; loss to follow-up; December 31, 2019; or occurrence of an SPC, whichever came first. If a patient had multiple SPCs, the first type was considered the SPC. The primary outcomes were incidence (per 10,000 person-years) and relative risk (standardized incidence ratio [SIR]) of SPCs among 5-year survivors of FPCs. The Korean National Cancer Statistics 2015 [13], the middle year of the present cohort, was used to determine the expected cancer rates.

**Statistical Analysis**

Incidence rates of SPCs among the overall sample and each group of survivors were calculated by dividing the observed number of SPCs by the corresponding total person-years and then multiplying by 10,000. To investigate the risk of SPCs among the overall sample and each group of survivors compared with the general Korean population, we calculated the age-standardized incidence ratio by type of FPCs and sex using the Korean National Cancer Statistics 2015 [13], the middle year of the cohort, as the expected cancer rates. The 95% CI of the SIR was calculated using the Poisson distribution. All estimates were calculated for overall SPCs and for type-specific SPCs. The 12 smoking-related cancers, 12 obesity-related cancers, 7 alcohol-related cancers, and 6 infection-related cancers were considered in the subgroup analyses to assess environmental factors for developing SPCs [14-16]. The incidences and SIRs were estimated to quantify the risk of developing any types of each environment factor–related SPC. A 2-sided P<.05 was considered statistically significant; multiple comparison corrections were not conducted, as the analyses were exploratory. SAS Enterprise Guide version 9.4 (SAS Institute Inc) was used for all statistical analyses.
Ethical Considerations
This study was approved by the research ethics committee of the South Korea National Health Insurance Sharing Service (M20200902739) and the institutional review board of Severance Hospital (4-2020-0155). The need for informed consent was waived owing to the use of deidentified data. Access to the HIRA data is restricted to authorized individuals and is only permissible from designated computers.

Results
A total of 266,241 (71.7%) 5-year cancer survivors were identified among 371,181 patients who were diagnosed with FPCs between 2009 and 2010, accruing 1,003,008 person-years of follow-up 5 years after their FPCs (mean 3.8 years; median 4.3 years; Figure 1 and Table 1). Among 371,181 patients who were diagnosed with an FPC between January 2009 and December 2010 in South Korea, there were 266,241 (71.7%) 5-year cancer survivors, and among the survivors, 7348 (2.8%) patients had SPCs. The percentage of 5-year survivors varied by the type of FPC; survival was lowest with pancreatic cancer (2628/7392, 35.6%) and highest with thyroid cancer (66,087/69,211, 95.5%; Table S1 in Multimedia Appendix 1). Of the survivors, 56.1% (149,352/266,241) were female, and 12.5% (33,171/266,241) and 30.1% (80,249/266,241) were younger than 40 years and older than 65 years, respectively.

Among all survivors, 7348 SPC cases (73 per 10,000 person-years) were identified, representing a 26% lower risk of developing SPCs (SIR 0.74, 95% CI 0.72-0.76; Table 2). There were 4207 and 3141 SPC cases (99 and 54 per 10,000 person-years) in male and female survivors, respectively, which corresponded to 34% and 11% lower risks of developing SPCs compared with the risk in the general Korean population. When considering age at FPC diagnosis, the risk of developing SPCs differed by age, with a higher risk observed among young patients with FPCs (<40 years old), and the risk gradually decreased with age (Table S2 in Multimedia Appendix 1). Therefore, subgroups by age were considered for further analysis. The risk of developing SPCs was significantly higher when the patient had FPC before they reached 40 years old (men: SIR 1.43, 95% CI 1.10-1.84; women: SIR 1.26, 95% CI 1.12-1.40), while the risk was significantly lower in patients with an FPC at ≥40 years old (men: SIR 0.66, 95% CI 0.64-0.68; women: SIR 0.86, 95% CI, 0.83-0.89).

Overall, many survivor groups defined by FPC types and sex had lower-than-expected risks of developing SPCs compared with the risks in the Korean general population. Specifically, men with 14 of the 20 FPC types had a significantly lower risk of developing any SPCs, while women with 7 of the 21 FPC types had a significantly lower risk of developing any SPCs (Table 3).

Figure 1. Flow diagram for patient selection.
Table 1. Characteristics of survivors diagnosed with a first primary cancer in 2009-2010 at age ≥18 years among 5-year survivors in South Korea.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Survivors, n (%)</th>
<th>Person-years of follow-up at 5 years after the first primary cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Mean</td>
</tr>
<tr>
<td>Total sample</td>
<td>266,241 (100)</td>
<td>1,003,008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.3</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>116,889 (43.9)</td>
<td>424,836</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.2</td>
</tr>
<tr>
<td>Women</td>
<td>149,352 (56.1)</td>
<td>578,172</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.3</td>
</tr>
<tr>
<td>Age at first primary cancer diagnosis (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-39</td>
<td>33,171 (12.5)</td>
<td>128,648</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.3</td>
</tr>
<tr>
<td>40-49</td>
<td>54,689 (20.5)</td>
<td>210,840</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.3</td>
</tr>
<tr>
<td>50-64</td>
<td>98,132 (36.9)</td>
<td>367,554</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.3</td>
</tr>
<tr>
<td>≥65</td>
<td>80,249 (30.1)</td>
<td>295,966</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.2</td>
</tr>
<tr>
<td>Calendar year of first primary cancer diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>135,195 (50.8)</td>
<td>574,434</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.9</td>
</tr>
<tr>
<td>2010</td>
<td>131,046 (49.2)</td>
<td>428,574</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.8</td>
</tr>
</tbody>
</table>

Table 2. Risk of developing any type of subsequent primary cancer (SPC) among 5-year cancer survivors with their first diagnosis between 2009 and 2010 in South Korea.

<table>
<thead>
<tr>
<th>Category</th>
<th>Survivors, n (%)</th>
<th>Observed SPCs, n (%)</th>
<th>Expected SPCs, n</th>
<th>Incidence per 10,000 person-years</th>
<th>SIRa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at FPC &lt;40 years</td>
<td>33,171 (12.5)</td>
<td>381 (5.2)</td>
<td>297</td>
<td>30</td>
<td>1.28 (1.16-1.42)</td>
</tr>
<tr>
<td>Age at FPC ≥40 years</td>
<td>233,070 (87.5)</td>
<td>6,967 (94.8)</td>
<td>9614</td>
<td>80</td>
<td>0.73 (0.71-0.74)</td>
</tr>
<tr>
<td>Total</td>
<td>266,241 (100)</td>
<td>7348 (100)</td>
<td>9912</td>
<td>73</td>
<td>0.74 (0.72-0.76)</td>
</tr>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men with age at FPC &lt;40 years</td>
<td>8219 (7)</td>
<td>62 (1.5)</td>
<td>43</td>
<td>20</td>
<td>1.43 (1.10-1.84)</td>
</tr>
<tr>
<td>Men with age at FPC ≥40 years</td>
<td>10,867 (93)</td>
<td>4145 (98.5)</td>
<td>6326</td>
<td>106</td>
<td>0.66 (0.64-0.68)</td>
</tr>
<tr>
<td>Total</td>
<td>116,889 (43.9)</td>
<td>4207 (57.3)</td>
<td>6369</td>
<td>99</td>
<td>0.66 (0.64-0.68)</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women with age at FPC &lt;40 years</td>
<td>24,952 (16.7)</td>
<td>319 (10.2)</td>
<td>254</td>
<td>33</td>
<td>1.26 (1.12-1.40)</td>
</tr>
<tr>
<td>Women with age at FPC ≥40 years</td>
<td>124,400 (83.3)</td>
<td>2822 (89.8)</td>
<td>3289</td>
<td>59</td>
<td>0.86 (0.83-0.89)</td>
</tr>
<tr>
<td>Total</td>
<td>149,352 (56.1)</td>
<td>3141 (42.7)</td>
<td>3542</td>
<td>54</td>
<td>0.89 (0.86-0.92)</td>
</tr>
</tbody>
</table>

aSIR: standardized incidence ratio.
bFPC: first primary cancer.
Table 3. Risk of developing any type of subsequent primary cancers (SPCs) among 5-year survivors of a first primary cancer (FPC), with their first diagnosis between 2009 and 2010 in South Korea, by the type of their FPC.

<table>
<thead>
<tr>
<th>FPC</th>
<th>Survivors, n (%)</th>
<th>Age at FPC (years), mean</th>
<th>Age at SPC (years), mean</th>
<th>Observed SPCs, n (%)</th>
<th>Incidence per 10,000 person-years</th>
<th>Expected SPCs, n</th>
<th>SIR* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head and neck</td>
<td>2315 (2)</td>
<td>56.4</td>
<td>68.4</td>
<td>95 (2.3)</td>
<td>108</td>
<td>126</td>
<td>0.75 (0.61-0.92)</td>
</tr>
<tr>
<td>Esophagus</td>
<td>1458 (1.2)</td>
<td>64.9</td>
<td>73.3</td>
<td>63 (1.5)</td>
<td>111</td>
<td>120</td>
<td>0.52 (0.40-0.67)</td>
</tr>
<tr>
<td>Stomach</td>
<td>30,165 (25.8)</td>
<td>60.4</td>
<td>71.9</td>
<td>1286 (30.6)</td>
<td>108</td>
<td>1750</td>
<td>0.73 (0.70-0.78)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>22,008 (18.8)</td>
<td>60.8</td>
<td>71.6</td>
<td>1071 (25.5)</td>
<td>126</td>
<td>1322</td>
<td>0.81 (0.76-0.86)</td>
</tr>
<tr>
<td>Liver</td>
<td>10,116 (8.7)</td>
<td>57.8</td>
<td>67.7</td>
<td>129 (3.1)</td>
<td>42</td>
<td>438</td>
<td>0.30 (0.25-0.35)</td>
</tr>
<tr>
<td>Gallbladder and biliary</td>
<td>1334 (1.1)</td>
<td>64.4</td>
<td>71.0</td>
<td>44 (1)</td>
<td>85</td>
<td>105</td>
<td>0.42 (0.30-0.56)</td>
</tr>
<tr>
<td>Pancreas</td>
<td>1420 (1.2)</td>
<td>63.4</td>
<td>69.7</td>
<td>26 (0.6)</td>
<td>44</td>
<td>116</td>
<td>0.23 (0.15-0.33)</td>
</tr>
<tr>
<td>Larynx</td>
<td>1438 (1.2)</td>
<td>62.9</td>
<td>72.3</td>
<td>132 (3.1)</td>
<td>248</td>
<td>103</td>
<td>1.28 (1.07-1.51)</td>
</tr>
<tr>
<td>Lung</td>
<td>10,349 (8.9)</td>
<td>65.5</td>
<td>72.3</td>
<td>158 (3.8)</td>
<td>40</td>
<td>712</td>
<td>0.22 (0.19-0.26)</td>
</tr>
<tr>
<td>Breast</td>
<td>111 (0.1)</td>
<td>56.5</td>
<td>73.4</td>
<td>8 (0.2)</td>
<td>192</td>
<td>6</td>
<td>1.34 (0.58-2.64)</td>
</tr>
<tr>
<td>Prostate</td>
<td>10,628 (9.1)</td>
<td>68.0</td>
<td>76.2</td>
<td>362 (8.6)</td>
<td>132</td>
<td>566</td>
<td>0.64 (0.58-0.71)</td>
</tr>
<tr>
<td>Testis</td>
<td>408 (0.3)</td>
<td>33.3</td>
<td>55.8</td>
<td>6 (0.1)</td>
<td>38</td>
<td>4</td>
<td>1.50 (0.55-3.27)</td>
</tr>
<tr>
<td>Kidney</td>
<td>3482 (3.0)</td>
<td>55.4</td>
<td>67.7</td>
<td>156 (3.7)</td>
<td>127</td>
<td>163</td>
<td>0.96 (0.81-1.12)</td>
</tr>
<tr>
<td>Bladder</td>
<td>5033 (4.3)</td>
<td>63.0</td>
<td>74.2</td>
<td>252 (6.0)</td>
<td>155</td>
<td>307</td>
<td>0.82 (0.72-0.93)</td>
</tr>
<tr>
<td>Brain</td>
<td>1687 (1.4)</td>
<td>46.9</td>
<td>64.8</td>
<td>23 (0.5)</td>
<td>40</td>
<td>60</td>
<td>0.38 (0.24-0.58)</td>
</tr>
<tr>
<td>Thyroid</td>
<td>10,515 (9)</td>
<td>46.5</td>
<td>64.2</td>
<td>300 (7.1)</td>
<td>75</td>
<td>287</td>
<td>1.05 (0.93-1.17)</td>
</tr>
<tr>
<td>Hodgkin</td>
<td>164 (0.1)</td>
<td>43.1</td>
<td>74.7</td>
<td>3 (0.1)</td>
<td>47</td>
<td>5</td>
<td>0.60 (0.12-1.75)</td>
</tr>
<tr>
<td>Non-Hodgkin</td>
<td>2358 (2)</td>
<td>52.9</td>
<td>66.8</td>
<td>76 (1.8)</td>
<td>87</td>
<td>110</td>
<td>0.69 (0.55-0.87)</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>491 (0.4)</td>
<td>61.5</td>
<td>64.3</td>
<td>3 (0.1)</td>
<td>27</td>
<td>23</td>
<td>0.13 (0.03-0.38)</td>
</tr>
<tr>
<td>Leukemia</td>
<td>1409 (1.2)</td>
<td>47.5</td>
<td>62.0</td>
<td>14 (0.3)</td>
<td>34</td>
<td>46</td>
<td>0.31 (0.17-0.51)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>116,889 (100)</td>
<td>59.6</td>
<td>71.2</td>
<td>4207 (100)</td>
<td>99</td>
<td>6369</td>
<td>0.66 (0.64-0.68)</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head and neck</td>
<td>1176 (0.8)</td>
<td>54.5</td>
<td>63.1</td>
<td>25 (0.8)</td>
<td>57</td>
<td>33</td>
<td>0.76 (0.49-1.13)</td>
</tr>
<tr>
<td>Esophagus</td>
<td>138 (0.1)</td>
<td>64.9</td>
<td>64.0</td>
<td>4 (0.1)</td>
<td>74</td>
<td>5</td>
<td>0.75 (0.21-1.93)</td>
</tr>
<tr>
<td>Stomach</td>
<td>15,182 (10.2)</td>
<td>60.2</td>
<td>70.2</td>
<td>366 (11.7)</td>
<td>60</td>
<td>475</td>
<td>0.77 (0.69-0.86)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>14,610 (9.8)</td>
<td>61.9</td>
<td>69.7</td>
<td>394 (12.5)</td>
<td>68</td>
<td>448</td>
<td>0.88 (0.80-0.97)</td>
</tr>
<tr>
<td>Liver</td>
<td>3062 (2.1)</td>
<td>61.9</td>
<td>69.5</td>
<td>41 (1.3)</td>
<td>43</td>
<td>84</td>
<td>0.49 (0.35-0.66)</td>
</tr>
<tr>
<td>Gallbladder and biliary</td>
<td>1349 (0.9)</td>
<td>65.8</td>
<td>70.9</td>
<td>30 (1)</td>
<td>57</td>
<td>51</td>
<td>0.59 (0.40-0.85)</td>
</tr>
<tr>
<td>Pancreas</td>
<td>1208 (0.8)</td>
<td>64.7</td>
<td>67.8</td>
<td>12 (0.4)</td>
<td>24</td>
<td>48</td>
<td>0.25 (0.13-0.44)</td>
</tr>
<tr>
<td>Larynx</td>
<td>102 (0.1)</td>
<td>61.6</td>
<td>70.6</td>
<td>7 (0.2)</td>
<td>186</td>
<td>3</td>
<td>2.03 (0.81-4.17)</td>
</tr>
<tr>
<td>Lung</td>
<td>4210 (2.8)</td>
<td>63.9</td>
<td>72.3</td>
<td>49 (1.6)</td>
<td>33</td>
<td>128</td>
<td>0.38 (0.28-0.51)</td>
</tr>
<tr>
<td>Breast</td>
<td>29,866 (20)</td>
<td>50.6</td>
<td>62.6</td>
<td>599 (19.1)</td>
<td>53</td>
<td>608</td>
<td>0.99 (0.91-1.07)</td>
</tr>
<tr>
<td>Cervix</td>
<td>8901 (6)</td>
<td>51.9</td>
<td>63.3</td>
<td>222 (7.1)</td>
<td>60</td>
<td>249</td>
<td>0.89 (0.78-1.02)</td>
</tr>
<tr>
<td>Uterus</td>
<td>2529 (1.7)</td>
<td>51.7</td>
<td>61.5</td>
<td>80 (2.5)</td>
<td>79</td>
<td>68</td>
<td>1.18 (0.94-1.47)</td>
</tr>
<tr>
<td>Ovary</td>
<td>3089 (2.1)</td>
<td>48.7</td>
<td>59.8</td>
<td>80 (2.5)</td>
<td>72</td>
<td>68</td>
<td>1.18 (0.93-1.46)</td>
</tr>
<tr>
<td>Kidney</td>
<td>1739 (1.2)</td>
<td>57.7</td>
<td>69.8</td>
<td>54 (1.7)</td>
<td>84</td>
<td>52</td>
<td>1.03 (0.77-1.35)</td>
</tr>
<tr>
<td>Bladder</td>
<td>1146 (0.8)</td>
<td>65.3</td>
<td>72.5</td>
<td>37 (1.2)</td>
<td>94</td>
<td>39</td>
<td>0.95 (0.67-1.31)</td>
</tr>
<tr>
<td>Brain</td>
<td>1813 (1.2)</td>
<td>52.3</td>
<td>61.2</td>
<td>40 (1.3)</td>
<td>65</td>
<td>46</td>
<td>0.88 (0.63-1.19)</td>
</tr>
<tr>
<td>Thyroid</td>
<td>55,572 (37.2)</td>
<td>47.4</td>
<td>58.9</td>
<td>1022 (32.5)</td>
<td>47</td>
<td>1,053</td>
<td>0.97 (0.91-1.03)</td>
</tr>
</tbody>
</table>
Among FPC types, the risk of developing SPCs was significantly higher after laryngeal cancer in men (SIR 1.28, 95% CI 1.07-1.51; incidence: 248 per 10 000 person-years). For the 3 most common FPC types in 5-year cancer survivors, the overall SIRs in men were significantly lower for stomach, colorectal, and prostate cancers (Table 3). In women, the SIR was significantly lower for stomach cancer (SIR 0.77, 95% CI 0.69-0.86; incidence, 60 per 10 000 person-years) but not significantly lower for breast cancer and thyroid cancer.

Tables S3 and S4 in Multimedia Appendix 1 show the subgroup analyses by age (<40 and ≥40 years) for the risks of developing SPCs by FPC types in men and women. The risk of developing any type of SPC after kidney cancer (SIR 3.02, 95% CI 1.21-6.21) and thyroid cancer (SIR 1.37, 95% CI 1.18-1.58) was significantly higher in young (<40 years) men and women, respectively, than in the general population.

Figures 2 and 3 display the percentage contributions of different FPC and SPC combinations to the total incidence of SPCs, as well as the SIRs for men and women. Among all survivors, the 5 SPCs that contributed the most to the total incidence of SPCs were lung (21.6%), prostate (15.2%), stomach (12%), colorectal (9.5%), and liver (8.9%) cancers in men and breast (18.9%), lung (12.2%), stomach (10.4%), colorectal (10%), and thyroid (8.2%) cancers in women. These findings are represented in the row labeled “All FPCs” in Figure 2A and Figure 3A. In men, the 5 most significant SIRs for SPCs were observed among survivors of non-Hodgkin lymphoma for Hodgkin lymphoma (SIR 20.6), survivors of breast cancer for prostate cancer (SIR 6.07), survivors of esophageal cancer for head and neck cancer (SIR 5.62), laryngeal cancer survivors for head and neck cancer (SIR 5.19), and survivors of head and neck cancer survivors for esophageal cancer (SIR 4.27). These SIRs are shown in Figure 2B and Multimedia Appendix 2. For women, the 5 most significant SIRs for SPCs were observed among survivors of multiple myeloma for leukemia (SIR 15.11), survivors of laryngeal cancer for lung cancer (SIR 7.83), survivors of leukemia for non-Hodgkin lymphoma (SIR 7.11), survivors of non-Hodgkin lymphoma for leukemia (SIR 5.95), and survivors of colorectal cancer for uterus cancer (SIR 4.25). These findings are shown in Figure 3B and Multimedia Appendix 2. Figure 4 summarizes the heightened risk of SPCs across various FPC and SPC combinations in both men and women. Notably, the risks of brain cancer in colorectal cancer survivors, lung cancer in laryngeal cancer survivors, and both kidney cancer and leukemia in thyroid cancer survivors were significantly higher in both sexes.

The subgroup analysis to assess environmental factors for developing SPCs showed that the risks of developing each factor-related SPC were higher in young (<40 years old) patients with FPCs than older patients in general (Table S5 in Multimedia Appendix 1). Among middle-aged and older (≥40 years) survivors, some types of FPCs were associated with a higher risk of developing SPCs compared with the general population; for example, male survivors of laryngeal cancer had a higher risk of smoking-related SPCs (SIR 1.46, 95% CI 1.20-1.77), and survivors of uterus cancer had a higher risk of obesity-related SPCs (SIR 1.33, 95% CI 1.02-1.71). The results of additional subgroup analyses to assess the contribution of each FPC and SPC to the SIRs by age (<50 years and ≥50 years) in men and women are displayed in Figures S1 and S2 in Multimedia Appendix 3. Some considerable associations between the type of FPC and SPC that could be related to hereditary cancer syndromes such as hereditary breast and ovarian cancer, Lynch syndrome, and Von-Hippel-Lindau (VHL) syndrome were observed in patients with an FPC before 50 years of age (eg, uterus cancer in female breast cancer survivors, uterus cancer in female colorectal cancer survivors, pancreas cancer in male kidney cancer survivors; see Figures S1A and S2A in Multimedia Appendix 3).

---

**Table 3**

<table>
<thead>
<tr>
<th>FPC</th>
<th>Survivors, n (%)</th>
<th>Age at FPC (years), mean</th>
<th>Age at SPC (years), mean</th>
<th>Observed SPCs, n (%)</th>
<th>Incidence per 10,000 person-years</th>
<th>Expected SPCs, n</th>
<th>SIR* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hodgkin</td>
<td>82 (0.1)</td>
<td>38.4</td>
<td>47.5</td>
<td>2 (0.1)</td>
<td>61</td>
<td>1</td>
<td>1.37 (0.17-4.94)</td>
</tr>
<tr>
<td>Non-Hodgkin</td>
<td>1991 (1.3)</td>
<td>53.1</td>
<td>62.1</td>
<td>57 (1.8)</td>
<td>76</td>
<td>53</td>
<td>1.07 (0.81-1.39)</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>457 (0.3)</td>
<td>62.1</td>
<td>67.3</td>
<td>3 (0.1)</td>
<td>29</td>
<td>10</td>
<td>0.30 (0.06-0.87)</td>
</tr>
<tr>
<td>Leukemia</td>
<td>1130 (0.8)</td>
<td>49.0</td>
<td>58.5</td>
<td>17 (0.5)</td>
<td>52</td>
<td>21</td>
<td>0.80 (0.47-1.28)</td>
</tr>
<tr>
<td>Total</td>
<td>149,352 (100)</td>
<td>52.7</td>
<td>63.7</td>
<td>3141 (100)</td>
<td>54</td>
<td>3,542</td>
<td>0.89 (0.86-0.92)</td>
</tr>
</tbody>
</table>

*SIR: standardized incidence ratio.*
Figure 2. Risk of developing subsequent primary cancers (SPCs) among 5-year male survivors (A) calculated by dividing the observed number of SPCs of each cell by the total number of observed SPCs and the (B) standardized incidence ratios of SPCs by each first primary cancer (FPC). Statistically significant associations between the FPC and SPC are shown by bold boxes, gray cells indicate associations not tested due to a small number of observed SPCs, and blue and red colors represent lower- and higher-than-expected values, respectively, based on 195 eligible statistical tests (observed number of SPCs ≥5 or statistically significant association). GB: gallbladder.

Figure 3. Risk of developing subsequent primary cancers (SPCs) among 5-year female survivors (A) calculated by dividing the observed number of SPCs of each cell by the total number of observed SPCs and the (B) standardized incidence ratios of SPCs by each first primary cancer (FPC). Statistically significant associations between the FPC and SPC are shown by bold boxes, gray cells indicate associations not tested due to a small number of observed SPCs, and blue and red colors represent lower- and higher-than-expected values, respectively, based on 147 eligible statistical tests (observed number of SPCs ≥5 or statistically significant association). GB: gallbladder.
Discussion

Principal Findings

The results indicate that the risk of developing SPCs among adult-onset 5-year cancer survivors in South Korea is generally lower than in the general population. However, young (<40 years old) cancer survivors exhibited an increased risk of developing SPCs, and the risk of SPCs also varied according to the patient’s sex and the type of FPC. Survivors from smoking-related cancers have a significant risk of developing subsequent smoking-related cancers. These findings highlight the necessity for tailored approaches to cancer screening, prevention, and management strategies for cancer survivors.

Environmental factors such as smoking, alcohol use, exposure to ultraviolet light, and viral infections are well-known risk factors for the development of cancer [17]. In addition, an estimated 5% to 10% percent of cancers have a heritable component that is caused by genetic susceptibility [18]. Additionally, Tomasetti and Vogelstein [9,19] have suggested that “bad luck,” random mutations occurring during stem cell division, contributes to two-thirds of the adult cancer incidence; based on the results of their statistical model, there is a high correlation between the number of stem cell divisions of a given tissue and the lifetime risk of cancer in that tissue. The epidemiological results indicate a 26% lower risk of developing SPCs (SIR 0.74, 95% CI 0.72-0.76) in the overall sample of Korean cancer survivors when compared with the risk of developing cancer in the general population. This can be explained by the fact that cancer-causing mutations are stochastic in nature; therefore, the probability of “very bad luck,” the development of a SPC after an FPC in cancer survivors, is lower than “bad luck” in the general Korean population.

Although the overall risk of developing SPCs in cancer survivors is low, a high risk of developing well-known smoking-related cancers [20,21] was observed among cancer survivors with...
smoking-related cancers, including laryngeal, head and neck, and esophageal cancers in men and laryngeal and lung cancers in women. This result highlights the importance of environmental factors in the occurrence of SPCs, and primary prevention should be a priority for controlling both FPCs and SPCs, as a substantial cancer burden could be prevented through lifestyle modifications [22]. Smoking is one of the strongest carcinogens, causing various types of cancer, and quitting smoking has been shown to reduce the risk of cancer and improve survival even after a cancer diagnosis [23,24]. Therefore, it is crucial for cancer survivors to quit smoking to improve their survival and lower their risk of subsequent cancers. However, smoking cessation interventions for cancer survivors have not been successful in the past [25,26]. Clinicians need to communicate with and educate cancer survivors about the high risk of secondary cancer and guide them on smoking cessation. Collaboration between oncologists and primary physicians is essential to provide high-quality comprehensive smoking cessation interventions, as smoking cessation services are mainly provided by primary physicians [27].

The findings of this study indicate that young cancer survivors are at a higher risk of developing SPCs, implying that factors beyond “bad luck,” such as hereditary predispositions, may play a significant role in the development of SPCs in young adults. There were significant associations between specific types of SPCs and FPCs that may be linked to hereditary cancer syndrome. In men, young kidney cancer survivors were at a high risk of developing SPCs, with pancreas and brain cancers being the most common types. VHL syndrome, caused by germline mutations of the VHL gene, is characterized by tumors of the central nervous system, kidney, retina, and pancreas. Only a small number of studies have evaluated VHL syndrome in Korea; the incidence of VHL disease in patients with renal cell carcinoma was reported at 0.7%, and it was male-dominant (70%) with a median age of onset of 33 years (14-59 years) [28,29]. In women, the SIRs were significantly high among breast cancer survivors for ovarian cancer and colorectal cancer survivors for uterus cancers, possibly due to hereditary breast and ovarian cancer and Lynch syndrome harboring a germline mutation in BRCA1/2 and mismatch repair–related genes, respectively. Identifying hereditary cancer syndrome is crucial for cancer prevention and treatment, as both the individual and one-half of first-degree family members are at high risk of cancer [30]. Hereditary cancer syndromes may represent a broader clinical spectrum than previously understood [31,32]. Further genetic studies targeting young cancer survivors with a high risk of SPCs are worth conducting.

The most common types of FPCs were mainly observed as SPCs in cancer survivors, with lung cancer and prostate cancer in men being the most dominant SPCs. This suggests that cancer screening programs for cancer survivors need to focus more on these cancer types. For example, low-dose computed tomography over chest x-ray for lung cancer screening [33] in both sexes and prostate ultrasonography in addition to prostate-specific antigen tests for male cancer survivors may be worthy of consideration. Furthermore, the risk of SPCs varied by the types of FPC and sex, with young cancer survivors being more vulnerable to developing SPCs. These findings highlight the need for tailored cancer screening programs for this population. Based on these results, there is a need to develop cancer type-, age-, and sex-specific national screening programs for cancer survivors. Further studies are necessary to assess the benefits of new evidence-based tailored screening programs for this population.

An interesting association was observed in male breast cancer survivors, as they had a higher risk of developing prostate cancer as an SPC. The positive association has been reported in male patients [34,35], and 2 hypotheses have been proposed to explain it [36]. The first hypothesis is that 4% to 40% of male breast cancer patients harbor a BRCA2 germline mutation [37,38]. Since BRCA1/2 mutations are associated with both breast and prostate cancers, the increased risk of prostate cancer in male breast cancer survivors could be due to their germline susceptibility [39,40]. The second hypothesis is that most male breast cancers are estrogen receptor–positive, which makes them good candidates for hormone treatment with aromatase inhibitors [35,41]. The suppression of estrogen could cause an imbalance of estrogen and testosterone, which hypothetically increases the risk of prostate cancer [42]. This association highlights the need for germline testing and careful prostate cancer screening in male breast cancer survivors.

**Limitations**

Although this study provides novel findings on the risk of SPCs in Korean cancer survivors, several limitations need to be acknowledged. Factors such as latency, family cancer history, clinicopathologic and sociodemographic factors, and treatments for FPCs, including chemoradiation therapy, could also influence the risk of SPCs [43]. The results showed strong positive associations among non-Hodgkin lymphoma, Hodgkin lymphoma, and leukemia, which may be due to the effect of chemotherapy [44]. However, due to the unavailability of relevant data, this study did not examine their effects. Additionally, the lack of information on other known risk factors such as smoking, alcohol consumption, diet, physical activity level, excess body weight, environmental carcinogens, and cancer-associated infection status such as human papilloma virus infection, is a limitation. The incidence of SPCs at the same organ as the FPC could be underestimated as they were excluded to minimize bias due to misidentification of FPC recurrence. Furthermore, the incidence of SPCs could be underestimated in cancer survivors with FPC types associated with high mortality, such as pancreatic and lung cancers, as deceased patients within 5 years after the FPC diagnosis were not included. Heightened medical surveillance following the FPC diagnosis may have increased the chance of detecting SPCs, particularly in cases where the sites of the FPC and SPC were anatomically close (eg, thyroid and head and neck or breast and lung cancers). However, to minimize this bias, SPCs occurring within 5 years of the FPC diagnosis were excluded, as ICBP covers the medical cost for only 5 years. Some of the significant associations could be biased by multiple testing, and the statistical power was also limited, especially for the subgroup analysis in young patients with FPCs due to the small sample size. Finally, some minor malignancies that could have been related to genetic susceptibility, such as pheochromocytoma and paraganglioma syndrome [45], were excluded from the
analysis as they were classified as other cancers (other C00-96). The American Cancer Society defines cancer survivors as “anyone who has ever been diagnosed with cancer no matter where they are in the course of their disease” [46], whereas we focused on 5-year cancer survivors in this study.

**Comparison With Prior Work**

The results of this study differ significantly from those of a previous study that used SEER data [4]. In that study, the risks of SPCs in survivors of adult-onset cancers were higher in several types of primary cancer compared with the general population in the United States. Several factors could explain this difference. First, the same group conducted another study that focused on adolescent and young adult (AYA; aged 15-39 years) cancer survivors using SEER data [47], and the risk of SPCs was generally higher in AYA than in adult cancer survivors (Table S6 in Multimedia Appendix 1). Second, although the SEER study covered 9% to 13% of the US population [4], HIRA covers nearly the entire Korean population. Third, only 8.1% of the population in the SEER study Asian or Pacific Islander descent, while nearly all Koreans are Asian. Fourth, the incidences of cancer and frequent cancer types differ between Korea and the United States (Table S7 in Multimedia Appendix 1) [48]. Fifth, the mean follow-up durations were 7.3 years in the SEER study and 3.8 years in our study, and the shorter follow-up duration in this study might have led to an underestimation of the risk of developing SPCs. Last, cancer screening rates vary by cancer type and nation [49,50].

**Conclusion**

The varying risk of SPCs by age, sex, and the type of FPC in cancer survivors suggests the necessity for ongoing efforts to develop tailored prevention and screening programs for cancer survivors. Lifestyle modifications, including smoking cessation, are essential to reduce the risk of SPCs in cancer survivors. In addition, genetic testing, along with proactive cancer screening and prevention strategies, should be implemented for young cancer survivors because of their elevated risk of developing SPCs.

**Acknowledgments**

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00302584) and the Soonchunhyang University Research Fund.

**Data Availability**

Data sharing is not applicable to this article, as the Health Insurance Review and Assessment (HIRA) data are managed by the Korean government. Access to the data is restricted to authorized individuals after review of the study protocol and is only permissible from designated computers.

**Authors' Contributions**

YYC, ML, IJ, and JHC designed the study. YYC, ML, EHK, and JEL collected the data. YYC, ML, EHK, and IJ analyzed the data. YYC and ML wrote the manuscript. IJ (IJUNG@yuhs.ac) and JHC (jcheong@yuhs.ac) contributed equally as co-corresponding authors. All authors approved the final version of the manuscript.

**Conflicts of Interest**

None declared.

**References**


9. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science 2015 Jan 02;347(6217):78-81 [FREE Full text] [doi: 10.1126/science.aaf9011] [Medline: 25953640]


22. Ordóñez-Mena JM, Schöttker B, Mons U, Jenab M, Freising H, Bueno-de-Mesquita B, Consortium on HealthAgeing: Network of Cohorts in Europe (SHARE). Quantification of the smoking-associated cancer risk with...


https://publichealth.jmir.org/2024/1/e48380


Abbreviations

- **AYA**: adolescent and young adult
- **FPC**: first primary cancer
- **HIRA**: Health Insurance Review and Assessment
- **ICBP**: Korean Individual Co-payment Beneficiaries Program
- **ICD**: International Classification of Diseases
- **SEER**: Surveillance, Epidemiology, and End Results Program
- **SIR**: standardized incidence ratio
- **SPC**: subsequent primary cancer
- **VHL**: Von-Hippel-Lindau

©Yoon Young Choi, Myeongjee Lee, Eun Hwa Kim, Jae Eun Lee, Inkyung Jung, Jae-Ho Cheong. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 08.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Metabolic Score for Insulin Resistance and New-Onset Type 2 Diabetes in a Middle-Aged and Older Adult Population: Nationwide Prospective Cohort Study and Implications for Primary Care

Hui Cheng¹*, MD; Zhihui Jia¹*, MD; Yu Ting Li²,³, MPH; Xiao Yu¹, MM; Jia Ji Wang⁴, MPH; Yao Jie Xie⁵, PhD; Jose Hernandez⁶,⁷, DPhil; Harry H X Wang¹,⁸,⁹, PhD

¹School of Public Health, Sun Yat-Sen University, Guangzhou, China
²State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, China
³Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, China
⁴School of Public Health, Guangzhou Medical University, Guangzhou, China
⁵School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, China (Hong Kong)
⁶Faculty of Medicine and Health, EDU, Digital Education Holdings Ltd, Kalkara, Malta
⁷Green Templeton College, University of Oxford, Oxford, United Kingdom
⁸Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, New Territories, China (Hong Kong)
⁹Usher Institute, Deanery of Molecular, Genetic & Population Health Sciences, The University of Edinburgh, Edinburgh, United Kingdom

*these authors contributed equally

Corresponding Author:
Harry H X Wang, PhD
School of Public Health
Sun Yat-Sen University
No 74 Zhongshan Road 2
Guangzhou, 510080
China
Phone: 86 20 87330672
Email: wanghx27@mail.sysu.edu.cn

Abstract

Background: The metabolic score for insulin resistance (METS-IR) has emerged as a noninsulin-based index for the approximation of insulin resistance (IR), yet longitudinal evidence supporting the utility of METS-IR in the primary prevention of type 2 diabetes mellitus (T2DM) remains limited.

Objective: We aimed to investigate the longitudinal association between METS-IR, which combines fasting plasma glucose (FPG), lipid profiles, and anthropometrics that can be routinely obtained in resource-limited primary care settings, and the incidence of new-onset T2DM.

Methods: We conducted a closed-cohort analysis of a nationwide, prospective cohort of 7583 Chinese middle-aged and older adults who were free of T2DM at baseline, sampled from 28 out of 31 provinces in China. We examined the characteristics of participants stratified by elevated blood pressure (BP) at baseline and new-onset T2DM at follow-up. We performed Cox proportional hazard regression analysis to explore associations of baseline METS-IR with incident T2DM in participants overall and in participants stratified by baseline BP. We also applied net reclassification improvement and integrated discrimination improvement to examine the incremental value of METS-IR.

Results: During a mean follow-up period of 6.3 years, T2DM occurred in 527 participants, among which two-thirds (332/527, 62.9%; 95% CI 58.7%-67.1%) had baseline FPG<110 mg/dL. A SD unit increase in baseline METS-IR was associated with the first incidence of T2DM (adjusted hazard ratio [aHR] 1.33, 95% CI 1.22-1.45; P<.001) in all participants. We obtained similar results in participants with normal baseline BP (aHR 1.41, 95% CI 1.22-1.62; P<.001) and elevated baseline BP (aHR 1.29, 95% CI 1.16-1.44; P<.001). The predictive capability for incident T2DM was improved by adding METS-IR to FPG. In study participants with new-onset T2DM whose baseline FPG was <126 mg/dL and <110 mg/dL, 62.9% (332/527; 95% CI 60%-65.9%) and 58.1% (193/332; 95% CI 54.3%-61.9%) of participants had baseline METS-IR above the cutoff values, respectively.
Conclusions: METS-IR was significantly associated with new-onset T2DM, regardless of baseline BP level. Regular monitoring of METS-IR on top of routine blood glucose in clinical practice may add to the ability to enhance the early identification of primary care populations at risk for T2DM.

doi:10.2196/49617

KEYWORDS
metabolic score for insulin resistance; type 2 diabetes mellitus; blood pressure; longitudinal study; primary care

Introduction

Type 2 diabetes mellitus (T2DM) exerts a profound impact on the health and well-being of individuals, families, and communities worldwide because of its high prevalence, concomitant increase in risks of complications and treatment costs, and reduced quality of life [1-4]. Data from the European Association for the Study of Diabetes, the American Diabetes Association, and the International Diabetes Federation consistently indicate a rising epidemic of T2DM regionally and globally [5,6]. In low- and middle-income countries, the rapid progress of urbanization, aging, lifestyle transformations, and a lack of sustainable health education have further posed significant challenges to the prevention of T2DM [6-8]. Moreover, patients at older ages with poor glycemic control may suffer from progressive pathological as well as functional decline [9]. Evidence suggests that the progression of T2DM and its complications can largely be delayed through population-based early preventive public health programs [10,11].

Insulin resistance (IR) is a common feature of prediabetes and prehypertension and also a precursor to the development of both conditions [12]. Recommendations stress that the initiation of intervention during the stage of IR, instead of after a T2DM diagnosis, is more effective in reducing treatment burden and overall health care costs [13]. The hyperinsulinemic-euglycemic clamp (HEC) has been widely considered the “gold standard” experimental method for direct determination of IR, although less popular in daily practice due to its invasiveness, expensiveness, and complex procedures [14]. The homeostatic model assessment for IR and the quantitative insulin sensitivity check index, which are both fasting insulin-based indexes [15], are also less commonly performed as part of regular checkups in low- and middle-income countries’ primary care settings.

The metabolic score for insulin resistance (METS-IR), a novel index that considers fasting plasma glucose (FPG), lipid profiles, and obesity index that can be easily obtained during routine examination, has been proposed as a simple and high-accuracy measure to assess IR in the western population [14]. Evaluation of METS-IR may reduce the cost associated with the immunoassay of insulin and the heterogeneity between different analytical methods [14,16]. The hypothesis that METS-IR is predictive of new-onset T2DM over time needs further testing in the Chinese middle-aged and older adult population, which accounts for 42.2% of the total population [17]. From a multimorbidity perspective, whether the association between METS-IR and first-incident T2DM is persistent among participants whose blood pressure (BP) falls within the normal range also remains uncertain.

We aimed to investigate the association between METS-IR at baseline and the incidence of new-onset T2DM at follow-up in a middle-aged and older adult population. We also sought to examine whether the predictive capability of METS-IR on top of routine blood glucose for incident T2DM differed between participants with and those without elevated baseline BP.

Methods

Study Design

We conducted a cohort analysis of nationally representative survey data retrieved from the China Health and Retirement Longitudinal Study between 2011 and 2018, run by Peking University in partnership with the National Natural Science Foundation of China, the National Institute on Aging, and the World Bank. The data collected during the study period are made available in the public domain. The China Health and Retirement Longitudinal Study was originally designed after the US Health and Retirement Study and other internationally developed aging-related surveys to collect high-quality data on various aspects of health and social care for middle-aged and older adults in China [18-21].

Setting and Data Source

The nationwide baseline survey commenced in 2011 (wave 1). Follow-up surveys were repeatedly conducted in 2013 (wave 2), 2015 (wave 3), and 2018 (wave 4). The details of the sampling frame and household interview procedures were described elsewhere [18,19]. In brief, multistage, region-stratified, probability proportional to size sampling was adopted for sample selection from 150 county-level districts in 28 out of a total of 31 provinces across China. The baseline participants were drawn from over 10 thousand households in 450 neighborhood- and village-level units. Information on the individual’s sociodemographics, lifestyles, and health status was collected through face-to-face computer-assisted interviews in each wave. Data on anthropometric and clinical parameters were obtained by physical examinations, including blood drawing, for subsequent laboratory tests in waves 1 and 3 [19-21].

Participants

A total of 17,708 participants were enrolled at baseline, of whom 17,314 were aged 45 years and older. We excluded individuals who had any of the following: (1) a clinical diagnosis of T2DM, FPG≥126 mg/dL (7 mmol/L), or glycated hemoglobin A1c (HbA1c) ≥6.5% (n=2414); (2) missing information on the diagnosis of T2DM (n=130); (3) unknown BP levels (n=2490); (4) incomplete data on BMI, FPG, triglycerides, and high-density lipoprotein cholesterol (HDLC) to calculate
baseline METS-IR (n=4454); and (5) nonattendance at any of the follow-ups (n=243). This yielded a total of 7583 participants who met the eligibility criteria and were included in the closed-cohort analysis (Figure S1 in Multimedia Appendix 1).

**Study Variables and Measurements**

Information on age, sex, place of residence, education level, household income, and living relationships was collected by centrally trained interviewers with an internationally comparable, validated survey instrument [19]. Self-reported lifestyles included current cigarette smoking and engagement in alcohol drinking at least once per month regularly. Anthropometric parameters were measured with participants in light clothing and without shoes, using a portable stadiometer (Seca 213 stadiometer [Seca Trading]) and a calibrated digital scale (Omron HN-286 scale [Krell Precision]) [19]. BMI was calculated as weight in kilograms divided by the square of height in meters (kg/m²). The BP of the arm with a higher value was measured in a seated position using a routinely validated automatic BP monitor (Omron HEM-7112/7200 monitor [Omron]), and the average of 3 BP readings taken at 45-second intervals was recorded [19]. A venous blood sample at fasting was collected for blood-based bioassays according to standard operating procedures [20,21]. The Hexokinase method was used for FPG measurement (mg/dL), while lipid panel profiles including total cholesterol (mg/dL) and triglycerides (mg/dL) were measured using the Oxidase method, and low-density lipoprotein cholesterol (mg/dL) and HDL-C (mg/dL) were determined using the direct method [21]. Glucose and lipid parameters, together with BMI, were taken into account in the assessment of METS-IR, which was calculated as: Ln((2xFPG)+triglyceride)×BMI/Ln(HDL-C) [14]. Elevated BP at baseline was defined as systolic BP≥120 mm Hg or diastolic BP≥80 mm Hg (or both) or the presence of physician-diagnosed hypertension [22]. This included both hypertension and high normal BP. The primary outcome of this study was the first incidence of physician-diagnosed T2DM during follow-up, which was double-verified using the information documented in the previous follow-up wave. Participants enrolled at baseline were followed until they had newly diagnosed T2DM or the recorded attendance at the most recent wave of follow-up, whichever came first.

**Statistical Analysis**

Data are presented as n (%) values for categorical variables and as mean (SD) or median (IQR; 25th to 75th percentiles) values where appropriate for continuous variables. We examined the characteristics of participants stratified by the presence of elevated BP at baseline and new-onset T2DM at follow-up. The 2-tailed 2-sample t test, the nonparametric Wilcoxon rank sum test, or the chi-square test, where appropriate, was used for between-group comparisons in participants with and those without new-onset T2DM. Participants were further divided into quartiles of METS-IR, where the lowest quartile was used as the reference group. The cumulative hazard of T2DM was determined by the Kaplan-Meier plot, and the 2-sided logrank test was used for the overall comparison of curves across METS-IR quartiles. Cox proportional hazard regression models were constructed to estimate the risk of new-onset T2DM in participants overall and in participants stratified by baseline BP after adjusting for known important covariates, including demographics, socioeconomic status, lifestyles, and anthropometric measurements. The proportional hazards assumption for model fit was tested using the scaled Schoenfeld residuals. The adjusted hazard ratios (aHR) with a 95% CI were estimated for a unit increase in METS-IR per SD and for each METS-IR quartile. We further modeled the data as restricted cubic splines with 4 knots, located at the 5th, 35th, 65th, and 95th percentiles following the Akaike information criterion [23], of METS-IR to assess the shape of the association between METS-IR and the risk of T2DM. All models were adjusted for confounders included in the Cox models.

We also assessed the predictive capability of baseline METS-IR for new-onset T2DM. We applied net reclassification improvement and integrated discrimination improvement [24] to examine whether adding METS-IR to FPG may have incremental value in improving the predictive accuracy for T2DM. Furthermore, the optimal cutoff points with the highest Youden index were derived from the receiver operating characteristic (ROC) analysis. We calculated the proportion of people with new-onset T2DM who had baseline METS-IR above the corresponding cutoff value in all participants and in the subgroup without baseline impaired fasting glucose (ie, FPG<110 mg/dL [25]).

A series of sensitivity analyses were performed to ensure the reliability of the results. We repeated the aforementioned Cox models fitted in the main analysis under 3 different scenarios. First, participants who had an incident of T2DM at the first follow-up visit were excluded to account for possible reverse causality bias. Second, participants with the presence of general obesity (ie, BMI ≥28 kg/m² according to the Working Group on Obesity in China [26]) at baseline were excluded given the strong association between obesity and incident diabetes. Third, we used laboratory test results of FPG and HbA1c, where available, to supplement the information on the occurrence of T2DM (ie, FPG≥126 mg/dL [7 mmol/L] or HbA1c≥6.5%) to ascertain whether associations between METS-IR and T2DM may vary from the main analysis. We also treated age, socioeconomic status, lifestyles, and anthropometric measurements as time-dependent variables in the Cox models built in the main analysis to take into account the potential time-varying effects of covariates. In addition, the demographic characteristics of participants included in the analysis were compared with those of participants excluded due to missing values to understand the potential selection bias and its impact on the generalizability of the findings. Analyses were conducted using SAS (version 9.4; SAS Institute Inc) and R (version 4.0.2; R Core Team). A 2-tailed P value of less than .05 was considered statistically significant.

**Ethical Considerations**

All study participants provided written consent at enrolment. Data are publicly archived by Peking University, where ethics approval was obtained (IRB00001052-11014 and IRB00001052-11015). The ethics of the present analysis were sought from the Biomedical Research Ethics Review Committee at Sun Yat-Sen University (SPH2019123). Data anonymization...
was performed by removing all participant identifiers before data analysis.

Results

Study Population and Baseline Characteristics

Data were analyzed among a total of 7583 participants free of T2DM at baseline (Figure S1 in Multimedia Appendix 1). Participants ranged in age between 45 and 96 years, with a mean age of 59.1 (SD 9.3) years, sampled from 28 out of 31 provinces. Around one-third (2502/7583) of participants had a BP<120/80 mm Hg at baseline. T2DM occurred in 527 participants (ie, 214 male participants and 313 female participants) over a mean of 6.3 years of follow-up, among which nearly two-thirds (332/527, 62.9%; 95% CI 58.7%-67.1%) had baseline FPG<110 mg/dL (6.1 mmol/L). The baseline METS-IR of participants who had T2DM at follow-up was significantly higher than that of nonincident T2DM counterparts (36.1, SD 8.4 vs 32.6, SD 6.6; P<.001 among those with normal baseline BP and 38.9, SD 8 vs 35.3, SD 7.8; P<.001 among those with elevated baseline BP; Table 1). In participants with normal baseline BP, those who developed T2DM had significantly higher levels of BMI, waist circumference, FPG, and triglyceride but lower HDL-C at baseline when compared to those free of new-onset T2DM. A similar pattern was observed in participants with elevated baseline BP (Table S1 in Multimedia Appendix 1). Relative to participants who were excluded due to missing information (METS-IR, BP, FPG, and lost to follow-up), this study sample had a greater proportion of female participants (4051/7583, 53.4% vs 3572/7317, 48.8%), rural residents (5029/7583, 66.3% vs 3948/7317, 54%), those with lower education level (5325/7583, 70.2% vs 4558/7317, 62.5%), and current smokers (2317/7583, 30.7% vs 1908/7317, 28.7%), despite similar mean age (59.1 vs 59.2 years; P=.84) and alcohol drinking profile (2522/7583, 33.3% vs 2478/7317, 34.4%; P=.15; Table S2 in Multimedia Appendix 1).

The 2-sample t test or the chi-square test, where appropriate, was used for between-group comparisons among participants with and those without new-onset T2DM (ie, T2DM at follow-up vs non-T2DM at follow-up). Normal BP was defined as systolic BP<120 mm Hg and diastolic BP<80 mm Hg. Elevated BP was defined as systolic BP≥120 mm Hg or diastolic BP≥80 mm Hg (or both) or the presence of physician-diagnosed hypertension according to the 2018 Chinese Guidelines for the Management of Hypertension.
Table 1. Characteristics of participants by blood pressure (BP) at baseline and new-onset type 2 diabetes mellitus (T2DM) at follow-up.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Normal BP at baseline</th>
<th>Elevated BP at baseline</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T2DM at follow-up</td>
<td>Non-T2DM at follow-up</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n=116)</td>
<td>(n=2386)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T2DM at follow-up</td>
<td>Non-T2DM at follow-up</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n=411)</td>
<td>(n=4670)</td>
<td></td>
</tr>
<tr>
<td>Age groups (years), n (%)</td>
<td>.10</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>40 (34.5)</td>
<td>1052 (44.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>135 (32.8)</td>
<td>1390 (29.8)</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>50 (43.1)</td>
<td>928 (38.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>171 (41.6)</td>
<td>1793 (38.4)</td>
<td></td>
</tr>
<tr>
<td>≥65</td>
<td>26 (22.4)</td>
<td>406 (17)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>105 (25.6)</td>
<td>1487 (31.8)</td>
<td></td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td>.35</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>47 (40.5)</td>
<td>1073 (44.97)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>167 (40.6)</td>
<td>2245 (48.1)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>69 (59.5)</td>
<td>1313 (55.03)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>244 (59.4)</td>
<td>2425 (51.9)</td>
<td></td>
</tr>
<tr>
<td>Place of residence, n (%)</td>
<td>.21</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>30 (25.9)</td>
<td>749 (31.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>139 (33.8)</td>
<td>1636 (35)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>86 (74.1)</td>
<td>1637 (68.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>272 (66.2)</td>
<td>3034 (65)</td>
<td></td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td>.07</td>
<td>.74</td>
<td></td>
</tr>
<tr>
<td>Elementary school or below</td>
<td>88 (75.9)</td>
<td>1612 (67.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>290 (70.7)</td>
<td>3335 (71.4)</td>
<td></td>
</tr>
<tr>
<td>Middle school or above</td>
<td>28 (24.1)</td>
<td>774 (32.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>120 (29.3)</td>
<td>1335 (28.6)</td>
<td></td>
</tr>
<tr>
<td>Annual household income, n (%)</td>
<td>.36</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>Quartile 1a (most deprived)</td>
<td>23 (20.2)</td>
<td>502 (21.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>110 (27)</td>
<td>1250 (26.9)</td>
<td></td>
</tr>
<tr>
<td>Quartile 2b</td>
<td>36 (31.6)</td>
<td>588 (24.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>103 (25.25)</td>
<td>1161 (25)</td>
<td></td>
</tr>
<tr>
<td>Quartile 3c</td>
<td>30 (26.3)</td>
<td>630 (26.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>105 (25.7)</td>
<td>1112 (24)</td>
<td></td>
</tr>
<tr>
<td>Quartile 4d (most affluent)</td>
<td>25 (21.9)</td>
<td>648 (27.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>90 (22.1)</td>
<td>1120 (24.1)</td>
<td></td>
</tr>
<tr>
<td>Living relationships, n (%)</td>
<td>.67</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td>Living with a partner</td>
<td>103 (88.8)</td>
<td>2087 (87.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>341 (83)</td>
<td>3815 (81.7)</td>
<td></td>
</tr>
<tr>
<td>Living alone</td>
<td>13 (11.2)</td>
<td>299 (12.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>70 (17)</td>
<td>855 (18.3)</td>
<td></td>
</tr>
<tr>
<td>Cigarette smoking, n (%)</td>
<td>.93</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Current smoker</td>
<td>35 (30.2)</td>
<td>726 (30.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>92 (22.5)</td>
<td>1464 (31.5)</td>
<td></td>
</tr>
<tr>
<td>Nonsmoker</td>
<td>81 (69.8)</td>
<td>1650 (69.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>317 (77.5)</td>
<td>3190 (68.5)</td>
<td></td>
</tr>
<tr>
<td>Alcohol drinking, n (%)</td>
<td>.60</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Regular drinker</td>
<td>36 (31)</td>
<td>796 (33.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>107 (26)</td>
<td>1583 (33.9)</td>
<td></td>
</tr>
<tr>
<td>Nondrinker</td>
<td>80 (69)</td>
<td>1588 (66.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>304 (74)</td>
<td>3084 (66.1)</td>
<td></td>
</tr>
<tr>
<td>METS-IRe index, mean (SD)</td>
<td>36.1 (8.4)</td>
<td>32.6 (6.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>38.9 (8)</td>
<td>35.3 (7.8)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aQuartile 1: ≤CNY 5200 (≤US $719.79).
bQuartile 2: ≥CNY 5201 to CNY 15,300 (≥US $719.93-$2117.85).
cQuartile 3: ≥CNY 15,301 to CNY 34,176 (≥US $2117.99-$4730.70).
dQuartile 4: ≥CNY 34,177 (≥US $4730.84).
eMETS-IR: metabolic score for insulin resistance.

The First Incidence of Physician-Diagnosed T2DM at Follow-Up

The incidence of new-onset T2DM was reached at a rate of 11 per 1000 person-years of follow-up in overall participants (ie, 7.2 and 13 per 1000 person-years for participants with normal baseline BP and elevated baseline BP, respectively; Table 2). The highest cumulative hazard of T2DM was observed at 7 years in participants with the highest quartile of baseline METS-IR (13.3%) when compared to that in those with the lowest quartile (4.1%), with a difference of 9.2 percentage points (95% CI 7.4-11; Figure 1).
Table 2. Incidence rate of physician-diagnosed type 2 diabetes mellitus (T2DM) at follow-up and associations between baseline metabolic score for insulin resistance (METS-IR) and new-onset T2DM.

<table>
<thead>
<tr>
<th>Variables</th>
<th>T2DM cases, n</th>
<th>Incidence rate&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Crude model</th>
<th>Adjusted model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HR&lt;sup&gt;b&lt;/sup&gt; (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>METS-IR, per SD unit increase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participants at baseline</td>
<td>527</td>
<td>11</td>
<td>1.44 (1.35-1.53)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Participants with normal blood pressure at baseline</td>
<td>116</td>
<td>7.2</td>
<td>1.38 (1.23-1.54)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Participants with elevated blood pressure at baseline</td>
<td>411</td>
<td>13</td>
<td>1.41 (1.31-1.52)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>METS-IR, quartiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participants at baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quartile 1 (≤29.18)</td>
<td>70</td>
<td>5.9</td>
<td>1 (Reference)</td>
<td>N/A&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Quartile 2 (29.19-33.47)</td>
<td>87</td>
<td>7.2</td>
<td>1.22 (0.89-1.67)</td>
<td>.22</td>
</tr>
<tr>
<td>Quartile 3 (33.48-38.97)</td>
<td>135</td>
<td>11.3</td>
<td>1.93 (1.44-2.57)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Quartile 4 (≥38.98)</td>
<td>235</td>
<td>19.9</td>
<td>3.39 (2.6-4.43)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>P value for trend</td>
<td>N/A</td>
<td>N/A</td>
<td>1.08 (1.07-1.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Participants with normal blood pressure at baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quartile 1 (≤28.22)</td>
<td>22</td>
<td>5.5</td>
<td>1 (Reference)</td>
<td>N/A</td>
</tr>
<tr>
<td>Quartile 2 (28.23-31.7)</td>
<td>13</td>
<td>3.2</td>
<td>0.59 (0.3-1.16)</td>
<td>.13</td>
</tr>
<tr>
<td>Quartile 3 (31.71-36.18)</td>
<td>33</td>
<td>8.1</td>
<td>1.48 (0.86-2.53)</td>
<td>.16</td>
</tr>
<tr>
<td>Quartile 4 (≥36.19)</td>
<td>48</td>
<td>12.1</td>
<td>2.22 (1.34-3.68)</td>
<td>.002</td>
</tr>
<tr>
<td>P for trend</td>
<td>N/A</td>
<td>N/A</td>
<td>1.08 (1.04-1.12)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Participants with elevated blood pressure at baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quartile 1 (≤29.87)</td>
<td>55</td>
<td>7.0</td>
<td>1 (Reference)</td>
<td>N/A</td>
</tr>
<tr>
<td>Quartile 2 (29.88-34.59)</td>
<td>70</td>
<td>8.8</td>
<td>1.25 (0.88-1.77)</td>
<td>.22</td>
</tr>
<tr>
<td>Quartile 3 (34.6-40.26)</td>
<td>110</td>
<td>13.9</td>
<td>1.98 (1.43-2.73)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Quartile 4 (≥40.27)</td>
<td>176</td>
<td>22.2</td>
<td>3.17 (2.34-4.28)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>P for trend</td>
<td>N/A</td>
<td>N/A</td>
<td>1.07 (1.06-1.09)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>per 1000 person-years.

<sup>b</sup>HR: hazard ratio.

<sup>c</sup>aHR: adjusted hazard ratio.

<sup>d</sup>Not applicable.
Normal BP at baseline was defined as systolic BP<120 mm Hg and diastolic BP<80 mm Hg. Elevated BP at baseline was defined as systolic BP≥120 mm Hg or diastolic BP≥80 mm Hg (or both) or the presence of physician-diagnosed hypertension according to the 2018 Chinese Guidelines for the Management of Hypertension. Cox proportional hazard model adjusted for age, sex, place of residence, education level, annual household income, living relationships, cigarette smoking, alcohol drinking, BP, and waist circumference. Tests for trend are based on variables containing the median value for each quartile.

**Associations Between Baseline METS-IR and First Incident T2DM**

An SD unit increase in baseline METS-IR was associated with the first incidence of T2DM (aHR 1.32, 95% CI 1.22-1.45; \(P<.001\)) in all participants. The adjusted risk of new-onset T2DM in participants with the highest quartile of baseline METS-IR was 2.72-fold higher than that in those with the lowest quartile of METS-IR. Similar trends were obtained from both crude and adjusted models, irrespective of baseline BP levels (Table 2). The 4-knot curve demonstrated a nonlinear association between baseline METS-IR and new-onset T2DM in participants overall and in participants with elevated baseline BP (Figure 2). The same models were repeated in the sensitivity analyses with consistent results (Tables S3-S6 in Multimedia Appendix 1).
Figure 2. Restricted cubic spline estimates of the relationship between the metabolic score for insulin resistance (METS-IR) and new-onset type 2 diabetes mellitus (T2DM). BP: blood pressure; HR: hazard ratio.

Restricted cubic splines with 4 knots, located at the 5th, 35th, 65th, and 95th percentiles of METS-IR for (A) all participants, (B) participants with normal BP, and (C) participants with elevated BP. The solid black line represents the fitted curve, and the gray bands represent the 95% CI bands. The model was adjusted for age, sex, place of residence, education level, annual household income, marital relationship, cigarette smoking, alcohol drinking, BP, and waist circumference. Normal BP was defined as systolic BP<120 mm Hg and diastolic BP<80 mm Hg. Elevated BP was defined as systolic BP≥120 mm Hg or diastolic BP≥80 mm Hg (or both) or the presence of physician-diagnosed hypertension according to the 2018 Chinese Guidelines for the Management of Hypertension.

The Predictive Capability of Baseline METS-IR for New-Onset T2DM

The ROC curves based on the predicted probabilities for incident T2DM showed that the area under the curve ranged from 0.633 to 0.708 across all models. Both net reclassification improvement and integrated discrimination improvement for predicting new-onset T2DM increased by adding METS-IR to FPG, irrespective of baseline elevated BP (Table S7 in Multimedia Appendix 1). In all participants with new-onset T2DM whose baseline BP levels were <126 mg/dL (7 mmol/L) and <110 mg/dL (6.1 mmol/L), 62.9% (332/527; 95% CI 60%-65.9%) and 58.1% (193/332; 95% CI 54.3%-61.9%) had baseline METS-IR above cutoff values (ie, ≥35.3 for all participants, ≥31.7 for participants with normal BP, and ≥35.5 for participants with elevated BP), respectively. The proportion of participants with baseline METS-IR above the corresponding cutoff values was similar between participants with and those without elevated baseline BP (Figure 3).

Figure 3. Proportion of participants with new-onset type 2 diabetes mellitus (T2DM) at follow-up by blood pressure (BP) at baseline. FPG: fasting plasma glucose.
Discussion

Main Findings

We provided longitudinal evidence supporting the potential utility of METS-IR in the prediction of T2DM in a large, nationally representative sample of the Chinese middle-aged and older adult population. We demonstrated significant associations between increased baseline METS-IR and a higher incidence of T2DM, regardless of the presence of baseline elevated BP. The predictive capability for incident T2DM was improved by adding METS-IR to FPG. In participants with incident T2DM whose baseline FPG fell within the normal range, around two-thirds had baseline METS-IR above the cutoff values, implying that the use of METS-IR on top of routine blood glucose in clinical practice may add additional value to early identification of populations at risk for T2DM.

Relationship With Other Studies

The role that METS-IR plays as a complement to previously validated risk prediction models is an emerging area of research interest. As a novel noninsulin-based index for estimating insulin sensitivity, METS-IR combines fasting laboratory values and anthropometric measurements that are routinely available in primary care instead of relying on insulin-based measurements that are laborious, expensive, and of high variability due to different immunoassay techniques [16]. Recent evidence indicates a satisfactory validity of METS-IR against the adjusted whole-body glucose disposal rate (ie, M value) derived from the “gold standard” HEC and the reliability of METS-IR when compared to both fasting insulin-based indexes (eg, homeostatic model assessment for IR and quantitative insulin sensitivity check index) and noninsulin-based indexes (eg, triglyceride-glucose index, triglyceride-glucose-BMI index, and triglyceride/HDL-C ratio) for predicting T2DM in a Mexican outpatient cohort [14]. The relationship between METS-IR and the occurrence of T2DM observed in the Latin-American population was similarly documented in subsequent studies among nonobese adults recruited in the Japanese NAFld in the Gifu Area, Longitudinal Analysis cohort [27,28] and among rural adults in Henan, northern China [29]. Further to our previous cross-sectional investigation in an urbanized township in southern China [30], we extended the analysis by using a large cohort of middle-aged and older adults that are nationally representative, while taking into account the urban-rural heterogeneity in lifestyles, income, and education profiles, as well as health care disparities that are prevailing in developing countries [31]. We found a consistent relationship between increased baseline METS-IR and risks for new-onset T2DM, irrespective of BP levels, stressing that normotensive people should not be neglected in diabetes prevention.

The potential association between METS-IR and incident T2DM may have several mechanisms. METS-IR was found to be significantly correlated with visceral, intrahepatic, and intrapancreatic fat content that may contribute to pathophysiological alterations of glucose and lipid homeostasis [14,32-35]. This may translate into chronically elevated levels of glucose and fatty acids, thereby inducing toxic states in pancreatic islets and progressive worsening of beta-cell function due to glucotoxicity, lipotoxicity, and glucolipotoxicity [36]. In addition to plasma triglyceride and HDL-C that are related to insulin-mediated glucose disposal, the inclusion of BMI as a surrogate for visceral adiposity takes into account the causal relationship between adipose tissue inflammation and the development of IR and ultimately T2DM [37-39]. The nonlinear relationship we observed between baseline METS-IR and new-onset T2DM might be partly explained by the multifactorial interactions among adiposity, dysglycemia, and inflammation in metabolic pathways and vascular biologic processes [13,39-42]. Our findings echo previous literature on the nonlinear relationship between FPG and T2DM in the European population [43]. Given that splines were constructed from piecewise polynomials, the shape of the restricted cubic splines curve could be largely influenced by the location and number of knots. It is also worth noting that the cutoff values derived from the ROC analysis should not be interpreted as a rigid diagnostic threshold but rather as a reference level above which regular monitoring of METS-IR would probably yield additional value in targeted interventions despite normal blood glucose and therefore may be meaningful for risk assessment and risk communication in diabetes prevention. The use of METS-IR has also been extended to T2DM-related cardiovascular conditions in Mexican and South Korean populations [44,45], indicating the potential for wider applicability in community-based practice where a valid, easy-to-measure, and less resource-consuming prediction tool is preferable.

Strengths and Weaknesses of the Study

We examined the longitudinal relationships of baseline METS-IR and first incident T2DM in a large, nationally representative sample while evaluating potential nonlinear associations. The analyses were systematically performed using METS-IR as a continuous variable and in quartiles, with a consistent methodology adopted to deal with confounding and reverse causality. A fairly extensive range of sensitivity analyses made little difference in the estimated associations between baseline METS-IR and new-onset T2DM. A particular strength was that the size of the study allowed us to retain power while stratifying participants by the presence of baseline elevated BP and thus considering the challenge of addressing the most common modality of multimorbidity encountered in the real-world clinical setting. This study has some possible limitations that merit consideration. First, the Cox models were built on a specific cohort of population samples who were middle-aged and older adults, and thus the generalizability of our findings to a wider population, or patients in certain health care settings, should be interpreted with caution. Selection bias may occur as nearly half of participants from the source cohort were not included in the present analysis given the eligibility criteria, although comparable in population mean age and drinking profile. Second, the HEC was not available for direct assessment of insulin sensitivity due to time and budget constraints, which may inevitably affect the accuracy of the assessment. However, complex and expensive procedures for
determining insulin levels may limit implementation feasibility in large-scale studies in low-resource settings. Third, undiagnosed diabetes at baseline might have led to an underestimation of hazard ratios for new-onset T2DM. Fourth, we were not able to take account of the unmeasured confounders such as dietary regimes [46], body constitutions [47], environmental exposures, and health care usage that may exert cohort effects. We also acknowledge that time-varying confounding is impossible to fully overcome given that past exposures were not adequately captured. Fifth, we cannot rule out the possibility of random measurement errors in a large-scale multisite study, albeit with a standardized procedure for data collection. Last but not least, the blood drawing was voluntary and not performed in waves 2 and 4, and thus we were unable to cross-check the incidence of physician-diagnosed T2DM based on FPG and HbA1c. Nevertheless, the supplement use of biomedical records where available at follow-up in the sensitivity analyses yielded consistent results, suggesting the robustness of our findings.

**Implications for Research and Clinical Practice**

This study carries a clinically meaningful message for primary care physicians with a potentially greater opportunity for recognition of middle-aged and older adults early in the course of their diabetes. Physicians working in collaboration with other health care professionals can help those at-risk individuals enhance their capacity to optimize blood glucose, BP, BMI, and lipid profiles through health education paired with effective strategies such as lifestyle coaching and skill building. This is in line with recent recommendations from the United States and Europe, advocating that early and sustainable preventative efforts could spare the expense of managing T2DM and its long-term complications [48,49]. Middle-aged and older adults whose blood glucose falls within the normal range but have a higher METS-IR might remain at risk for T2DM and thus should still be monitored for disease occurrence through regular follow-up along with individualized risk assessment. This also provides an impetus for future studies to explore whether the biological underpinnings of METS-IR-T2DM may vary among individuals with different phenotypes of visceral obesity and genetic predispositions. The age of 45 years was used as a cutoff for middle age, which is in line with the eligibility criteria widely adopted in other large-scale population-based cohort studies such as the Healthy, Aging, and Retirement in Thailand Study [50], the Swedish EpiHealth Study [51], and the Canadian Longitudinal Study on Aging [52]. The Rotterdam Study, which was originally comprised of participants aged 45 years or older, has expanded the cohort that targeted participants aged 40 years or older since 2016 [53]. This echoes the shift toward earlier onset of T2DM [54,55], and thus future pragmatic trials to evaluate the use of METS-IR may be extended to the younger population. Alongside the ongoing efforts to translate team-based care into practice on a global scale [56], further steps to integrate the monitoring of METS-IR, rather than blood glucose alone, into the existing approach to diabetes prevention are especially promising. This would help inform broad public health policy and community-oriented preventive strategies that include early identification of individuals at risk of T2DM through innovative surveillance tools based on readily available routine physical examination indicators.

In conclusion, METS-IR was significantly associated with new-onset T2DM, regardless of baseline BP level. Regular monitoring of METS-IR on top of routine blood glucose in resource-limited settings may add to the ability to enhance early recognition and appropriate management of individuals at risk for T2DM in primary care.

**Acknowledgments**

We wish to acknowledge the tremendous support from the National School of Development at Peking University to ensure high-quality data collection. We also thank the Behavioral and Social Research Division of the National Institute on Aging of the National Institute of Health, the Natural Science Foundation of China, and the World Bank for their strong support of the original study. The present work was supported by the National Natural Science Foundation of China (72061137002). The study sponsor and funder were not involved in the design of the study, the collection, analysis, and interpretation of data, or writing the report, and they did not impose any restrictions regarding the publication of the report.

**Data Availability**

The original China Health and Retirement Longitudinal Study data were made available in the public domain by Peking University. The data sets generated or analyzed during the current study are available from the corresponding author on reasonable request.

**Authors’ Contributions**

HHXW and HC were responsible for the conceptualization of the work and writing the original draft. HC and ZJ were responsible for data curation. HC was responsible for the formal analysis. HHXW was responsible for methodology and supervision. YTL and JJW were responsible for project administration. ZJ and XY contributed to validation. YTL, YJX, and JH contributed to review and editing. All authors contributed to the interpretation of the data and read and approved the final manuscript.

**Conflicts of Interest**

None declared.
Multimedia Appendix 1
Diagram of study flow; Clinical parameters of participants by BP at baseline and new-onset T2DM at follow-up; Comparison of demographic characteristics between participants included in the final analysis and those excluded due to missing values; Sensitivity analyses; Predictive capability of baseline METS-IR on top of blood glucose for new-onset T2DM.

References


Abbreviations

- aHR: adjusted hazard ratio
- BP: blood pressure
- FPG: fasting plasma glucose
- HbA1c: hemoglobin A1c
- HDL-C: high-density lipoprotein cholesterol
- HEC: hyperinsulinemic-euglycemic clamp
IR: insulin resistance
METS-IR: metabolic score for insulin resistance
ROC: receiver operating characteristic
T2DM: type 2 diabetes mellitus

Edited by D Islek Yaras; submitted 03.06.23; peer-reviewed by Z Zhu, N Kakaletsis; comments to author 11.12.23; revised version received 29.02.24; accepted 29.03.24; published 03.06.24.

Please cite as:
Cheng H, Jia Z, Li YT, Yu X, Wang JJ, Xie YJ, Hernandez J, Wang HHX. Metabolic Score for Insulin Resistance and New-Onset Type 2 Diabetes in a Middle-Aged and Older Adult Population: Nationwide Prospective Cohort Study and Implications for Primary Care. JMIR Public Health Surveill 2024;10:e49617
URL: https://publichealth.jmir.org/2024/1/e49617
doi:10.2196/49617
PMID:38569189

©Hui Cheng, Zhihui Jia, Yu Ting Li, Xiao Yu, Jia Ji Wang, Yao Jie Xie, Jose Hernandez, Harry H X Wang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 03.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
The Incidence and Outcomes of Out-of-Hospital Cardiac Arrest During the COVID-19 Pandemic in South Korea: Multicenter Registry Study

Heekyung Lee1,2, MD, PhD; Jaehoon Oh1,3, MD, PhD‡; Hyuk Joong Choi1,2, MD, PhD; Hyungoo Shin1,2, MD, PhD; Yongil Cho1,3, MD, PhD; Juncheol Lee1,3, MD, PhD

1Department of Emergency Medicine, Hanyang University College of Medicine, Seoul, Republic of Korea
2Department of Emergency Medicine, Hanyang University Guri Hospital, Gyeonggi-do, Republic of Korea
3Department of Emergency Medicine, Hanyang University Hospital, Seoul, Republic of Korea
‡The Korean Cardiac Arrest Research Consortium (KoCARC) Investigators

Corresponding Author:
Jaehoon Oh, MD, PhD
Department of Emergency Medicine
Hanyang University College of Medicine
222 Wangsimni-ro, Seongdong-gu
Seoul, 04763
Republic of Korea
Phone: 82 2 2290 9829
Fax: 82 2 2290 9280
Email: ojjai@hanyang.ac.kr

Abstract

Background: The COVID-19 pandemic has profoundly affected out-of-hospital cardiac arrest (OHCA) and disrupted the chain of survival. Even after the end of the pandemic, the risk of new variants and surges persists. Analyzing the characteristics of OHCA during the pandemic is important to prepare for the next pandemic and to avoid repeated negative outcomes. However, previous studies have yielded somewhat varied results, depending on the health care system or the specific characteristics of social structures.

Objective: We aimed to investigate and compare the incidence, outcomes, and characteristics of OHCA during the prepandemic and pandemic periods using data from a nationwide multicenter OHCA registry.

Methods: We conducted a multicenter, retrospective, observational study using data from the Korean Cardiac Arrest Resuscitation Consortium (KoCARC) registry. This study included adult patients with OHCA in South Korea across 3 distinct 1-year periods: the prepandemic period (from January to December 2019), early phase pandemic period (from July 2020 to June 2021), and late phase pandemic period (from July 2021 to June 2022). We extracted and contrasted the characteristics of patients with OHCA, prehospital time factors, and outcomes for the patients across these 3 periods. The primary outcomes were survival to hospital admission and survival to hospital discharge. The secondary outcome was good neurological outcome.

Results: From the 3 designated periods, a total of 9031 adult patients with OHCA were eligible for analysis (prepandemic: n=2728; early pandemic: n=2954; and late pandemic: n=3349). Witnessed arrest (P<.001) and arrest at home or residence (P=.001) were significantly more frequent during the pandemic period than during the prepandemic period, and automated external defibrillator use by bystanders was lower in the early phase of the pandemic than during other periods. As the pandemic advanced, the rates of the first monitored shockable rhythm (P=.10) and prehospital endotracheal intubation (P<.001) decreased significantly. Time from cardiac arrest cognition to emergency department arrival increased sequentially (prepandemic: 33 min; early pandemic: 35 min; and late pandemic: 36 min; P<.001). Both survival and neurological outcomes worsened as the pandemic progressed, with survival to discharge showing the largest statistical difference (prepandemic: 385/2728, 14.1%; early pandemic: 355/2954, 12%; and late pandemic: 392/3349, 11.7%; P=.01). Additionally, none of the outcomes differed significantly between the early and late phase pandemic periods (all P>.05).

Conclusions: During the pandemic, especially amid community COVID-19 surges, the incidence of OHCA increased while survival rates and good neurological outcome at discharge decreased. Prehospital OHCA factors, which are directly related to
OHCA prognosis, were adversely affected by the pandemic. Ongoing discussions are needed to maintain the chain of survival in the event of a new pandemic.

**Trial Registration:** ClinicalTrials.gov NCT03222999; https://classic.clinicaltrials.gov/ct2/show/NCT03222999

**KEYWORDS**
heart arrest; cardiopulmonary resuscitation; SARS-CoV-2; mortality; outpatient; cardiac arrest; multicenter registry study; out-of-hospital cardiac arrest; heart attack; observational study; adult; older adults; analysis; pandemic; prepandemic; endemic; defibrillator; COVID-19

**Introduction**
At the end of December 2019, the outbreak of SARS-CoV-2 emerged in Wuhan, China, leading to the global pandemic of COVID-19 [1,2]. As of August 2023, more than 769 million confirmed cases and almost 7 million cumulative fatalities have been attributed to COVID-19 [3]. The emergence of SARS-CoV-2 variants and mutations has played a significant role in the pandemic’s persistence [4], and even after the end of the global pandemic, the risk of new variants and surges remains [5].

According to the current out-of-hospital cardiac arrest (OHCA) guidelines, the chain of survival (CoS) has been constantly emphasized to improve outcomes for patients with OHCA [6]. The OHCA CoS includes rapid recognition and activation of emergency response, early high-quality cardiopulmonary resuscitation (CPR) and defibrillation, and effective advanced life support interventions [6]. Each prehospital step experienced concurrent disruptions during the COVID-19 pandemic, resulting in a disruption of the OHCA CoS worldwide [7-9]. Notably, there was an increase in the incidence of OHCA, while rates of bystander CPR and automated external defibrillator (AED) use rate declined, especially during periods of heightened community transmission [10,11]. Additionally, a study using emergency medical services (EMS) data of South Korea reported that the proportion of prehospital return of spontaneous circulation (ROSC) decreased during the pandemic period [12].

During the pandemic era, outcomes of OHCA, including survival and neurological outcomes, deteriorated [9,10,13]. A previous meta-analysis reported that the ROSC rate, survival to hospital admission, and hospital discharge with good neurologic outcome (GNO) were significantly lower during the pandemic compared to the prepandemic period [10]. One recent registry-based cohort study reported that the survival rate was significantly decreased while the bystander CPR rate was stable [14]. Another study reported higher incidence of OHCA and decreased bystander CPR in regions with a high burden of pandemic [15]. A population-based nationwide study in Japan also reported that the pandemic is related to poorer neurologic outcomes and less AED use in patients with OHCA [16]. These poor outcomes could be closely related to changes in the prehospital phase during the pandemic, especially when considering the nature of OHCA.

Previous research has explored the impact of the COVID-19 pandemic on OHCA incidence, outcomes, and prehospital factors [10,17-19]. Analyzing the characteristics of OHCA during the pandemic is important to prepare for the next pandemic and to avoid repeated negative outcomes. However, previous studies have yielded somewhat varied results, depending on the health care system or the specific characteristics of social structures. Here, we aimed to investigate and compare the incidence, outcomes, and characteristics of OHCA between the prepandemic and pandemic periods, using data from a nationwide multicenter registry.

**Methods**
**Study Design**
We conducted a multicenter, retrospective, observational study using data from the Korean Cardiac Arrest Resuscitation Consortium (KoCARC) registry. This nationwide registry collates OHCA cases in South Korea in alignment with Utstein-style templates and involves a collaborative research network among multiple hospitals. The study included patients with OHCA who were transported to the emergency department (ED) by EMS and underwent resuscitation. Inclusion was limited to patients diagnosed with a medical cause identified by emergency physicians. Meanwhile, individuals with terminal illnesses, in hospice care, who are pregnant, or with a “Do Not Resuscitate” card were excluded. Patients with OHCA resulting from nonmedical causes, such as trauma, poisoning, burns, drowning, asphyxia, or hanging, were also excluded. Data were collected using a standardized registry form and uploaded to a web-based electronic database. The quality of the registry was monitored using a quality management committee to improve its reliability and integrity.

**Ethical Considerations**
The project was registered at ClinicalTrials.gov with the identifier NCT03222999 and received ethical approval from the institutional review boards of the 62 participating hospitals. Given the nature of this study, the requirement for informed consent was waived by the institutional review boards. All personal information of patients with OHCA enrolled in the registry was anonymized. There was no direct compensation for the included patients.

**Study Population**
This study included OHCA cases in South Korea registered in the KoCARC database in 3 distinct 1-year periods: the prepandemic period, early phase of the pandemic, and late phase of the pandemic. The KoCARC releases a year’s worth of data sets annually, with the period starting in July and ending in June of the following year. The prepandemic period was defined as...
the year leading up to the reporting of the first COVID-19 case (from January 1 to December 31, 2019). The early phase of the pandemic corresponds to the first data set released after the declaration of the pandemic (from July 1, 2020, to June 30, 2021). The late phase of the pandemic encompasses the most recent data set (from July 1, 2021, to June 30, 2022). Patients aged <18 years were excluded.

Data Extraction and Definition

Previously published literature has provided comprehensive details regarding the KoCARC database, including information on data elements and quality assurance [20]. We extracted the following data from the KoCARC database: (1) patient characteristics: age, sex, witnessed arrest, arrest location at home or residence, bystander response (CPR and AED), first monitored shockable rhythm, and prehospital management (defibrillation, adrenaline use, and advanced airway [supraglottic airway or endotracheal intubation]); and (2)prehospital time factors: cardiac arrest recognition to ED arrival, EMS call to EMS arrival, arrest place arrival to departure, arrest place departure to ED arrival, and cardiac arrest recognition toprehospital ROSC.

Outcome Variables

Primary outcomes were survival until hospital admission and survival to hospital discharge. We compared survival rates between the prepandemic period and the early and late pandemic periods and investigated the factors affecting survival rates. The secondary outcome was a good neurological outcome (GNO), which was defined as a Cerebral Performance Category (CPC) score of 1 or 2 at the time of hospital discharge. A CPC score of 1 denotes good cerebral performance, meaning patients are conscious, alert, and capable of functioning with only minor neurological or psychological deficits. A CPC score of 2 indicates moderate cerebral disability, with patients being conscious, possessing adequate cerebral function for independent daily activities, and having the ability to work in sheltered environments. A CPC score of 5 corresponds to mortality, specifically defined as death or brain death.

Statistical Analysis

This study used data from the KoCARC registry, which were meticulously compiled and organized using a spreadsheet application (Excel 365; Microsoft). Continuous baseline variables were presented as median and IQR, and their normal distribution was assessed using the Shapiro-Wilk test. In instances where the data deviated from the normality assumptions, the Kruskal-Wallis test was used to compare the groups and the Mann-Whitney U test was used for a post hoc test. Categorical variables were expressed as absolute counts and percentages and analyzed using the chi-square test. A significance threshold of \( P < .05 \) was adopted to establish statistical significance.

To ascertain the independent association between variables and neurological and survival outcomes, multivariate analysis was performed using logistic regression. Furthermore, to validate the calibration of the logistic model, the Hosmer-Lemeshow test was applied. The statistical software packages SPSS (version 27.0 for Windows; IBM) and MedCalc (version 17.2; MedCalc Software) were used to perform all statistical computations.

Results

From the 3 designated periods, a total of 9209 patients with OHCA were registered. After excluding patients aged <18 years, 9031 patients with OHCA were deemed eligible. The number of patients included in each period was as follows: 2728 in the prepandemic period, 2954 in the early phase of the pandemic, and 3349 in the late phase of the pandemic (Figures 1 and 2). There was a sequential increase in the incidence of OHCA at the participating hospitals throughout the pandemic.
The baseline and prehospital OHCA characteristics from the 3 periods, along with intergroup comparisons, are summarized in Table 1. Age was significantly higher in the late phase pandemic period compared to the prepandemic or early phase pandemic periods (prepandemic: 72 y; early pandemic: 72 y; and late pandemic: 73 y; \( P = .02 \)). However, there was no significant difference in the sex distribution across the periods \( (P = .55) \). Witnessed arrests \( (P < .001) \) and arrests at home or residence \( (P = .001) \) were significantly higher during the pandemic compared to the prepandemic period. Although bystander CPR was not statistically different among the 3 periods \( (P = .43) \), AED use by bystanders was lower in the early phase of the pandemic compared to the other periods. The rate of the first monitored shockable rhythm and prehospital defibrillation decreased sequentially throughout the study period. The prehospital endotracheal intubation rate significantly decreased as the pandemic progressed (prepandemic: 246/2728, 9%; early pandemic: 154/2954, 5.2%; and late pandemic: 116/3349, 3.5%; \( P < .001 \)), resulting in a significant increase in supraglottic airway use (prepandemic: 1685/2728, 61.9%; early pandemic: 2158/2954, 73.3%; and late pandemic: 2408/3349, 72.6%; \( P < .001 \)).

During the pandemic, pre–hospital phase time factors were extended compared to the prepandemic period. The time from cardiac arrest recognition to ED arrival increased sequentially (prepandemic: 33 min; early pandemic: 35 min; and late pandemic: 36 min; \( P < .001 \)). Other time-related factors also worsened during the pandemic, including EMS call to arrival, arrest place arrival to departure, and arrest place departure to ED arrival.

Primary and secondary outcomes are summarized in Table 2. As the pandemic progressed, both survival and neurological outcomes worsened. The survival-to-discharge rate exhibited the most pronounced statistical variation (prepandemic: 385/2728, 14.1%; early pandemic: 355/2954, 12%; and late pandemic: 392/3349, 11.7%; \( P = .01 \)). The survival-to–hospital admission rate was 24.3% (815/3349) in the late phase of the pandemic, which was significantly lower than that in the prepandemic period \( (P = .02) \). However, there was no statistically significant difference between the prepandemic and early phase pandemic periods \( (P = .08) \). Both the survival-to-discharge \( (P = .02 \) and \( P = .005 \), respectively) and GNO \( (both \ P = .03) \) rates were significantly lower in the early and late phase pandemic periods than in the prepandemic period. Additionally, none of the outcomes differed significantly between early and late phase pandemic periods \( (all \ P > .05) \).
<table>
<thead>
<tr>
<th>Variables</th>
<th>Prepandemic period&lt;sup&gt;a&lt;/sup&gt; (n=2728)</th>
<th>Early phase pandemic period&lt;sup&gt;b&lt;/sup&gt; (n=2954)</th>
<th>Late phase pandemic period&lt;sup&gt;c&lt;/sup&gt; (n=3349)</th>
<th>P value&lt;sup&gt;d,e&lt;/sup&gt; (pre- vs early pandemic)</th>
<th>P value&lt;sup&gt;d,e&lt;/sup&gt; (pre- vs late pandemic)</th>
<th>P value&lt;sup&gt;d,e&lt;/sup&gt; (early vs late pandemic)</th>
<th>P value&lt;sup&gt;d,e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y), median (IQR)</strong>&lt;sup&gt;f&lt;/sup&gt;</td>
<td>72 (59-81)</td>
<td>72 (59-81)</td>
<td>73 (60-82)</td>
<td>.45</td>
<td>.06</td>
<td>.005</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Sex (male), n (%)</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1780 (65.2)</td>
<td>1950 (66)</td>
<td>2230 (66.6)</td>
<td>.56</td>
<td>.28</td>
<td>.63</td>
<td>.55</td>
</tr>
<tr>
<td><strong>Witnessed arrest, n (%)</strong></td>
<td>1580 (58)</td>
<td>1860 (63.1)</td>
<td>2086 (62.9)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.92</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Arrest location, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home or residence</td>
<td>1625 (59.6)</td>
<td>1899 (64.4)</td>
<td>2066 (62.3)</td>
<td>&lt;.001</td>
<td>.03</td>
<td>.09</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Bystander response, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bystander CPR&lt;sup&gt;i&lt;/sup&gt;</td>
<td>1443 (56.1)</td>
<td>1572 (54.5)</td>
<td>1798 (55.8)</td>
<td>.23</td>
<td>.80</td>
<td>.31</td>
<td>.43</td>
</tr>
<tr>
<td>AED&lt;sup&gt;j&lt;/sup&gt; use by bystander</td>
<td>43 (1.7)</td>
<td>26 (0.9)</td>
<td>54 (1.7)</td>
<td>.01</td>
<td>.98</td>
<td>.008</td>
<td>.02</td>
</tr>
<tr>
<td><strong>First monitored rhythm, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shockable rhythm</td>
<td>496 (19.1)</td>
<td>518 (18.1)</td>
<td>538 (16.9)</td>
<td>.38</td>
<td>.03</td>
<td>.21</td>
<td>.10</td>
</tr>
<tr>
<td>Prehospital defibrillation, n (%)</td>
<td>621 (23.5)</td>
<td>665 (23)</td>
<td>726 (22.3)</td>
<td>.67</td>
<td>.27</td>
<td>.51</td>
<td>.54</td>
</tr>
<tr>
<td>Prehospital adrenaline use, n (%)</td>
<td>505 (19.6)</td>
<td>720 (25.1)</td>
<td>656 (20.3)</td>
<td>&lt;.001</td>
<td>.52</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Prehospital airway, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supraglottic airway</td>
<td>1685 (61.9)</td>
<td>2158 (73.3)</td>
<td>2408 (72.6)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.56</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Endotracheal intubation</td>
<td>246 (9)</td>
<td>154 (5.2)</td>
<td>116 (3.5)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Cardiac arrest cognition to ED&lt;sup&gt;k&lt;/sup&gt; arrival (min), median (IQR)</td>
<td>33 (26-44)</td>
<td>35 (29-45)</td>
<td>36 (28-46)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.39</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>EMS&lt;sup&gt;l&lt;/sup&gt; call to EMS arrival (min), median (IQR)</td>
<td>7 (5-10)</td>
<td>9 (7-12)</td>
<td>9 (7-13)</td>
<td>.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Arrest place arrival to departure (min), median (IQR)</td>
<td>13 (9-18)</td>
<td>15 (11-19)</td>
<td>15 (11-19)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.45</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Arrest place departure to ED arrival (min), median (IQR)</td>
<td>9 (6-13)</td>
<td>10 (7-14)</td>
<td>10 (7-15)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>.31</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Cardiac arrest cognition to prehospital ROSC&lt;sup&gt;m&lt;/sup&gt; (min), median (IQR)</td>
<td>16 (11-25)</td>
<td>20 (12-28)</td>
<td>19 (13-26)</td>
<td>.003</td>
<td>.03</td>
<td>.29</td>
<td>.008</td>
</tr>
</tbody>
</table>

<sup>a</sup>Prepandemic period: from January to December 2019.

<sup>b</sup>Early phase pandemic period: from July 2020 to June 2021.

<sup>c</sup>Late phase pandemic period: from July 2021 to June 2022.

<sup>d</sup>Mann-Whitney U test (for continuous variables).

<sup>e</sup>P<.05 was significant.

<sup>f</sup>Kruskal-Wallis test (for continuous variables).

<sup>g</sup>Continuous variables are presented as median (IQR) and tested by using the Mann-Whitney U or Kruskal-Wallis test.

<sup>h</sup>Categorical variables are presented as n (%) and tested by using the chi-squared test.

<sup>i</sup>CPR: cardiopulmonary resuscitation.

<sup>j</sup>AED: automated external defibrillator.

<sup>k</sup>ED: emergency department.

<sup>l</sup>EMS: emergency medical services.

<sup>m</sup>ROSC: return of spontaneous circulation.
### Table 2. Comparison of primary and secondary outcomes between prepandemic and pandemic periods using the nationwide, multicenter, out-of-hospital cardiac arrest registry of South Korea

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Prepandemic period</th>
<th>Early phase pandemic period</th>
<th>Late phase pandemic period</th>
<th>P value (pre vs early pandemic)</th>
<th>P value (pre vs late pandemic)</th>
<th>P value (early vs late pandemic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival to hospital admission, n (%)</td>
<td>734 (26.9)</td>
<td>734 (24.8)</td>
<td>815 (24.3)</td>
<td>.08</td>
<td>.02</td>
<td>.64</td>
</tr>
<tr>
<td>Survival to discharge, n (%)</td>
<td>385 (14.1)</td>
<td>355 (12)</td>
<td>392 (11.7)</td>
<td>.02</td>
<td>.005</td>
<td>.70</td>
</tr>
<tr>
<td>Good neurologic outcome, n (%)</td>
<td>260 (9.5)</td>
<td>234 (7.9)</td>
<td>265 (7.9)</td>
<td>.03</td>
<td>.03</td>
<td>.99</td>
</tr>
</tbody>
</table>

- Prepandemic period: from January 2019 to December 2019.
- Early phase pandemic period: from July 2020 to June 2021.
- Late phase pandemic period: from July 2021 to June 2022.

Multivariate analyses of the factors related to survival and neurological outcomes were performed to adjust for confounders that could affect the primary outcome (Table 3). Significant independent effects on all outcomes were observed for age (P<.001), witnessed arrest (P<.001), arrest at home or residence (P<.001), first monitored shockable rhythm (P<.001), prehospital adrenaline use (P<.001 to P=.003), and supraglottic airway (P<.001). Cardiac arrest cognition at ED arrival significantly affected GNO after adjusting for confounders (adjusted odds ratio 1.001, 95% CI 1.000-1.002; P=.04). The difference in period (prepandemic, early phase, or late phase) did not independently affect any of the 3 outcomes.

### Table 3. Multivariate analysis of factors affecting survival and neurological outcomes using the nationwide, multicenter, out-of-hospital cardiac arrest registry of South Korea.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Survival to hospital admission</th>
<th>Survival to discharge</th>
<th>Good neurologic outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted OR (95% CI)</td>
<td>P value</td>
<td>Adjusted OR (95% CI)</td>
</tr>
<tr>
<td>Age (per year)</td>
<td>0.971 (0.967-0.974)</td>
<td>&lt;.001</td>
<td>0.965 (0.960-0.970)</td>
</tr>
<tr>
<td>Witnessed arrest (yes or no)</td>
<td>1.993 (1.747-2.275)</td>
<td>&lt;.001</td>
<td>2.122 (1.732-2.600)</td>
</tr>
<tr>
<td>Arrest at home or residence (yes or no)</td>
<td>0.709 (0.628-0.799)</td>
<td>&lt;.001</td>
<td>0.599 (0.507-0.708)</td>
</tr>
<tr>
<td>AEDb use by bystander (yes or no)</td>
<td>0.845 (0.513-1.396)</td>
<td>.51</td>
<td>1.530 (0.861-2.717)</td>
</tr>
<tr>
<td>First monitored rhythm (shockable or nonshockable)</td>
<td>4.667 (4.080-5.338)</td>
<td>&lt;.001</td>
<td>9.061 (7.662-10.716)</td>
</tr>
<tr>
<td>Prehospital adrenaline (used or not used)</td>
<td>0.756 (0.651-0.877)</td>
<td>&lt;.001</td>
<td>0.723 (0.583-0.896)</td>
</tr>
<tr>
<td>Prehospital supraglottic airway (performed or not performed)</td>
<td>0.750 (0.654-0.861)</td>
<td>&lt;.001</td>
<td>0.455 (0.380-0.545)</td>
</tr>
<tr>
<td>Prehospital endotracheal intubation (performed or not performed)</td>
<td>0.867 (0.657-1.145)</td>
<td>.32</td>
<td>0.556 (0.374-0.826)</td>
</tr>
<tr>
<td>Cardiac arrest cognition to EDc arrival (per min)</td>
<td>1.000 (1.000-1.001)</td>
<td>.48</td>
<td>1.001 (1.000-1.002)</td>
</tr>
</tbody>
</table>

**Period when OHCA occurred**

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>Reference</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepandemic period</td>
<td>Reference</td>
<td>—e</td>
<td>Reference</td>
</tr>
<tr>
<td>Early phase pandemic period</td>
<td>0.961 (0.867-1.065)</td>
<td>.45</td>
<td>0.995 (0.861-1.150)</td>
</tr>
<tr>
<td>Late phase pandemic period</td>
<td>1.015 (0.918-1.123)</td>
<td>.77</td>
<td>0.962 (0.836-1.107)</td>
</tr>
</tbody>
</table>

- OR: odds ratio.
- AED: automated external defibrillator.
- ED: emergency department.
- Not applicable.
Discussion

Principal Findings

In this study, we aimed to investigate changes in OHCA characteristics and outcomes during the SARS-CoV-2 pandemic using multicenter registry-based data. The incidence of OHCA increased, and neurologic and survival outcomes were worse during the pandemic era compared to the prepandemic era. Outcomes were particularly poor in the late phase when the number of SARS-CoV-2 infection cases spiked due to the spread of the Delta and Omicron variants [21].

During the pandemic, a significant increase in the global incidence of OHCA was observed compared to the prepandemic period. In cities such as New York, Detroit, and London, the number of OHCA cases increased proportionally with the number of COVID-19 cases during the initial surge [17-19]. In South Korea, the number of out-of-hospital sudden cardiac arrests increased by approximately 8%, from 60 per 100,000 population in 2019 (the year preceding the COVID-19) to 64.7 per 100,000 population in 2021 [22]. One meta-analysis that was reported early in the pandemic, which included 10 studies with a total of 35,000 participants published before October 2020, found that the incidence of OHCA increased by approximately 2.2 fold compared to the prepandemic period [10].

Globally, the outcomes of OHCA during the pandemic have consistently been reported to be less favorable than those observed before the pandemic [13]. For instance, in 2 regions of the United States, the survival-to-discharge rate decreased from 14.7% in 2019 to 7.9% in 2020 [23]. Meanwhile, in Singapore, while the incidence of OHCA increased during the pandemic, prehospital ROSC decreased [24]. In South Korea, the survival-to-discharge rate of OHCA, which had been steadily increasing for more than a decade before the pandemic, dropped from 8.7% in 2019 to 7.5% in 2020 and 7.3% in 2021. Moreover, the rate of brain function recovery at discharge reached 5.4% in 2019 but dropped to 4.9% in 2020 and 4.4% in 2021 [22]. In addition, a previous meta-analysis reported that the outcomes of OHCA were significantly poorer than before the pandemic in most of the included studies. The meta-analysis also showed a significant increase in mortality [10,25].

The characteristics of OHCA shifted during the pandemic periods in this study. Instances of witnessed OHCA and arrests at home or residence became more frequent in the pandemic era. This change has been observed in previous studies and can be attributed to the increased time spent at home during the pandemic due to quarantine and social distancing [10,26]. These results may also reflect the number of patients who progressed to OHCA due to being unable to visit a hospital in a timely manner as a result of reduced access to hospitals. Although bystander CPR decreased slightly during the pandemic, the difference was not statistically significant. This study also found that bystander use of AEDs decreased in the early phase of the pandemic but returned to prepandemic levels in the late phase of the pandemic. This is likely a reflection of bystander fears of COVID-19 transmission in the early phase of the pandemic [27]. A previous meta-analysis reported a decreased proportion of OHCAs with shockable rhythms during the pandemic [10,25]. In our research, the frequency of the first monitored shockable rhythm decreased progressively. The rise in respiratory arrests due to respiratory failure due to COVID-19 may have contributed to the diminished proportion of OHCA cases with shockable rhythms. Such cases would likely have a more unfavorable prognosis than OHCA cases of shockable rhythm with an etiology of cardiac origin [28]. Notably, the age demographic skewed older in the late phase of the pandemic, suggesting a possible increase in OHCA among older patients, who are more vulnerable to COVID-19.

Prehospital airway management has also been affected by the pandemic. An updated meta-analysis reported an increase in the use of both endotracheal intubation and supraglottic airway devices during the pandemic [25]. However, in this study, the supraglottic airway was favored, while endotracheal intubation progressively decreased during the pandemic. This is likely due to the EMS personnel’s fear of contracting COVID-19 during endotracheal intubation and the physical restrictions imposed by wearing protective clothing, face shields, and goggles. The summed rates of supraglottic airway and endotracheal intubation was higher than that before the pandemic, suggesting that more advanced resuscitation was performed in the prehospital setting.

Transport time, including the time from cardiac arrest diagnosis to ED arrival, significantly increased during the pandemic era. During the surge in patients with COVID-19, the lack of medical resources, poor access to hospitals, requirements for ambulance disinfection, and the need to wear protective equipment may have contributed to the increase in transportation time [29]. Difficulties in hospital selection led to prolonged, unnecessary prehospital treatment, and the frequency of adrenaline use increased during the pandemic. Further, prolonged transport time may have directly contributed to poorer outcomes in patients with OHCA during the pandemic. Furthermore, it emerged as an independent factor for poorer neurological outcomes in the multivariate analysis of this study. Such delays could compromise the continuity and quality of advanced life support from prehospital settings to hospitals, as underscored in the CoS, ultimately leading to suboptimal patient outcomes.

The prehospital factors consistently emphasized in the CoS, which play a key role in OHCA outcomes, were also heavily impacted by the pandemic. Maintaining a CoS in the event of future outbreaks and the emergence of new variants should be an ongoing discussion.

Limitations

This study has several limitations. First, depending on the country and region, the impact of the COVID pandemic on OHCA may vary. Our findings were derived using only patients with OHCA in South Korea, and there may be limits to extending the results globally. Second, although the KoCARC registry was prospectively collected using established protocols and methods, the aim to examine the impact of the pandemic on OHCA was not predetermined before the initiation of the registry. Third, we were unable to investigate whether SARS-CoV-2 infection was the direct cause of cardiac arrest in the enrolled patients with OHCA. However, the main contribution of this study was to investigate the impact of...
prehospital factors associated with the CoS during the pandemic. Forth, despite the comprehensive data collection employing Utstein-style templates within the registry, the potential remains for latent confounders to influence the results. Lastly, long-term prognoses were not included in this analysis due to the limited observation period.

Conclusions
The incidence of OHCA increased during the pandemic, while both the survival rate and GNO at discharge decreased, especially during community COVID-19 surges. Prehospital OHCA factors, which directly influence OHCA prognosis, were adversely affected by the pandemic. Ongoing discussions are needed to maintain the CoS in anticipation of future pandemics.

Acknowledgments
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the South Korea government (Ministry of Science and ICT [MSIT]; No.RS-2023-00277934). This work was also supported by the research fund of Hanyang University (HY-202300000001164).

We would like to acknowledge and thank the following:
• Investigators from all participating hospitals of the Korean Cardiac Arrest Resuscitation Consortium (KoCARC): Do Kyun Kim (Seoul National University Hospital), Sang Suk Han, Phil Cho Choi (Kangbuk Samsung Medical Center), Sang O Park, Jong Won Kim (Konkuk University Medical Center), Han Sung Choi (Kyung Hee University Hospital), Sung Hyuk Choi (Korea University Guro Hospital), Min Seob Sim, Gun Tak Lee (Samsung Medical Center), Shin Ahn (Asan Medical Center), Jong Whan Shin (SMG-SNU Boramae Medical Center), Sang Hyun Park, Keun Hong Park (Seoul Medical Center), In Cheol Park (Yonsei University Severance Hospital), Tae Young Kong (Yonsei University Gangnam Severance Hospital), Kyoung Won Lee, Chu Hyun Kim (Inje University Seoul Paik Hospital), Youngsuk Cho (Hallym University Kangdong Sacred Heart Hospital), Gu Hyun Kang, Yong Soo Jang (Hallym University Kangnam Sacred Heart Hospital), Seok Ran Yeom, Sang Kyoon Han (Pusan National University Hospital), Jae Hoon Lee (Dong-A University Hospital), Jeong Bae Park, Hyun Wook Ryoo (Kyoungpook National University Hospital), Kyung Woo Lee, Tae Chang Jang (Daegu Catholic University Medical Center), Jae-hyug Woo (Gachon University Gil Medical Center), Woon Jeong Lee, Seon Hee Woo (The Catholic University of Korea Incheon St. Mary's Hospital), Sung Hyun Yun, Tae Jin Cho (Catholic Kwandong University International St. Mary's Hospital), Sun Pyo Kim, Yong Jin Park (Chosun University Hospital), Jin Woong Lee, Wonjoon Jeong (Chungnam National University Hospital), Sung Soo Park, Jae Kwang Lee (Konyang University Hospital), Ryeok Ahn, Wook Jin Choi (Ulsan University Hospital), Young Gi Min, Eun Jung Park (Ajou University Hospital), Joong Hee Kim (Seoul National University Bundang Hospital), In Byung Kim, Ki Ok Ahn (Myongji Hospital), Han Jin Cho (Korea University Ansan Hospital), Seung Cheol Lee, Sang Hun Lee (Dongguk University Ilsan Hospital), Young Sik Kim, Young Rock Ha (Bundang Jesaeng Hospital), Jin Sik Park, Myoung Woo Lee (Sejong Hospital). Dai Han Wi (Wonkwang University Sanbon Hospital), Sang Ok Ha, Won Seok Yang (Hallym University Pyeongchon Sacred Heart Hospital), Ok Jun Kim, Tae Nyoung Chung (Cha University Bundang Medical Center), Soon Joo Wang, Hang A Park (Hallym University Dongtan Sacred Heart Hospital), Jun Hwi Cho, Chan Woo Park (Kangwon National University Hospital), An Muh Eob, Tae Hun Lee (Hallym University Chuncheon Sacred Heart Hospital), Sang Chul Kim, Hoon Kim (Chungbuk National University Hospital), Han Joo Choi, Chan Young Koh (Dankook University Hospital), Jung Won Lee, Dong Wook Lee (Soonchunhyang University Cheonan Hospital), Tae Oh Jung, Jae Choi Yoon (Chonbuk National University Hospital), Dae Hai Choi, Jung Tae Choi (Dongguk University Gyeongju Hospital), Jin Hee Jeong, Soo Hoon Lee (Gyeongsang National University Hospital), Ji Ho Ryu, Maeng Real Park (Pusan National University Yangsan Hospital), Won Kim (Cheju Halfa General Hospital), Sung Wook Song, Woo Jung Kim (Jeju National University Hospital), Joon-myoung Kwon, Eui Hyuk Kang (Mediplex Sejong Hospital), Sang Chan Jin, Tae-ekwon Kim (Keimyung University Dongsan Medical Center), and Seong Chun Kim (Gyeongsang National University Changwon Hospital)
• The steering committee, comprised of following individuals: Sung Oh Hwang (Chair, Wonju Severance Christian Hospital), Sang Do Shin (Chair of Steering Committee, Seoul National University hospital), Hyuk Jun Yang (Advisory Committee, Gachon University Gil hospital), Sung Phil Chung (Data Safety and Management Board, Yonsei University Gangnam Severance Hospital), Sung Woo Lee (Security and Monitoring Board, Korea University Anam hospital), Kyung Jun Song (Secretariat, SMG-SNU Boramae Medical Center), Seung Sik Hwang (Epidemiology and Prevention Research Committee, Seoul National University), Gyu Chong Cho (Community Resuscitation Research Committee, Hallym University Kangdong Sacred Heart Hospital), Sung Woo Moon (Emergency Medical Service Resuscitation Research Committee, Korea University Ansan Hospital), Kyoung Chul Cha (Hospital Resuscitation Research Committee, Wonju Severance Christian Hospital), Won Young Kim (Hypothermia and Post-Resuscitation Care Research Committee, Asan Medical Center), Sung Hoon Na (Cardiac Care Resuscitation Research Committee, Seoul National University Hospital), and Young Ho Kwack (Pediatric Resuscitation Research Committee, Seoul National University hospital),
• The members of the secretariat: Joo Yeong Kim (Korea University Ansan hospital), Jeong Ho Park (Seoul National University hospital), Sun Young Lee (Seoul National University hospital), and Jung Eun Kim (Seoul National University hospital),
• The National Fire Agency for providing prehospital emergency medical services (EMS) data, and
• The Korean Association of Cardiopulmonary Resuscitation (KACPR) for their support.
Data Availability
The data sets used and analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References


Abbreviations

AED: automated external defibrillator
CoS: chain of survival
CPC: Cerebral Performance Category
CPR: cardiopulmonary resuscitation
ED: emergency department
EMS: emergency medical services
GNO: good neurologic outcome
KoCARC: Korean Cardiac Arrest Resuscitation Consortium
OHCA: out-of-hospital cardiac arrest
ROSC: return of spontaneous circulation

Edited by A Mavragani; submitted 01.09.23; peer-reviewed by S Schnaubelt, E Kawamoto; comments to author 08.12.23; revised version received 28.12.23; accepted 16.05.24; published 24.06.24.

Please cite as:
URL: https://publichealth.jmir.org/2024/1/e52402
doi:10.2196/52402
PMID:
Original Paper

Association Between Socioeconomic Inequalities in Pain and All-Cause Mortality in the China Health and Retirement Longitudinal Study: Longitudinal Cohort Study

Zhuo Zhang¹, PhD; Dongmei Xue¹, MSc; Ying Bian¹, PhD
Institute of Chinese Medical Sciences, State Key Laboratory of Quality Research in Chinese Medicine, University of Macau, Taipa, Macao

Corresponding Author:
Ying Bian, PhD
Institute of Chinese Medical Sciences
State Key Laboratory of Quality Research in Chinese Medicine
University of Macau
Room 1048, Building E12
Avenida da Universidade
Taipa, 999078
Macao
Phone: 853 88228537
Email: bianyingum@163.com

Abstract

Background: Few studies focus on the equality of pain, and the relationship between pain and death is inconclusive. Investigating the distribution of pain and potential mortality risks is crucial for ameliorating painful conditions and devising targeted intervention measures.

Objective: Our study aimed to investigate the association between inequalities in pain and all-cause mortality in China.

Methods: Longitudinal cohort data from waves 1 and 2 of the China Health and Retirement Longitudinal Study (2011-2013) were used in this study. Pain was self-reported at baseline, and death information was obtained from the 2013 follow-up survey. The concentration index and its decomposition were used to explain the inequality of pain, and the association between pain and death was analyzed with a Cox proportional risk model.

Results: A total of 16,747 participants were included, with an average age of 59.57 (SD 9.82) years. The prevalence of pain was 32.54% (8196/16,747). Among participants with pain, the main pain type was moderate pain (1973/5426, 36.36%), and the common pain locations were the waist (3232/16,747, 19.3%), legs (2476/16,747, 14.78%) and head (2250/16,747, 13.44%). We found that the prevalence of pain was concentrated in participants with low economic status (concentration index –0.066, 95% CI –0.078 to –0.054). Educational level (36.49%), location (36.87%), and economic status (25.05%) contributed significantly to the inequality of pain. In addition, Cox regression showed that pain was associated with an increased risk of all-cause mortality (hazard ratio 1.30, 95% CI 1.06-1.61).

Conclusions: The prevalence of pain in Chinese adults is concentrated among participants with low economic status, and pain increases the risk of all-cause death. Our results highlight the importance of socioeconomic factors in reducing deaths due to pain inequalities by implementing targeted interventions.

(JMIR Public Health Surveill 2024;10:e54309) doi:10.2196/54309

KEYWORDS
pain; equality; all-cause mortality; concentration index; decomposition

Introduction

Body pain is both a sensory and emotional experience linked to actual or potential tissue damage, incorporating physiological as well as emotional and psychological dimensions [1]. Pain is now recognized as an independent pathological condition, not merely a symptom of other diseases. The biopsychosocial framework for understanding pain emphasizes the complex and interactive relationships between biological, psychological, and social determinants. Pain creates a heavy burden of disease, with chronic pain being the predominant type [2]. According to data from the 2021 US National Health Interview Survey,
approximately 51.6 million US adults (20.9%) have chronic pain and 17.1 million (6.9%) experience high-impact chronic pain [3]. This scenario not only escalates the demand for health care resources and medical expenditures but also leads to diminished work efficiency and productivity loss, thereby exacerbating the economic burden. There is a significant correlation between pain and psychological health disorders, such as anxiety and depression, which further impacts the quality of life of patients [4]. Family members also experience additional pressures due to the caregiving responsibilities for individuals with pain. The management of pain, especially chronic pain, often involves the use of opioid medications, which increases the risk of drug dependency [5]. Notably, considering factors such as time, economic conditions, and pain perception, many middle-aged and older adults choose self-management strategies, including medication, acupuncture, and massage, to alleviate pain. Research also shows a significant rise in pain incidence in the last years of life, climbing from 26% to 46% [6].

The global trend of population aging is becoming increasingly significant. With advancing age, the issues of chronic musculoskeletal joint pain, neuropathic pain, vascular-related pain, and pain triggered by psychological factors are gradually intensifying. Research on the factors influencing pain and its treatment behavior in the older population in the northeast region of China indicates that the prevalence of physical pain is as high as 32.5% among people aged 60 years or older [7]. The Global Burden of Disease study also revealed that chronic pain ranks first among high-risk factors for nonlethal health impairments [8]. In the 2019 Global Burden of Disease study, it was further revealed that migraine ranks second among factors causing disabilities, and among women younger than 50 years, migraine becomes the leading cause of disability [9]. Each year, more than 10 million new cases of malignant tumors occur globally, with 30% to 60% of patients experiencing varying degrees of pain. In late-stage patients, the prevalence of pain can even reach up to 90%. Cancer pain, as one of the major complications of malignant tumors, not only greatly affects the quality of life of patients but also becomes one of the primary reasons for seeking medical attention. Healthy People 2030 emphasizes assisting individuals with high-impact chronic pain in managing their pain safely and reducing its effects [10]. Consequently, cancer pain control has been listed as one of the 4 key priorities in the World Health Organization’s cancer treatment planning [11].

Investigating the distribution of pain is crucial for ameliorating painful conditions and devising targeted intervention measures. Previous research has predominantly focused on the unequal distribution of pain across gender, race, and disease management. This is chiefly due to the understanding of pain as an issue intricately connected to fairness and justice, set against a societal framework marked by discrimination and structural violence [12]. However, there has been relatively scant attention given to studies investigating the distribution of pain among different socioeconomic status (SES) groups, especially in developing countries. Socioeconomic factors encompass indicators such as education, occupation, economic status, and social support. The association between socioeconomic factors and chronic diseases has been reported in numerous studies, linking lower SES to higher prevalence and severity of pain. However, results pertaining to the association between socioeconomic factors and pain are sometimes contradictory, potentially due to the variables chosen for SES, the methods used, and the procedures implemented. Macro data indicates that the prevalence of any type of pain in lower- and middle-income countries is twice that in high-income countries. Mortality rates may be influenced by pain, as evidenced by a cohort study and meta-analysis comprising over 500,000 research participants, which found that individuals with chronic widespread pain have a higher mortality rate [13]. On the other hand, some studies have arrived at opposite conclusions. A meta-analysis revealed only a moderate and nonsignificant association between chronic pain and all-cause mortality, with the degree of association being slightly stronger for widespread pain, though the combined estimate still remained nonsignificant [14].

Despite previous studies analyzing the relationship between pain and all-cause mortality, research focusing on the inequality of pain and its relation to all-cause mortality rates is scarce, particularly in developing countries. Moreover, even in large cohorts, results are not entirely consistent, necessitating further research. Therefore, we have used data from the China Health and Retirement Longitudinal Study (CHARLS) survey to quantify the association between pain in Chinese adults and all-cause mortality, and additionally, we have explored the unequal conditions of pain.

Methods

Data Sources and Study Population

The data used in this study came from CHARLS, which conducted a national baseline survey in 2011 and followed up in 2013, 2015, and 2018. The project conducted multistage sampling in 150 counties and 450 communities (villages) in 28 provinces (autonomous regions and municipalities directly under the central government), and conducted questionnaire surveys and face-to-face interviews with middle-aged and older individuals (aged ≥45 years) and their families, aiming to analyze the current situation of aging in China and promote high-quality research to solve problems related to aging. This study used baseline data from 2011 (wave 1), excluding participants with missing information on pain, economic status, and sociodemographic characteristics (n=469) and those aged <45 years (n=492); a total of 16,747 participants were included. Considering that data were available for interview status (ie, indicating that the participant was either alive or dead) and exact time of death in 2013 (wave 2), we used the 2013 data to account for all-cause deaths; a total of 394 all-cause deaths were followed. The screening flow chart is shown in Multimedia Appendix 1, Figure S1.

Pain, Economic Status, and All-Cause Mortality

Pain was the primary independent variable in this study. Pain was measured using a questionnaire that asked participants whether they had pain, as well as the severity and location of the pain. This study used the CHARLS 2011 question “Are you often troubled with any body pains?” to determine whether the
participants had pain problems and the question “How bad is your pain?” to determine the severity of pain. The response options included mild, moderate, and severe. Moreover, the study used the question “In what part of your body do you feel pain?” to determine the location of pain, including head, shoulders, arms, wrists, fingers, chest, stomach, back, waist, buttocks, legs, knees, ankles, toes, and neck. In the Constructed Expenditure, Income, and Wealth Database published by CHARLS in 2017, the indicator IM_PCE was used for economic status. All-cause death was the primary outcome of the study, and information on deaths came from the 2013 survey. There were no secondary outcome events, and follow-up was terminated when an outcome event occurred.

Covariates
We defined covariates using information collected from the baseline survey. The covariates and grouping in this study included age (45-60, 60-75, and ≥75 years), gender (male and female), education level (illiterate, primary school, secondary/high school, and university or above), marital status (never married; married; and separated, divorced, or widowed), smoking status (nonsmokers and current smokers), drinking status (never, occasionally, and regularly), BMI (<25 kg/m², 25-30 kg/m², and ≥30 kg/m²), and location (rural village or urban community). BMI was calculated by dividing weight in kilograms by the square of height in meters (kg/m²).

Statistical Analysis
This study describes the pain status and sociodemographic characteristics of the participants. Using CHARLS sampling weights (ind_weight_ad2), the data were extrapolated to estimate the percentage of Chinese adults aged ≥45 years experiencing pain. Continuous variables were represented as the mean and SD, and a 2-tailed t test was used for comparisons between groups. Categorical variables were represented as numbers and percentages, and comparisons between groups were performed using the χ² test. Logistic regression was used to describe the impact of sociodemographic characteristics on pain. Second, the concentration index was used to analyze the equality of pain, and the economic status indicator was per capita expenditure. The concentration index was defined as twice the area between the curve and the diagonal, ranging from –1 to 1. A positive concentration index value indicated that the disease distribution showed a trend to occur among groups with higher economic status, while a negative value indicated the opposite. The closer the concentration index was to 0, the more equal was the disease distribution [15]. Decomposition analysis was further used to describe the contribution of each factor to inequality. The reliability of using the concentration index and its decomposition to measure pain inequalities has been well established in previous studies [16]. Finally, a Cox proportional hazards regression model was used to explore the relationship between pain and all-cause death, and further stratified analysis was conducted based on sociodemographic characteristics. All reported P values were 2-sided, and statistical analyses were conducted with Stata (version 16.0; StataCorp).

The concentration index was calculated in detail as follows:

In the formula, R_i denotes the proportion of individuals ranked by SES, y_i denotes the prevalence of pain, and μ denotes the average prevalence of pain.

The decomposition method for the concentration index allowed decomposing the index into contributions from various influencing factors. The associated linear regression model is specified below:

In the formula, β_k denotes the regression coefficient. The concentration index was decomposed using the equation provided below:

In the formula, C_k, β_k, and indicate the concentration index, regression coefficient, and contribution, respectively.

Ethical Considerations
All practices involving human participants in this study were conducted in strict adherence to the principles of the Declaration of Helsinki. Ethical clearance for the initial CHARLS data collection was obtained from the Peking University Ethical Review Committee (IRB00001052-11015). Moreover, every participant in the CHARLS cohort willingly provided their written informed consent, ensuring ethical compliance and participant awareness. The present analysis received approval from the University of Macau’s research ethics subpanel committee (BSERE21-APP012-ICMS). All respondents signed informed consent forms, and the project was reviewed by the Institutional Review Board of Peking University.

Results
Pain Prevalence and Sociodemographic Characteristics of the Participants
A total of 16,747 participants were included in the study, with an average age of 59.57 (SD 9.82) years, of whom 48.94% (8,196/16,747) were male. A total of 31% (95% CI 29.36-32.69) of participants had pain, of whom 36.36% (1973/5426) had moderate pain. The top 3 pain locations were the waist (3232/16,747, 19.3%), legs (2476/16,747, 14.78%), and head (2250/16,747, 13.44%). There were differences in pain status among different groups by age, gender, education status, marital status, smoking, alcohol consumption, economic status, and location (Table 1). Female gender and regular drinking increased the risk of pain. Higher education, higher economic status, and living in an urban community reduced the risk of pain (Multimedia Appendix 1, Table S1).

https://publichealth.jmir.org/2024/1/e54309
Table 1. Sample characteristics of adults aged ≥45 years from China Health and Retirement Longitudinal Study by pain status.

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>Total (N=16,747)</th>
<th>No pain (n=11,298)</th>
<th>Pain (n=5449)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>59.57 (9.82)</td>
<td>59.29 (9.86)</td>
<td>60.17 (9.71)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Age groups (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-60</td>
<td>9545 (57)</td>
<td>6587 (58.3)</td>
<td>2958 (54.29)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>60-75</td>
<td>5789 (34.57)</td>
<td>3774 (33.4)</td>
<td>2015 (36.98)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≥75</td>
<td>1413 (8.44)</td>
<td>937 (8.29)</td>
<td>476 (8.74)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>8196 (48.94)</td>
<td>6042 (53.48)</td>
<td>2154 (39.53)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8551 (51.06)</td>
<td>5256 (46.52)</td>
<td>3295 (60.47)</td>
<td></td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Illiterate</td>
<td>4631 (27.65)</td>
<td>2750 (24.34)</td>
<td>1881 (34.52)</td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>6574 (39.25)</td>
<td>4239 (37.52)</td>
<td>2335 (42.85)</td>
<td></td>
</tr>
<tr>
<td>Secondary/high school</td>
<td>5146 (30.73)</td>
<td>3961 (35.06)</td>
<td>1185 (21.75)</td>
<td></td>
</tr>
<tr>
<td>University or above</td>
<td>396 (2.36)</td>
<td>348 (3.08)</td>
<td>48 (0.88)</td>
<td></td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never married</td>
<td>156 (0.93)</td>
<td>107 (0.95)</td>
<td>49 (0.9)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>14,567 (86.98)</td>
<td>9929 (87.88)</td>
<td>4638 (85.12)</td>
<td></td>
</tr>
<tr>
<td>Separated, divorced, or widowed</td>
<td>2024 (12.09)</td>
<td>1262 (11.17)</td>
<td>762 (13.98)</td>
<td></td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Nonsmokers</td>
<td>11,377 (67.93)</td>
<td>7458 (66.01)</td>
<td>3919 (71.92)</td>
<td></td>
</tr>
<tr>
<td>Current smokers</td>
<td>5370 (32.07)</td>
<td>3840 (33.99)</td>
<td>1530 (28.08)</td>
<td></td>
</tr>
<tr>
<td><strong>Drinking status</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>9779 (58.39)</td>
<td>6406 (56.7)</td>
<td>3373 (61.9)</td>
<td></td>
</tr>
<tr>
<td>Occasionally</td>
<td>1705 (10.18)</td>
<td>1187 (10.51)</td>
<td>518 (9.51)</td>
<td></td>
</tr>
<tr>
<td>Regularly</td>
<td>5263 (31.43)</td>
<td>3705 (32.79)</td>
<td>1558 (28.59)</td>
<td></td>
</tr>
<tr>
<td><strong>BMI (kg/m²)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>&lt;25</td>
<td>12,701 (75.84)</td>
<td>8613 (76.23)</td>
<td>4088 (75.02)</td>
<td></td>
</tr>
<tr>
<td>25-30</td>
<td>3372 (20.13)</td>
<td>2232 (19.76)</td>
<td>1140 (20.92)</td>
<td></td>
</tr>
<tr>
<td>≥30</td>
<td>674 (4.02)</td>
<td>453 (4.01)</td>
<td>221 (4.06)</td>
<td></td>
</tr>
<tr>
<td><strong>Economic status</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tertile 1</td>
<td>5583 (33.34)</td>
<td>3556 (31.47)</td>
<td>2027 (37.2)</td>
<td></td>
</tr>
<tr>
<td>Tertile 2</td>
<td>5583 (33.34)</td>
<td>3648 (32.29)</td>
<td>1935 (35.51)</td>
<td></td>
</tr>
<tr>
<td>Tertile 3</td>
<td>5581 (33.33)</td>
<td>4094 (36.24)</td>
<td>1487 (27.29)</td>
<td></td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Rural village</td>
<td>12,888 (76.96)</td>
<td>8291 (73.38)</td>
<td>4597 (84.36)</td>
<td></td>
</tr>
<tr>
<td>Urban community</td>
<td>3859 (23.04)</td>
<td>3007 (26.62)</td>
<td>852 (15.64)</td>
<td></td>
</tr>
<tr>
<td><strong>Severity of pain</strong></td>
<td></td>
<td></td>
<td></td>
<td>_a</td>
</tr>
<tr>
<td>Mild</td>
<td>1358 (25.03)</td>
<td>—</td>
<td>1358 (25.03)</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>1973 (36.36)</td>
<td>—</td>
<td>1973 (36.36)</td>
<td></td>
</tr>
<tr>
<td>Severe</td>
<td>2095 (38.61)</td>
<td>—</td>
<td>2095 (38.61)</td>
<td></td>
</tr>
<tr>
<td><strong>Location of pain</strong></td>
<td></td>
<td></td>
<td></td>
<td>_b</td>
</tr>
<tr>
<td>Head</td>
<td>2250 (13.44)</td>
<td>—</td>
<td>2250 (13.44)</td>
<td></td>
</tr>
</tbody>
</table>
Baseline characteristics & Total (N=16,747) & No pain (n=11,298) & Pain (n=5449) & \(P\) value

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=16,747)</th>
<th>No pain (n=11,298)</th>
<th>Pain (n=5449)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulders</td>
<td>2038 (12.17)</td>
<td>—</td>
<td>2038 (12.17)</td>
<td></td>
</tr>
<tr>
<td>Arms</td>
<td>1655 (9.88)</td>
<td>—</td>
<td>1655 (9.88)</td>
<td></td>
</tr>
<tr>
<td>Wrists</td>
<td>1057 (6.31)</td>
<td>—</td>
<td>1057 (6.31)</td>
<td></td>
</tr>
<tr>
<td>Fingers</td>
<td>990 (5.91)</td>
<td>—</td>
<td>990 (5.91)</td>
<td></td>
</tr>
<tr>
<td>Chest</td>
<td>1068 (6.38)</td>
<td>—</td>
<td>1068 (6.38)</td>
<td></td>
</tr>
<tr>
<td>Stomach</td>
<td>1436 (8.57)</td>
<td>—</td>
<td>1436 (8.57)</td>
<td></td>
</tr>
<tr>
<td>Back</td>
<td>1533 (9.15)</td>
<td>—</td>
<td>1533 (9.15)</td>
<td></td>
</tr>
<tr>
<td>Waist</td>
<td>3232 (19.3)</td>
<td>—</td>
<td>3232 (19.3)</td>
<td></td>
</tr>
<tr>
<td>Buttocks</td>
<td>633 (3.78)</td>
<td>—</td>
<td>633 (3.78)</td>
<td></td>
</tr>
<tr>
<td>Legs</td>
<td>2476 (14.78)</td>
<td>—</td>
<td>2476 (14.78)</td>
<td></td>
</tr>
<tr>
<td>Knees</td>
<td>2178 (13.01)</td>
<td>—</td>
<td>2178 (13.01)</td>
<td></td>
</tr>
<tr>
<td>Ankles</td>
<td>960 (5.73)</td>
<td>—</td>
<td>960 (5.73)</td>
<td></td>
</tr>
<tr>
<td>Toes</td>
<td>563 (3.36)</td>
<td>—</td>
<td>563 (3.36)</td>
<td></td>
</tr>
<tr>
<td>Neck</td>
<td>1135 (6.78)</td>
<td>—</td>
<td>1135 (6.78)</td>
<td></td>
</tr>
</tbody>
</table>

*Not applicable.*

**Concentration and Its Decomposition**

The concentration index of pain among Chinese adults was –0.066 (95% CI –0.078 to –0.054), which indicates that the prevalence of pain was higher among those with low economic status (Figure 1). Decomposition analysis found that old age, female gender, lower educational level, never being married or having another marriage status, living in a rural community, smoking, and having a lower economic status increased the inequality of pain and never or sometimes drinking and lower BMI decreased the inequality of pain. Among these factors, educational level (36.49%), location (36.87%), and economic status (25.05%) contributed significantly to the inequality of pain (Table 2). Further analysis showed that pain severity and the different parts of the body affected by pain (head, shoulders, arms, wrists, fingers, chest, stomach, back, waist, buttocks, legs, knees, ankles, toes, and neck) were concentrated in participants with low economic status, and the legs exhibited the most severe inequality (concentration index –0.121, 95% CI –0.141 to –0.101; Multimedia Appendix 1, Table S2).

**Figure 1.** Concentration curve of pain among Chinese adults aged ≥45 years from the China Health and Retirement Longitudinal Study, 2011-2012. The shadow around the curve represents the 95% CI.
Table 2. Decomposition analysis on the inequality of pain among Chinese adults aged ≥45 years.

<table>
<thead>
<tr>
<th>Variables and categories</th>
<th>Elasticity</th>
<th>Concentration index</th>
<th>Contribution</th>
<th>Contribution rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, (years)</td>
<td></td>
<td></td>
<td></td>
<td>0.9</td>
</tr>
<tr>
<td>60-75</td>
<td>0.077</td>
<td>−0.075</td>
<td>−0.006</td>
<td></td>
</tr>
<tr>
<td>≥75</td>
<td>−0.018</td>
<td>−0.123</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Female gender</td>
<td>0.890</td>
<td>−0.004</td>
<td>−0.003</td>
<td>0.9</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
<td></td>
<td>36.49</td>
</tr>
<tr>
<td>Illiteracy</td>
<td>0.803</td>
<td>−0.204</td>
<td>−0.164</td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>1.105</td>
<td>−0.055</td>
<td>−0.061</td>
<td></td>
</tr>
<tr>
<td>Secondary/high school</td>
<td>0.439</td>
<td>0.202</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>Marriage status</td>
<td></td>
<td></td>
<td></td>
<td>0.38</td>
</tr>
<tr>
<td>Never married</td>
<td>0.001</td>
<td>−0.178</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Separated, divorced, or widowed</td>
<td>0.037</td>
<td>−0.036</td>
<td>−0.001</td>
<td></td>
</tr>
<tr>
<td>Smoker</td>
<td>0.031</td>
<td>−0.016</td>
<td>−0.001</td>
<td>0.14</td>
</tr>
<tr>
<td>Drinking status</td>
<td>−0.256</td>
<td>−0.008</td>
<td>0.002</td>
<td>−0.42</td>
</tr>
<tr>
<td>Never</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occasionally</td>
<td>−0.026</td>
<td>0.019</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m^2)</td>
<td></td>
<td></td>
<td></td>
<td>−0.35</td>
</tr>
<tr>
<td>&lt;25</td>
<td>0.026</td>
<td>−0.014</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>25-30</td>
<td>0.045</td>
<td>0.037</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Economic status</td>
<td></td>
<td></td>
<td></td>
<td>25.05</td>
</tr>
<tr>
<td>Tertile 1</td>
<td>0.141</td>
<td>−0.667</td>
<td>−0.094</td>
<td></td>
</tr>
<tr>
<td>Tertile 2</td>
<td>0.162</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Location: rural village</td>
<td>1.073</td>
<td>−0.129</td>
<td>−0.138</td>
<td>36.87</td>
</tr>
</tbody>
</table>

Association of Pain With Risks of All-Cause Mortality

After 3 years of follow-up, 394 death events were reported, for a mortality rate of 2.35%. Participants with pain had a higher mortality rate than those without pain (log-rank test; \( P = .002 \)) (Figure 2). After adjusting for all covariates, Cox regression found that pain was associated with an increased risk of all-cause mortality (hazard ratio [HR] 1.30, 95% CI 1.06-1.61; Figure 3). Severity of pain also increased the risk of all-cause mortality (moderate: HR 1.36, 95% CI 1.02-1.83; severe: HR 1.42, 95% CI 1.07-1.87). The analysis focusing on specific types of pain showed that chest pain was significantly associated with all-cause mortality (Multimedia Appendix 1, Table S3). In addition, we performed stratified analyses according to sociodemographic characteristics, and the results for each group differed slightly from the overall results, but the different groups showed a trend that pain was associated with a higher risk of all-cause mortality across different groups (Figure 3).
**Discussion**

**Principal Results**

This study found that there was inequality in the prevalence of pain in nationally representative data from CHARLS. This inequality was concentrated in people with lower economic conditions. Education level, location, and economic status contributed significantly to this inequality. We further found that pain increased the risk of all-cause mortality. These findings outline the current inequitable status of pain and potential mortality risks in China.

**Comparison With Prior Work**

Previous studies have revealed correlations between pain and economic status. A meta-analysis found that compared to people with high SES, people with low and medium SES had an increased risk of pain [16]. We reached similar conclusions...
using the concentration index. In addition, studies on the association between SES and pain have been mainly concentrated in high-income countries, with fewer prospective studies conducted in low- and middle-income countries, and our results make up for this deficiency. In addition, we found that economic status, education level, and location were associated with greater pain inequality. All 3 variables are common indicators of SES [17-19]. Individuals with lower SES tend to report higher prevalence of pain [20,21]. Individuals with low SES often face barriers to accessing medical resources, including appropriate resources for pain management [19]. Limited access to medical resources may lead to delayed diagnosis and treatment of pain symptoms, exacerbating pain. Also, lower SES is associated with heavy physical work or a sedentary lifestyle [22,23]. These characteristics may contribute to the development of musculoskeletal pain. Chronic exposure to economic stress, including unemployment, housing instability, and food insecurity, could lead to chronic psychosocial stress. This stress is associated with the development of pain, involving the hypothalamic-pituitary-adrenal axis and neuroinflammatory status [16,24].

Pain is a sensory and emotional experience with great biological, psychological, and sociological complexity, which not only seriously affects the quality of life, but may also adversely affect longevity [1,25,26]. Song and Chung [27] used data from the Korean Longitudinal Study of Aging (2006-2016) and found there was a statistically significant relationship between pain and mortality risk, and this risk increased with the intensity of pain. Smith et al [28] also found this association in the English Longitudinal Study of Ageing. Our research led to similar conclusions. There have also been studies reporting an association between different pain sites and death. A systematic review and meta-analysis of cohort study data found that back pain was associated with increased all-cause mortality in women and people with more severe back pain [29]. Cleveland et al [30] found that knee pain was associated with increased mortality in a community-based cohort study. However, it is hard to say whether pain directly increases the risk of death. Chen et al [31], using data from the UK Biobank, found that at least half of the association between chronic musculoskeletal pain and increased all-cause mortality may be mediated by 4 factors: physical activity, smoking status, alcohol consumption, and opioid use. The effect of pain on all-cause death is a complex issue and more research is needed to further understand this association. Current evidence highlights the importance of pain management and effective pain management strategies, including comprehensive treatments and attention to the patient’s lifestyle and mental health, which may help reduce the adverse health outcomes associated with pain [32-36]. With the increase of aging in China, it is necessary to accurately understand pain and conduct policy interventions for pain.

Limitations
Our findings highlight the importance of equality-oriented policies for pain to reduce deaths associated with pain inequality. However, our study has some potential limitations. First, information about pain, its severity, and its location were self-reported, which could lead to recall bias. Second, the data for estimating pain equality in this study were from 2011, which may introduce bias relative to the current situation. Third, some studies have found that new-onset pain may significantly increase the risk of death [27], and our study did not consider the duration of pain, which led to insufficiently detailed results. Fourth, some factors, such as recent injuries, may influence the association between pain and all-cause mortality. However, CHARLS did not investigate this, potentially leading to an overestimation of the effect of pain on all-cause mortality. Finally, the study participants were aged 45 to 75 years, which limits our estimation of the equality of pain prevalence in people older than 75 years. Nevertheless, our study has important implications for highlighting the impact of pain on health.

Conclusion
The prevalence of pain in Chinese adults was concentrated among participants with low economic status, and pain increased the risk of all-cause death. Our results highlight the importance of socioeconomic factors in reducing deaths due to pain inequality by implementing targeted interventions to reduce pain inequality.

Acknowledgments
We extend our gratitude to all participants in the CHARLS study and the trial investigators. This article was crafted using research materials sourced from the Peking University Open Research Data Platform. ZZ contributed to conceptualization, formal analysis, and writing (original draft). DX contributed to formal analysis and writing (original draft). YB contributed to conceptualization, writing (review and editing), and supervision. All authors read and approved the final version of the manuscript. This study was funded by 3 research grants awarded to YB from the University of Macau (UM; MYRG2019-00044-ICMS, QRCM-IRG2022-001, and SKL-QRCM-IRG2023-032) and the UM Talent Program. This work also received funding from the Shaanxi Provincial Sports Bureau Regular Project (2023108), the Teacher Education Reform and Development Research Project of Xi’an Medical University (2023JFY-31), a school-level scientific research fund from Xi’an Medical University (2023BS21), and a Shaanxi provincial philosophy and social science research project (2024QN271). The sponsors did not play any role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Data Availability
All data sets are accessible through the China Health and Retirement Longitudinal Study database [37]. Researchers interested in using these data can register for an account on the website.
References


Abbreviations

CHARLS: China Health and Retirement Longitudinal Study
HR: hazard ratio
SES: socioeconomic status
A Multimorbidity Analysis of Hospitalized Patients With COVID-19 in Northwest Italy: Longitudinal Study Using Evolutionary Machine Learning and Health Administrative Data

Dayana Benny¹,², PhD; Mario Giacobini³, PhD; Alberto Catalano¹,⁴, PhD; Giuseppe Costa¹, MD; Roberto Gnavi⁵, MD; Fulvio Ricceri¹, PhD

¹Centre for Biostatistics, Epidemiology, and Public Health, Department of Clinical and Biological Sciences, University of Turin, Turin, Italy
²Modeling and Data Science, Department of Mathematics, University of Turin, Turin, Italy
³Data Analysis and Modeling Unit, Department of Veterinary Sciences, University of Turin, Turin, Italy
⁴Department of Translational Medicine, University of Piemonte Orientale, Novara, Italy
⁵Unit of Epidemiology, Regional Health Service, Local Health Unit Torino 3, Turin, Italy

Corresponding Author:
Dayana Benny, PhD
Centre for Biostatistics, Epidemiology, and Public Health
Department of Clinical and Biological Sciences
University of Turin
Regione Gonzole 10
Orbassano
Turin, 10043
Italy
Phone: 39 0116705440
Email: dayana.benny@unito.it

Abstract

Background: Multimorbidity is a significant public health concern, characterized by the coexistence and interaction of multiple preexisting medical conditions. This complex condition has been associated with an increased risk of COVID-19. Individuals with multimorbidity who contract COVID-19 often face a significant reduction in life expectancy. The postpandemic period has also highlighted an increase in frailty, emphasizing the importance of integrating existing multimorbidity details into epidemiological risk assessments. Managing clinical data that include medical histories presents significant challenges, particularly due to the sparsity of data arising from the rarity of multimorbidity conditions. Also, the complex enumeration of combinatorial multimorbidity features introduces challenges associated with combinatorial explosions.

Objective: This study aims to assess the severity of COVID-19 in individuals with multiple medical conditions, considering their demographic characteristics such as age and sex. We propose an evolutionary machine learning model designed to handle sparsity, analyzing preexisting multimorbidity profiles of patients hospitalized with COVID-19 based on their medical history. Our objective is to identify the optimal set of multimorbidity feature combinations strongly associated with COVID-19 severity. We also apply the Apriori algorithm to these evolutionarily derived predictive feature combinations to identify those with high support.

Methods: We used data from 3 administrative sources in Piedmont, Italy, involving 12,793 individuals aged 45-74 years who tested positive for COVID-19 between February and May 2020. From their 5-year pre–COVID-19 medical histories, we extracted multimorbidity features, including drug prescriptions, disease diagnoses, sex, and age. Focusing on COVID-19 hospitalization, we segmented the data into 4 cohorts based on age and sex. Addressing data imbalance through random resampling, we compared various machine learning algorithms to identify the optimal classification model for our evolutionary approach. Using 5-fold cross-validation, we evaluated each model’s performance. Our evolutionary algorithm, utilizing a deep learning classifier, generated prediction-based fitness scores to pinpoint multimorbidity combinations associated with COVID-19 hospitalization risk. Eventually, the Apriori algorithm was applied to identify frequent combinations with high support.

Results: We identified multimorbidity predictors associated with COVID-19 hospitalization, indicating more severe COVID-19 outcomes. Frequently occurring morbidity features in the final evolved combinations were age>53, R03BA (glucocorticoid inhalants), and N03AX (other antiepileptics) in cohort 1; A10BA (biguanide or metformin) and N02BE (anilides) in cohort 2;
N02AX (other opioids) and M04AA (preparations inhibiting uric acid production) in cohort 3; and G04CA (Alpha-adrenoreceptor antagonists) in cohort 4.

**Conclusions:** When combined with other multimorbidity features, even less prevalent medical conditions show associations with the outcome. This study provides insights beyond COVID-19, demonstrating how repurposed administrative data can be adapted and contribute to enhanced risk assessment for vulnerable populations.

(International Public Health Surveillance 2024;10:e52353) doi: 10.2196/52353

**KEYWORDS** machine learning; evolutionary algorithm; multimorbidity; data analysis; epidemiology; feature bins; COVID-19; long COVID; ICD; ATC; polypharmacy; sparse binary data; feature engineering; public health; severity; epidemiology; coronavirus; SARS-CoV-2; risk assessments; risk assessment; data; data mining; big data; longitudinal study; longitudinal analysis; longitudinal analyses; health data; Italy

**Introduction**

**Background**

COVID-19, classified as a highly infectious disease, poses a severe threat to vulnerable populations, making it a critical public health concern and a significant global epidemiological issue. The first Italian case of COVID-19 was diagnosed in the Lombardy region on February 21, 2020. The virus quickly spread across the country, leading to a nationwide lockdown and overwhelming the health care system. Italy was among the countries hardest hit by the COVID-19 pandemic, with Piedmont, a region in the northwest, experiencing a high number of cases during the first wave.

Multimorbidity refers to the presence of multiple coexisting medical conditions in a patient, which interact with each other, resulting in a complex and multidimensional health condition [1]. At the population level, it has been established that interactions between diseases can increase the severity of the overall medical condition and complicate the treatment of other diseases within the combination [2,3]. In people infected with SARS-CoV-2, multimorbidity can increase the severity of the infection [4,5]. Therefore, it is important to identify specific disease combinations that could impact the severity of COVID-19 in individuals with multimorbidity.

It is necessary to point out that having one or more of these chronic health conditions does not guarantee severe COVID-19 development, but it does increase the risk. Different diseases may affect COVID-19 outcomes differently. Therefore, identifying specific disease combinations and studying interactions between different chronic health conditions are essential for understanding the severity of COVID-19 among individuals with multimorbidity. This can help health care professionals identify those at the highest risk of severe complications and provide appropriate prevention, care, and treatment.

Studying multimorbidity using traditional methods can be labor-intensive, requiring the identification of high-dimensional combinatorial features. Also, there is no universally accepted list of medical conditions to define multimorbidity. To address these challenges, efforts must focus on identifying low-dimensional representations of multimorbidity features for effective outcome prediction. High-order input features make machine learning models more prone to overfitting, and identifying meaningful high-order combinatorial features requires extensive effort from experts with domain knowledge.

Contrary to misconceptions, the concept of combinatorial features requires extensive effort from experts with domain knowledge. Contrary to misconceptions, the concept of multimorbidity analysis in patients with COVID-19 is not outdated. Our study introduces a cutting-edge tool designed to analyze the complex interactions among diverse chronic health conditions and their collective impact, which could be valuable in situations similar to recent health crises.

Traditionally, research on multimorbidity has focused on counting the total number of chronic conditions rather than considering individual experiences and the effects of different combinations of diseases [6]. Count-based measures of multimorbidity have been utilized to predict emergency hospitalizations [7,8]. Common combinations of medical conditions have been documented to delineate patterns of multimorbidity [9,10]. Previous studies have explored multimorbidity combinations using methods such as latent class analysis [11], cluster analysis [12], network analysis [13], factor analysis [14], association rules, and tree-based analysis [13,15].

Rare features, such as diseases and drugs with low occurrence rates in the data, can pose significant challenges for both statistical and machine learning analyses. Their lower prevalence in the data can result in sparsity, which may lead to poorer predictions. Some studies using machine learning to investigate multimorbidity patterns tackled data sparsity by strategies such as removing sparsity-inducing features [16], consolidating feature categories after one-hot encoding [17], or clustering rare features [18]. However, while these methods can alleviate sparsity, they may also result in the loss of important information and impede the meaningful interpretation of multimorbidity features [19].

**Importance of the Proposed Method for Multimorbidity Research**

With the increasing prevalence of electronic health records and other large data sets, there is a rising demand for efficient and effective methods to analyze and comprehend multimorbidity. By utilizing machine learning algorithms and other advanced computational techniques, researchers can gain deeper insights into the underlying mechanisms and risk factors associated with multimorbidity. This understanding can significantly inform more effective prevention and treatment strategies.
An evolutionary algorithm coupled with deep learning–based feature scoring represents a powerful approach for analyzing multimorbidity data [20]. This method involves multiple steps aimed at identifying the most relevant features for predicting the target variable while minimizing the number of features used.

Initially, the data set undergoes preprocessing by using a feature binning approach to generate various subsets or bins of the multimorbidity features. This step aims to reduce sparsity in the data and facilitates more effective feature scoring. Subsequently, deep learning is utilized to score the features within each subset based on their relevance for predicting the target variable. The result of this process is a feature score assigned to each feature within every subset. Next, an evolutionary algorithm is applied to select the optimal subset of features based on their scores. The algorithm initiates by generating a population of candidate feature subsets and iteratively enhances this population through selection, crossover, and mutation operations [21]. The fitness of each candidate solution is assessed using a fitness function that integrates the deep learning–based feature scores derived from each subset or bin of features.

The output of the evolutionary algorithm is a subset of features that are most relevant for predicting the target variable. These features can be utilized for further analysis, such as constructing a predictive model or uncovering the underlying associations of multimorbidity patterns. In summary, using an evolutionary algorithm with deep learning–based feature scoring offers a robust approach to analyzing multimorbidity data, pinpointing the most influential features for predicting the target variable. This approach can result in enhanced model performance, faster training times, and improved interpretability in complex data sets featuring multimorbidity [22].

**Study Goal**

This study aims to identify multimorbidity patterns that may serve as predictors of COVID-19 hospitalization (as a proxy for a more severe COVID-19 outcome) using evolutionary algorithms. To assess the effectiveness of our approach, we conducted a comparative analysis to justify the use of deep learning as a classifier over other established machine learning algorithms in terms of prediction accuracy. The evolutionary model may excel not only because of its superior predictive performance but also because it effectively manages sparsity. This capability could result in better identification of key features, more stable predictions, or enhanced performance within specific subgroups of the data [23].

A logistic model might reveal multimorbidity patterns that exhibit complexity [24]. However, linear models, despite their high interpretability, may struggle in sparse data sets where effective feature selection is relevant [25]. In such cases, evolutionary algorithms, particularly genetic algorithms, excel by adeptly handling feature interactions with complexity and identifying optimal feature subsets, a task that proves challenging for linear models in sparse data scenarios [26].

The selected models are interpreted using Shapley Additive Explanations (SHAP) values [27] to understand the relationship between the multimorbidity features across different cohort data and hospitalization outcomes. The proposed method also generates a feature-engineered data set containing a specified number of outcome-associated combinations or bins of multimorbidity. Then, the best performing bins are analyzed to explore the frequency of various multimorbidity patterns across all cohorts.

This study demonstrates how our innovative tool has the potential to revolutionize traditional risk assessment approaches. By incorporating complex combinations of diseases, the tool aims to improve the accuracy of predicting severe outcomes for individuals with multiple chronic conditions. With its adaptable design, it ensures applicability even in evolving scenarios involving different communicable diseases, highlighting its ongoing relevance. This study focuses on investigating the complexities of disease interactions, demonstrating how our tool could reshape risk assessment in similar contexts.

**Methods**

**Study Design**

This retrospective cohort study is designed to exhaustively examine the presence or absence of various multimorbidities in patients over a 5-year period leading up to the onset of COVID-19. The core of our analysis is the longitudinal tracking of these multimorbidities, relevant for understanding their impact on subsequent health outcomes, particularly hospitalizations due to COVID-19. Central to the study’s design is its longitudinal nature, involving systematic analysis of patient data collected over a specified time frame to assess how individual and collective health conditions influence the risk and severity of COVID-19–related hospitalizations. The retrospective cohort framework enables the use of existing medical records, including hospital discharge summaries and drug prescription data, to construct a comprehensive picture of each patient’s health status in the years leading up to the pandemic. Through an analysis of these long-term health patterns, our aim is to understand how preexisting conditions influence the severity of COVID-19.

This study involves examining individuals’ multimorbidity history over a 5-year period, encompassing both pre– and post–COVID-19 diagnosis periods. It further investigates the relationship between multimorbidity history, COVID-19 positivity, and the subsequent severity of COVID-19. This design allows for observing participants over an extended time frame and evaluating outcomes not only before but also after the critical event of COVID-19 diagnosis.

**Multimorbidity Data Set**

Data for the multimorbidity analyses were collected from the Piedmont Longitudinal Study (PLS), a health administrative cohort comprising anonymized records linked at the individual level from various social, health, and administrative databases [5,19]. Since February 2020, the PLS has been augmented by the regional COVID-19 platform, which collects data on COVID-19 infections. From these databases, we utilized registers for (1) hospital discharges, (2) drug prescription data, and (3) COVID-19 hospitalizations of individuals diagnosed...
with SARS-CoV-2 infection for the first time between February 22, 2020, and May 31, 2020. We retrieved the 5-year medical history of patients positive for COVID-19 from these data sets. The extracted data comprises 12,793 individuals aged 45-74 years who tested positive for the first time for SARS-CoV-2 infection. Our study specifically focused on this age group to eliminate potential influences from both younger and older individuals on the results. Also, this approach helped to mitigate bias associated with patients residing in nursing homes.

As the study was integrated into the National Statistical Plan, no ethical approvals or permits were required, and the database used for the analyses contained only anonymized data. Further information regarding ethical considerations and data availability can be found in the “Ethical Considerations” section.

**Ethical Considerations**

This study is part of the PLS, a specific project within the Italian National Statistical Program, proposed by the National Statistical System (SISTAN), integrated into the National Statistical Program (Programma Statistico Nazionale [PSN]), an initiative endorsed by the National Institute of Statistics (Istituto Nazionale di Statistica [ISTAT]) in Italy. This project is annually approved by the Italian Parliament. Since 2003, a dedicated form (PIE-00001 “Monitoring of socio-economic differences in mortality and morbidity through longitudinal studies”) has been included in the PSN, currently effective for the 3-year period from 2020 to 2022 [28] and recently renewed for 2023-2025.

Ethical approval or permits from the ethics committee are not necessary for this research. Access to PLS data within the responsible institution does not require informed consent, as stipulated by the Presidential Decree published in the Official Gazette of the Italian Republic N. 122 of May 26, 2022, under the PSN [29].

Consequently, informed consent from an ethical committee is not required for this study. All analyses adhered to the principles of the World Medical Association’s Declaration of Helsinki, and to preserve privacy, the data used for analysis underwent deidentification.

**Construction of the Exposure’s Variables**

In this longitudinal cohort study, patients’ multimorbidity status over the past 5 years (2015-2019) was compared in relation to a specific outcome (hospitalization due to COVID-19). Multimorbidity was defined using records from hospital discharge and drug prescription registers. In the data sets for hospital discharges and drug prescriptions, multiple entries exist for each patient with COVID-19. The drug prescriptions data set comprises approximately 1 million records, while the hospital discharges data set includes around 19,000 entries. From the drug prescriptions data set, the Anatomical Therapeutic Chemical (ATC) classification system codes were used. All distinct ATC codes up to the 4th level (the first 5 digits of the ATC codes) were considered in this study. One-hot encoding was applied to convert categorical codes into separate feature columns with binary values (0 or 1) indicating the absence or presence of drugs in each patient’s prescription history. Similarly, from the hospital discharge data, the 9th International Classification of Diseases-Clinical Modification (ICD-9-CM) codes [30] (as diagnosis codes) were used, and one-hot encoding was applied. Following these transformations, only drug codes and diagnosis codes meeting the criterion “at least 100 patients with this code in the COVID-19-positive patients’ data” are retained. Consequently, 194 features were derived from drug codes (n=112) and diagnosis codes (n=82) as multimorbidity features from the entire data set, where the presence and absence of these features are denoted as 1 and 0, respectively. Also, 2 features—age and sex—are included, with sex coded as 1 for females and 0 for males. Subsequently, the preprocessed data were segmented into 4 data sets based on age and sex. The data set transformation steps are illustrated in Figure 1.

Subsets of various cohorts were obtained by considering the study population falling within the age criteria of “aged 45-59 years” and “aged 60-74 years.” This subdivision is made because individuals aged 60 and above are often categorized as part of the older population [5]. Median values within each age range are used as threshold values for discretizing the age feature in this study. This approach involves categorizing or binning based on median values within each specified age range. For example, to discretize the age feature into groups such as “45-59” and “60-74,” we used medians (53 for “45-59” and 68 for “60-74”) as thresholds.

In the data sets for the younger cohorts (cohorts 1 and 2), the age feature was converted into a binary variable, where 1 represents age≥53 and 0 represents age<53. The age values were derived from the 2020 COVID-19 data, and the age of 53 years was used as a threshold to divide the younger population into 2 subgroups (45-53 and 54-59 years). Similarly, the older population was divided into 2 subgroups (60-68 and 69-74 years), where the age feature was converted into a binary variable, with 1 indicating age>68 and 0 indicating age≤68. All 4 cohort data sets were treated as distinct binary classification problems. The input variables, comprising multimorbidity history and age, along with the outcome variable indicating whether a patient was hospitalized due to COVID-19, were represented as binary values.

In our study, multimorbidity features included the presence and absence of prescribed drugs and diagnosed diseases, as well as patient age and sex. However, due to the rarity of many medical conditions in the study population, the resulting data set became sparse when encoding absence as 0 values.
Data Imbalance Rectification

A significant challenge when working with clinical data is predicting rare events, which can result in an imbalance problem when the target variable has more observations in one class than in others. Therefore, it is beneficial to handle imbalanced raw data properly to prevent bias toward a particular class. All data sets used in this study exhibit imbalance, and resampling is recommended as a solution. To address this, randomly balanced samples were drawn from the unbalanced original data set to achieve class balance. Subsequently, a statistical hypothesis test, specifically the one-proportion $z$-test, was performed. This test compares the proportion of the sampled population with that of the raw data population, ensuring the representativeness.
of the randomly balanced sample data compared with the original cohort data set and mitigating potential biases.

The steps performed to obtain an unbiased balanced data set with significant features are as follows:

- Extract all minority and majority samples attributed to the outcome value from the original cohort data set.
- Randomly select samples belonging to the majority class so that they are equal in number to the minority class to achieve a balanced data set.
- Calculate the prevalence of each feature in the randomly selected samples and compare it with the prevalence in the original population.
- Perform a one-proportion $z$-test on all nonzero variables to assess whether the frequency distribution of a feature in the resampled data is representative of the same feature in the original cohort data set, using a significance level of .05.
- Evaluate the results of the one-proportion $z$-test, considering the test statistic and $P$ values, to determine the significance and eliminate nonsignificant features from the sampled data.

**Model Development**

**Machine Learning Algorithms**

To select the best model, we evaluated the performance of various supervised machine learning algorithms. Using labeled health records enables the application of supervised learning, specifically binary classification to classify a patient’s multimorbidity profile. Deep learning and other machine learning algorithms were applied to all cohort data sets, as depicted in Figure 2. Results were compared using a scoring grid with average cross-validated scores.

**Figure 2.** Selecting best ML model for each cohort data set: A streamlined process of selecting the optimal ML model for cohort data sets, using supervised algorithms for binary classification of multimorbidity profiles, with comparison based on a scoring grid featuring average cross-validated scores. AUC: area under the curve; ML: machine learning.
SHAP Analysis

SHAP values were used to elucidate the contribution of individual features in predicting hospitalization outcomes across all cohorts. These SHAP values for all features were plotted, with their positions on the y-axis indicating their impact on the model outcome. Beeswarm plots of SHAP values were used to explore the distribution of influence that each feature has on the model outcome, with features of greater importance positioned higher on the graph. Each data point for a feature corresponds to a single patient, with the position of the data point (SHAP value) on the x-axis indicating the effect of that feature on the model outcome for that specific patient. In the SHAP beeswarm plots, when multimorbidity is present (indicated by a feature value of 1 in red), a higher positive SHAP value suggests that this feature acts as a risk factor for hospitalization. Conversely, a more negative SHAP value in the presence of multimorbidity indicates that this feature acts as a protective factor against hospitalization risk for the patient. These findings are illustrated in Figure 3.

Figure 3. SHAP beeswarm plots illustrating the impact of all features on COVID-19 hospitalization for all 4 models. SHAP: Shapley Additive Explanations.
Deep Learning With Sparse Data

This study addresses a sparse health care data set that includes rare medical conditions and drugs, posing challenges for statistical and machine learning analyses due to their low prevalence [23]. To tackle this issue, the study utilizes sequential deep learning with the Adaptive Gradient Algorithm (AdaGrad), an optimization algorithm well-suited for handling sparse data [31]. AdaGrad’s adaptive scaling of the learning rate eliminates the need for manual tuning and enhances robustness compared with stochastic gradient descent. Also, the study uses early stopping functionality to improve the model’s performance.

In all deep learning models, dropout has been used as a regularization technique to mitigate overfitting during training [32]. Specifically, a dropout layer with a 20% dropout rate has been introduced after the first and second layers in the sequential model. Given the binary classification nature of the problem, the default loss function used is binary cross-entropy loss [33].

Feature Selection for Discovering the Optimal Set of Multimorbidity Features

Feature selection as a preprocessing method eliminates irrelevant and redundant information, aiding in dimensionality reduction [34]. There are 3 main methods of feature selection: filter-based, embedded, and wrapper-based methods. Filter-based methods typically generate models with reduced predictive performance compared with the other 2 methods. The embedded method performs an optimal feature subset search while constructing the model, whereas the wrapper method selects the best feature subset based on the classifier’s performance. In our study, we used a wrapper method that utilizes deep learning as the classifier algorithm and an evolutionary algorithm as the search strategy to generate feature subsets (bins). The best performing bin is determined using the area under the curve (AUC) metric and selected as the optimal subset of multimorbidity features highly associated with COVID-19 hospitalization.

Evolutionary Machine Learning

The use of evolutionary algorithms represents a promising approach for extracting a reduced set of meaningful and accurate rare associations, particularly beneficial for addressing challenges such as sparse data, epistatic associations with features, and high-dimensional representations of features. Evolutionary machine learning is a hybrid method that leverages evolutionary computation to overcome challenges encountered in various machine learning tasks [35]. Compared with traditional algorithms that rely on exhaustive search-based techniques, evolutionary algorithms offer a more robust solution.

Several key considerations arise when performing feature engineering with evolutionary algorithms: (1) a feature’s lack of prevalence does not necessarily imply irrelevance; it could still strongly influence the outcome; (2) addressing data sparsity poses a challenge for many machine learning methods, particularly concerning features with near-0 variance; and (3) evaluating combinations of features may yield greater predictive power than assessing isolated features alone, emphasizing the significance of exploring feature interactions.

We used a genetic algorithm to create an optimized feature matrix. Initially, features were randomly grouped into bins, each forming a feature matrix. These bins were then regrouped using a genetic algorithm and a wrapper-based method interacting with a classifier. The study adopted the elitism principle to preserve the best-performing bins as checkpoints. The final feature matrix represents the evolved engineering matrix after all iterations, designed to address issues of data sparsity and incorporate interactions among various multimorbidity features.

The proposed evolutionary approach in the study is an evolutionary algorithm–based wrapper method, illustrated in Figure 4. It is a modified version of an existing evolutionary algorithm known as the Relevant Association Rare-variant-bin Evolver [23]. The proposed method differs from the existing approach in several ways: it utilizes a prediction-based method with separate training and testing phases, incorporates a deep learning technique with an AdaGrad optimizer, and estimates the frequency of occurrence of specific features within the best performing feature combinations. Also, the scores produced by the deep learning model serve as fitness scores to assess the performance of multimorbidity combinations in each cycle.
**Figure 4.** Illustration of the evolutionary approach carried out in this study: The process begins by randomly initializing subsets of features. Through the execution of evolutionary computation, a final feature matrix is generated. Subsequently, frequently occurring combinations and features are identified.

**Frequent Multimorbidity Features**

The most prevalent multimorbidity combinations were identified to discern patterns among patients with COVID-19 using the Apriori algorithm. Applied to the final bins data set, which includes various multimorbidity feature combinations obtained from the evolutionary algorithm, the Apriori algorithm utilized the support measure to gauge the commonality of feature combinations across rows in the final bins. To focus on relevant feature combinations, only the most common multimorbidity patterns were analyzed. Frequent combinations of features were examined using a minimum support threshold ($s_{min}$) set at 0.5 to derive frequent itemsets.
Results

Characteristics of the COVID-19 Population

Table 1 summarizes the characteristics of the COVID-19 population, while Table 2 presents the distribution of hospitalized and nonhospitalized patients.

Table 1. Characteristics of the COVID-19 population.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age groups (years), n/N (%)</td>
<td></td>
</tr>
<tr>
<td>45-53&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4179/7324 (57.06)</td>
</tr>
<tr>
<td>54-59&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3145/7324 (42.94)</td>
</tr>
<tr>
<td>60-68&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3296/5469 (60.27)</td>
</tr>
<tr>
<td>69-74&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2173/5469 (39.73)</td>
</tr>
<tr>
<td>Female, n/N (%)</td>
<td></td>
</tr>
<tr>
<td>Younger age group</td>
<td>4477/7324 (61.13)</td>
</tr>
<tr>
<td>Older age group</td>
<td>2355/5479 (42.98)</td>
</tr>
<tr>
<td>Male, n/N (%)</td>
<td></td>
</tr>
<tr>
<td>Younger age group</td>
<td>2847/7324 (38.87)</td>
</tr>
<tr>
<td>Older age group</td>
<td>3114/5479 (56.84)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Younger age group</td>
<td>52.3 (4.18)</td>
</tr>
<tr>
<td>Older age group</td>
<td>67 (4.55)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Considered the younger age group.

<sup>b</sup>Considered the older age group.

Table 2. Distribution of hospitalized and nonhospitalized patients with COVID-19.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Hospitalized</th>
<th>Nonhospitalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group (years)</td>
<td>Male, n/N</td>
<td>Female, n/N</td>
</tr>
<tr>
<td>45-59</td>
<td>1101/1717</td>
<td>616/1717</td>
</tr>
<tr>
<td>45-53</td>
<td>522/825</td>
<td>303/825</td>
</tr>
<tr>
<td>54-59</td>
<td>579/892</td>
<td>313/892</td>
</tr>
<tr>
<td>60-74&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1974/2927</td>
<td>953/2927</td>
</tr>
<tr>
<td>60-68&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1073/1585</td>
<td>512/1585</td>
</tr>
<tr>
<td>69-74&lt;sup&gt;a&lt;/sup&gt;</td>
<td>901/1342</td>
<td>441/1342</td>
</tr>
</tbody>
</table>

<sup>a</sup>Considered the older age group.

One-Proportion z-Test Results

The one-proportion z-test was conducted on all features, and the results comparing randomly sampled data with the original cohort data sets are presented in Multimedia Appendix 1.

Performance of Machine Learning Models and Model Selection

Table 3 illustrates the performance evaluation of the deep learning model used across all 4 cohorts. The evaluation of other machine learning models is presented in Multimedia Appendix 2.

For each cohort, as depicted in Figure 5, 2 line plots were generated to validate the model’s effectiveness using cross-validation.
Table 3. Performance evaluation of the deep learning model.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>AUC score 5-fold CV&lt;sup&gt;a&lt;/sup&gt; (SD), %</th>
<th>Training AUC score (loss)</th>
<th>Test AUC score (loss)</th>
<th>Accuracy, %</th>
<th>Precision, %</th>
<th>Recall, %</th>
<th>F&lt;sub&gt;1&lt;/sub&gt;-score, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77 (1.87)</td>
<td>82% (0.28)</td>
<td>80% (0.29)</td>
<td>76</td>
<td>85</td>
<td>63</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td>68 (1.94)</td>
<td>71% (0.30)</td>
<td>67% (0.32)</td>
<td>62</td>
<td>62</td>
<td>61</td>
<td>62</td>
</tr>
<tr>
<td>3</td>
<td>67 (1.87)</td>
<td>74% (0.31)</td>
<td>69% (0.32)</td>
<td>67</td>
<td>70</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>61 (2.44)</td>
<td>65% (0.34)</td>
<td>62% (0.34)</td>
<td>63</td>
<td>62</td>
<td>68</td>
<td>65</td>
</tr>
</tbody>
</table>

<sup>a</sup>AUC: area under the curve.

<sup>b</sup>CV: coefficient of variation.

Figure 5. Model loss plot and AUC score over epochs—validation of model efficiency for each cohort through 2 line plots: The topmost plot depicts the binary cross-entropy loss for the epochs for the training and validation data sets, and the bottommost one presents the classification performance (AUC score) over epochs. AUC: area under the curve.
In cohort 1, it is evident that the model learns the problem efficiently and rapidly, achieving an AUC score of 82% on the training data set and 80% on the test data set. The close similarity between these scores suggests that the model is neither overfitting nor underfitting. The cross-entropy loss plot showed that the model has converged, with acceptable loss values observed on both data sets. The classification performance plot further indicated convergence. The model’s performance and convergence suggested that cross-entropy loss is suitable for effectively learning this neural network problem. In cohort 2, the model achieved performance scores of 71% on the training data set and 67% on the test data set, with reasonable loss values. The minimal difference between these scores indicated that the model learned the problem satisfactorily. In cohort 3, the model achieved a training score of 74% and a test AUC score of 69%. Observing that there was no significant improvement after 30 epochs, early stopping could be implemented during model training to prevent overfitting and stabilize the validation loss.

In cohort 4, although the loss plot appeared well-converged, the model showed slightly lower classification performance compared with the models in other cohorts.

**Influence of Individual Features on COVID-19 Hospitalization: Most Prevalent Multimorbidity Features in Evolved Bins**

The accuracy scores of the evolutionarily obtained final bins have been calculated. The highest accuracy was achieved for cohort 1 using the evolutionary approach to find outcome-associated best subsets of features, reaching 71.43% (95% CI 67.31-67.97) with 64 features. For cohort 2, the accuracy was 63% (95% CI 59.43-59.75) using 69 features. Cohort 3 achieved an accuracy of 62.38% (95% CI 59.84-60.09) with 53 features, while cohort 4 achieved an accuracy of 58% (95% CI 55.42-55.63) using 61 features. These results were then compared with the accuracy score of the deep learning model that utilized all features, as illustrated in Figure 6.

**Table 4** displays the multimorbidity features that occurred most frequently in the final bins data set across all cohorts, using a minimum support (smin) measure of 0.6. It includes the prevalence of these features in the sampled data set. Detailed statistics for all other features can be found in Multimedia Appendix 3.

**Figure 6.** Maximum classification accuracy achieved by a bin versus number of features in that bin using evolutionary approach (left side) and the accuracy score achieved exclusively by the deep learning model (right side) with all the available features in the cohort.

In cohort 1, frequently occurring multimorbidity features included age>53, R03BA (glucocorticoid inhalants), and N03AX (other antiepileptics). For cohort 2, A10BA (biguanide or metformin) and N02BE (anilides) were prevalent. Cohort 3 exhibited frequent occurrences of N02AX (other opioids) and M04AA (preparations inhibiting uric acid production), while G04CA (Alpha-adrenoreceptor antagonists) was notable in cohort 4.
Table 4. Frequently occurred morbidity features in the evolutionarily obtained final bins data set with support measure with corresponding P values, and the prevalence of the features in the sampled data set utilized for the predictive analysis.

<table>
<thead>
<tr>
<th>Description</th>
<th>P value</th>
<th>Support</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &gt;53</td>
<td>&lt;.001</td>
<td>0.84</td>
<td>41.15</td>
</tr>
<tr>
<td>Glucocorticoids R03BA</td>
<td>&lt;.001</td>
<td>0.85</td>
<td>15.5</td>
</tr>
<tr>
<td>Other antiepileptics N03AX</td>
<td>&lt;.001</td>
<td>0.82</td>
<td>5.6</td>
</tr>
<tr>
<td>Other antihistamines for systemic use R06AX</td>
<td>&lt;.001</td>
<td>0.79</td>
<td>6.74</td>
</tr>
<tr>
<td>Other antibacterials J01XX</td>
<td>&lt;.001</td>
<td>0.78</td>
<td>14.2</td>
</tr>
<tr>
<td>Sulfonamides, plain C03CA</td>
<td>&lt;.001</td>
<td>0.76</td>
<td>5.19</td>
</tr>
<tr>
<td>Other opioids N02AX</td>
<td>&lt;.001</td>
<td>0.74</td>
<td>6.9</td>
</tr>
<tr>
<td>Vitamin D and analogs A11CC</td>
<td>&lt;.001</td>
<td>0.73</td>
<td>23.05</td>
</tr>
<tr>
<td>Other nonorganic psychoses 298</td>
<td>.16</td>
<td>0.68</td>
<td>0.16</td>
</tr>
<tr>
<td>Other acute and subacute forms of ischemic heart disease 411</td>
<td>.32</td>
<td>0.62</td>
<td>0.08</td>
</tr>
<tr>
<td>Biguanides A10BA</td>
<td>&lt;.001</td>
<td>0.86</td>
<td>4.31</td>
</tr>
<tr>
<td>Anilides N02BE</td>
<td>&lt;.001</td>
<td>0.79</td>
<td>6.4</td>
</tr>
<tr>
<td>Nucleosides and nucleotides (excluding reverse transcriptase inhibitors) J05AB</td>
<td>&lt;.001</td>
<td>0.76</td>
<td>2.91</td>
</tr>
<tr>
<td>Sulfonamides, plain C03CA</td>
<td>&lt;.001</td>
<td>0.76</td>
<td>4.09</td>
</tr>
<tr>
<td>Preparations inhibiting uric acid production M04AA</td>
<td>&lt;.001</td>
<td>0.74</td>
<td>5.13</td>
</tr>
<tr>
<td>Angiotensin II receptor blockers, plain C09CA</td>
<td>&lt;.001</td>
<td>0.71</td>
<td>8.4</td>
</tr>
<tr>
<td>Alpha-adrenergic receptor antagonists C02CA</td>
<td>&lt;.001</td>
<td>0.65</td>
<td>3.22</td>
</tr>
<tr>
<td>Dihydropyridine derivatives C08CA</td>
<td>&lt;.001</td>
<td>0.65</td>
<td>7.4</td>
</tr>
<tr>
<td>Triazole and tetrazole derivatives J02AC</td>
<td>.03</td>
<td>0.64</td>
<td>6.18</td>
</tr>
<tr>
<td>Selective serotonin reuptake inhibitors N06AB</td>
<td>.03</td>
<td>0.63</td>
<td>8.58</td>
</tr>
<tr>
<td>Prostaglandin analogs S01EE</td>
<td>.07</td>
<td>0.62</td>
<td>0.68</td>
</tr>
<tr>
<td>Fatty acid derivatives N03AG</td>
<td>.08</td>
<td>0.61</td>
<td>2.5</td>
</tr>
<tr>
<td>Acetic acid derivatives and related substances M01AB</td>
<td>.001</td>
<td>0.6</td>
<td>18.21</td>
</tr>
<tr>
<td>Benzodiazepine derivatives N03AE</td>
<td>.17</td>
<td>0.6</td>
<td>1.54</td>
</tr>
<tr>
<td>Surgical or other procedures not carried out because of contraindications V64</td>
<td>&gt;.99</td>
<td>0.64</td>
<td>0.18</td>
</tr>
<tr>
<td>Other orthopedic aftercare V54</td>
<td>.26</td>
<td>0.64</td>
<td>0.32</td>
</tr>
<tr>
<td>Malignant neoplasm of the bladder I88</td>
<td>.32</td>
<td>0.63</td>
<td>0.18</td>
</tr>
<tr>
<td>Cohort, category, and features</td>
<td>Description</td>
<td>P value</td>
<td>Support</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>735</td>
<td>Acquired deformities of the toe</td>
<td>&gt;.99</td>
<td>0.6</td>
</tr>
<tr>
<td>454</td>
<td>Varicose veins of lower extremities</td>
<td>.83</td>
<td>0.6</td>
</tr>
<tr>
<td>820</td>
<td>Fractures of the neck of the femur</td>
<td>.32</td>
<td>0.6</td>
</tr>
</tbody>
</table>

### 3

**ATC#**

<table>
<thead>
<tr>
<th>ATC#</th>
<th>Description</th>
<th>Support</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>N02AX</td>
<td>Other opioids</td>
<td>.84</td>
<td>12.96</td>
</tr>
<tr>
<td>M04AA</td>
<td>Preparations inhibiting uric acid production</td>
<td>.82</td>
<td>8.5</td>
</tr>
<tr>
<td>C03EA</td>
<td>Low-ceiling diuretics and potassium-sparing agents</td>
<td>.76</td>
<td>5.35</td>
</tr>
<tr>
<td>A02BA</td>
<td>H2-receptor antagonists</td>
<td>.75</td>
<td>4.04</td>
</tr>
<tr>
<td>B01AB</td>
<td>Heparin group</td>
<td>.73</td>
<td>12.59</td>
</tr>
<tr>
<td>N03AX</td>
<td>Other antiepileptics</td>
<td>.7</td>
<td>11.7</td>
</tr>
<tr>
<td>N02AA</td>
<td>Natural opium alkaloids</td>
<td>.68</td>
<td>13.9</td>
</tr>
<tr>
<td>J05AB</td>
<td>Nucleosides and nucleotides (excluding reverse transcriptase inhibitors)</td>
<td>.65</td>
<td>5.77</td>
</tr>
<tr>
<td>A12AA</td>
<td>Calcium</td>
<td>.62</td>
<td>4.67</td>
</tr>
<tr>
<td>C07BB</td>
<td>Beta blocking agents, selective, and thiazides</td>
<td>.62</td>
<td>2.62</td>
</tr>
<tr>
<td>B03BB</td>
<td>Folic acid and derivatives</td>
<td>.61</td>
<td>9.23</td>
</tr>
<tr>
<td>R03AC</td>
<td>Selective beta-2-adrenoreceptor agonists</td>
<td>.6</td>
<td>7.19</td>
</tr>
</tbody>
</table>

### ICD#

<table>
<thead>
<tr>
<th>ICD#</th>
<th>Description</th>
<th>Support</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>295</td>
<td>Schizophrenic disorders</td>
<td>.68</td>
<td>0.73</td>
</tr>
<tr>
<td>813</td>
<td>Fractures of the radius and ulna</td>
<td>.68</td>
<td>0.84</td>
</tr>
</tbody>
</table>

### 4

**ATC#**

<table>
<thead>
<tr>
<th>ATC#</th>
<th>Description</th>
<th>Support</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G04CA</td>
<td>Alpha-adrenoreceptor antagonists</td>
<td>.8</td>
<td>25.75</td>
</tr>
<tr>
<td>J01CA</td>
<td>Penicillins with extended spectrum</td>
<td>.73</td>
<td>14.47</td>
</tr>
<tr>
<td>C09DA</td>
<td>Angiotensin II receptor blockers and diuretics</td>
<td>.66</td>
<td>13.11</td>
</tr>
<tr>
<td>C09AA</td>
<td>ACE inhibitors, plain</td>
<td>.66</td>
<td>26.32</td>
</tr>
<tr>
<td>B01AA</td>
<td>Vitamin K antagonists</td>
<td>.64</td>
<td>4.61</td>
</tr>
<tr>
<td>C03CA</td>
<td>Sulfonamides, plain</td>
<td>.62</td>
<td>16.49</td>
</tr>
</tbody>
</table>

### ICD#

<table>
<thead>
<tr>
<th>ICD#</th>
<th>Description</th>
<th>Support</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>995</td>
<td>Certain adverse effects not elsewhere classified</td>
<td>.61</td>
<td>0.44</td>
</tr>
</tbody>
</table>

\(^a\)Not available.

\(^b\)ATC: Anatomical Therapeutic Chemical.

\(^c\)ICD: 9th International Classification of Diseases.

The graph in Figure 7 illustrates the combinations derived from analyzing all 2-variable combinations with a minimum support (\(s_{min}\)) of 0.5. Detailed results for these combinations can be found in Multimedia Appendix 4.
Figure 7. Frequent outcome-associated multimorbidity feature combinations (2 variable combinations with smin=0.5) in each cohort.

We observed that certain multimorbidity features appear consistently across most outcome-associated bins. Additionally, some features are common and frequent across the final bins of various cohorts. Table 5 tabulates the features and combinations that frequently appeared in the final bins data set, using a support (s) threshold between 0.7 and 1.0. These findings are graphically presented in Figure 8.
Table 5. Frequently appeared features and combinations in the final bins data set when the support (s) is configured between 0.7 and 1.0.

<table>
<thead>
<tr>
<th>Support</th>
<th>Length of the combination</th>
<th>Frequent features</th>
<th>Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.85</td>
<td>1</td>
<td>ATC R03BA</td>
<td>1</td>
</tr>
<tr>
<td>0.84</td>
<td>1</td>
<td>Age&gt;53</td>
<td>1</td>
</tr>
<tr>
<td>0.82</td>
<td>1</td>
<td>ATC N03AX</td>
<td>1</td>
</tr>
<tr>
<td>0.79</td>
<td>1</td>
<td>ATC R06AX</td>
<td>1</td>
</tr>
<tr>
<td>0.78</td>
<td>1</td>
<td>ATC J01XX</td>
<td>1</td>
</tr>
<tr>
<td>0.76</td>
<td>1</td>
<td>ATC C03CA</td>
<td>1</td>
</tr>
<tr>
<td>0.74</td>
<td>1</td>
<td>ATC N02AX</td>
<td>1</td>
</tr>
<tr>
<td>0.74</td>
<td>2</td>
<td>Age&gt;53, ATC R03BA</td>
<td>1</td>
</tr>
<tr>
<td>0.73</td>
<td>1</td>
<td>ATC A11CC</td>
<td>1</td>
</tr>
<tr>
<td>0.72</td>
<td>2</td>
<td>ATC N03AX, ATC R03BA</td>
<td>1</td>
</tr>
<tr>
<td>0.72</td>
<td>2</td>
<td>Age&gt;53, ATC N03AX</td>
<td>1</td>
</tr>
<tr>
<td>0.86</td>
<td>1</td>
<td>ATC A10BA</td>
<td>2</td>
</tr>
<tr>
<td>0.79</td>
<td>1</td>
<td>ATC N02BE</td>
<td>2</td>
</tr>
<tr>
<td>0.76</td>
<td>1</td>
<td>ATC C03CA</td>
<td>2</td>
</tr>
<tr>
<td>0.76</td>
<td>1</td>
<td>ATC J05AB</td>
<td>2</td>
</tr>
<tr>
<td>0.74</td>
<td>1</td>
<td>ATC M04AA</td>
<td>2</td>
</tr>
<tr>
<td>0.71</td>
<td>1</td>
<td>ATC C09CA</td>
<td>2</td>
</tr>
<tr>
<td>0.84</td>
<td>1</td>
<td>ATC N02AX</td>
<td>3</td>
</tr>
<tr>
<td>0.82</td>
<td>1</td>
<td>ATC M04AA</td>
<td>3</td>
</tr>
<tr>
<td>0.76</td>
<td>1</td>
<td>ATC C03EA</td>
<td>3</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>ATC A02BA</td>
<td>3</td>
</tr>
<tr>
<td>0.73</td>
<td>1</td>
<td>ATC B01AB</td>
<td>3</td>
</tr>
<tr>
<td>0.71</td>
<td>2</td>
<td>ATC M04AA, ATC N02AX</td>
<td>3</td>
</tr>
<tr>
<td>0.7</td>
<td>1</td>
<td>ATC N03AX</td>
<td>3</td>
</tr>
<tr>
<td>0.8</td>
<td>1</td>
<td>ATC G04CA</td>
<td>4</td>
</tr>
<tr>
<td>0.73</td>
<td>1</td>
<td>ATC J01CA</td>
<td>4</td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

The primary findings of the study highlight prevalent multimorbidity patterns identified within the evolved data set. These patterns, characterized by specific ATC codes and ICD codes, show significant associations with hospitalization outcomes, particularly among distinct demographic groups. This analysis not only provides insights into COVID-19 but also suggests potential broader applications. Repurposing data originally collected for administrative purposes, this innovative approach shows promise for multimorbidity analysis in public health. It shows the adaptability and versatility of the methodology, capable of extracting valuable insights from existing data sets to inform effective public health strategies and interventions.

While our evolutionary machine learning model shows only marginal superiority compared with other prediction models, even slight improvements in predictive performance can hold significant value in real-world applications, particularly in critical fields such as health care where accuracy is mandatory. Moreover, we acknowledge that achieving the highest prediction performance may not be the sole objective of our study.

In the baseline method, variables are transformed into a binary format using one-hot encoding, which leads to the creation of a large, sparse matrix [26]. Evolutionary models typically excel in handling high-dimensional data compared with linear models, utilizing their enhanced ability to navigate and effectively utilize the search space [36]. Evolutionary approaches have drawbacks such as challenges in interpretability, computational efficiency, and a higher risk of overfitting [37]. However, despite the simplicity and clear interpretability of linear models, evolutionary models excel in managing complex, high-dimensional data and are proficient in handling feature interactions with complexity. This makes them particularly suitable for studies focused on detailed and complex aspects of multimorbidity patterns. Utilizing a novel evolutionary machine learning approach, we illustrate the ability to derive meaningful results even from rare events. Our model’s successful application in uncovering prevalent morbidity patterns linked to COVID-19 outcomes underscores its potential to yield valuable insights across diverse data sets, particularly where data sparsity poses challenges. While acknowledging its computational demands, we emphasize the model’s readiness and adaptability for analyzing complex medical data, highlighting its robustness as a powerful tool in medical research.

We identified prevalent morbidity patterns from the evolved data set, focusing on multimorbidity combinations or feature

Figure 8. Illustration of the features and combinations that frequently appeared in the final bins data set when configuring the support (s) between 0.7 and 1.0 as radar chart, with features presented in more than 1 cohort stacked.
subsets closely associated with the outcome. This research targets clinically significant patterns directly. Utilizing an evolutionary algorithm to identify these combinations ensures the analysis is grounded in a robust, data-driven process. Analyzing the frequency of these subsets provides a measure of their prevalence and significance in the studied population. This step helps validate the relevance of the identified combinations, ensuring that the observed patterns are not random but indicative of common trends in patient data. Focusing on the most prevalent combinations, the study aims to yield findings with practical implications for health care providers. These findings can inform clinical decision-making by helping practitioners identify patients at higher risk due to specific multimorbidity patterns, enabling them to tailor treatment approaches accordingly.

Multimorbidity features such as older age combined with specific ATC codes (N03AX and R03BA) were frequently observed in outcome-related bins, particularly among middle-aged females. Likewise, during the analysis of SHAP values in cohort 1, it was noted that the use of inhaled corticosteroid medication for asthma (R03BA) had a significantly positive impact on the likelihood of hospitalization. This observation aligns with findings from the Open SAFELY study, which identified asthma as a significant risk factor for mortality in patients with COVID-19. Specifically, it highlighted that individuals using inhaled corticosteroids face the highest risk in this context [38].

The ATC N03AX group encompasses various antiepileptic medications used in treating bipolar disorder, epilepsy, migraine, and sometimes schizophrenia. Individuals with severe mental illnesses have shown a slightly higher risk of severe clinical outcomes from COVID-19 compared with those without prior mental health conditions [39]. Also, there have been reports linking the use of antiepileptic medications with vitamin D deficiency [40]. In our study, the presence of A11CC (vitamin D and analogs) in the multimorbidity history makes middle-aged females more vulnerable to hospitalization. Conversely, for older-age females, the presence of this feature is associated with smaller SHAP values, indicating that its presence in their history is protective against hospitalization.

In a multimorbidity study of hospitalized patients with COVID-19 [41], the ATC group most closely associated with prolonged hospital stays is M04AA, which includes preparations inhibiting uric acid production. In our study, among older-age females, the combinations of M04AA and N03AX were notably frequent. M04AA also featured prominently in middle-aged males, while G04CA (alpha-adrenoreceptor antagonists), used for benign prostate hypertrophy, was notable among older-age males. Research indicates that male COVID-19 cohorts experience more unfavorable clinical outcomes compared with females [42,43]. Specifically, while patients with cancer are at an increased risk of SARS-CoV-2 infection, individuals undergoing androgen-deprivation therapy for prostate cancer appear to have some level of protection against the infection [43].

Strength and Limitations
Each row in the data set represents a comprehensive aggregation of each patient’s multimorbidity history over a 5-year period, including all relevant instances of diseases and conditions. This approach ensures a holistic view of each patient’s health status. To minimize subjectivity in the selection process, the criteria for including health records in the data set are consistent and objective. The aggregation process is governed by standardized criteria, uniformly applied across all patients. Also, aggregating multiple health records into a single patient instance helps mitigate bias that could arise from selectively choosing one entry over another.

In many clinical scenarios, understanding the implications of false positives and false negatives is a requisite beyond just disease probabilities. Although metrics such as Pietra and sBrier [44] and the average deviation about the probability threshold (ADAPT) index [45] are valuable, especially when patients seek to understand disease probabilities, we believe that traditional metrics such as AUC, accuracy, precision, recall, and $F_1$-score, along with a confusion matrix, offer a comprehensive evaluation of the prediction models in this study. The use of a confusion matrix as an evaluation tool enables us to customize model assessment to reflect different clinical priorities, which is particularly relevant when the prediction model informs treatment plans or risk assessments [46].

Evolutionary algorithms inherently favor the best performing choices available, despite their stochastic nature. These biases contribute to their improved performance. Each evolutionary cycle involves evaluating bin fitness and performing genetic operations to identify the best performing group of features. In this study, the evolutionary algorithm is used not only for feature selection in sparse data but also to indirectly assess epistatic associations between features in each evolutionary cycle. Multimorbidity features are grouped into bins and scored based on a deep learning classifier’s predictive ability for the outcome. The features within bins are regrouped iteratively after each evolutionary cycle.

Many studies using machine learning to investigate multimorbidity patterns focus on handling sparse data sets by either removing sparsity-generating features or merging feature categories to reduce sparsity. However, these methods often result in information loss and less precise interpretation of multimorbidity features [19]. Instead of relying solely on a sequential deep learning model, we aggregated all evolved bins to create a new data set. This allowed us to analyze the evolutionarily evolved bins and identify frequent multimorbidity features and combinations.

Analyzing all possible combinations of multimorbidity features in a data set can be computationally expensive, and many irrelevant combinations may not warrant further analysis. To address this, we applied an evolutionary algorithm to extract meaningful combinations, prioritizing even less prevalent features. Consequently, our focus shifted to investigating only the most common multimorbidity features found in the top bins.
Conclusions
When combined with other multimorbidity features, we identified associations with the outcome even for less prevalent medical conditions. Discovering hidden interconnections among different multimorbidity features opens new research pathways for studying multidimensional medical conditions in combination.

Using an innovative evolutionary machine learning approach, we identified prevalent morbidity patterns linked to hospitalization risk, especially among specific age and gender cohorts. Our findings highlight the adaptability of this methodology, demonstrating its ability to yield significant insights even in scenarios involving rare events. In addition to this, we repurposed administrative data for multimorbidity analysis, offering a novel path for public health research. This approach has the potential to influence future studies and interventions, encompassing areas such as polypharmacy and long COVID-19 research. By deepening our understanding of COVID-19 dynamics, this study emphasizes the broader utility of such methodologies in shaping effective public health strategies and interventions.

Data Availability
The data sets used in this study are not publicly available. As a result of ethical committee restrictions, raw data cannot be publicly or freely shared to ensure the privacy and protection of individual-level data. However, researchers may request access to aggregated data by contacting the corresponding author (DB) through a reasonable inquiry. The code utilized to derive the results in this study is accessible in Multimedia Appendix 5.

Conflicts of Interest
None declared.

Multimedia Appendix 1
One-proportion z-test results.
[PDF File (Adobe PDF File), 311 KB - publichealth_v10i1e52353_app1.pdf ]

Multimedia Appendix 2
Performance evaluation of other machine learning algorithms.
[PDF File (Adobe PDF File), 308 KB - publichealth_v10i1e52353_app2.pdf ]

Multimedia Appendix 3
Outcome association of the feature and the support.
[PDF File (Adobe PDF File), 261 KB - publichealth_v10i1e52353_app3.pdf ]

Multimedia Appendix 4
Most prevalent multimorbidity feature combinations in evolved bins.
[PDF File (Adobe PDF File), 206 KB - publichealth_v10i1e52353_app4.pdf ]

Multimedia Appendix 5
The code used to derive the study results.
[ZIP File (Zip Archive), 1220 KB - publichealth_v10i1e52353_app5.zip ]

References
28. Established Legislative Decree no. 322/1989 concerning National Statistical System organization the Sistan includes: the National Institute of Statistics (ISTAT); public bodies and statistical information bodies (INEA, ISFOL); the statistical offices of the State administrations and other public bodies, of the Government Offices of the Government, of the Regions and Autonomous Provinces, of the Provinces, of the Chambers of Commerce (CCIAA), of the Municipalities, single or associated, and the statistics offices of other public and private institutions that perform public interest functions. SISTAN. URL: https://www.sistan.it/index.php?id=422 [accessed 2023-06-06]

https://publichealth.jmir.org/2024/1/e52353
Abbreviations

AdaGrad: Adaptive Gradient Algorithm
ADAPT: average deviation about the probability threshold
ATC: Anatomical Therapeutic Chemical
AUC: area under the curve
ICD-9-CM: 9th International Classification of Diseases-Clinical Modification
ISTAT: Istituto Nazionale di Statistica
PLS: Piedmont Longitudinal Study
PSN: Programma Statistico Nazionale
SHAP: Shapley Additive Explanations
SISTAN: National Statistical System

Please cite as:
Benny D, Giacobini M, Catalano A, Costa G, Gnavi R, Ricceri F
A Multimorbidity Analysis of Hospitalized Patients With COVID-19 in Northwest Italy: Longitudinal Study Using Evolutionary Machine Learning and Health Administrative Data
JMIR Public Health Surveill 2024;10:e52353
URL: https://publichealth.jmir.org/2024/1/e52353
doi: 10.2196/52353
PMID: 39024001

©Dayana Benny, Mario Giacobini, Alberto Catalano, Giuseppe Costa, Roberto Gnavi, Fulvio Ricceri. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 18.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Spatiotemporal Pattern and Its Determinants for Newly Reported HIV/AIDS Among Older Adults in Eastern China From 2004 to 2021: Retrospective Analysis Study

Gang Huang1*, MPH; Wei Cheng2*, MPH; Yun Xu3, MPH; Jiezhe Yang2, MPH; Jun Jiang2, MPH; Xiaohong Pan2, MPH; Xin Zhou2, MPH; Jianmin Jiang2,3, PhD; Chengliang Chai2, MPH

1Health Science Center, Ningbo University, Ningbo, China
2Department of AIDS and STD Prevention and Control, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China
3Key Lab of Vaccine, Prevention and Control of Infectious Disease of Zhejiang Province, Hangzhou, China
*these authors contributed equally

Corresponding Author:
Chengliang Chai, MPH
Department of AIDS and STD Prevention and Control
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Road, Binjiang District
Hangzhou, 310051
China
Phone: 86 13777867711
Email: chlchai@cdc.zj.cn

Abstract

Background: In recent years, the number and proportion of newly reported HIV/AIDS cases among older adults have increased dramatically. However, research on the pattern of temporal and spatial changes in newly reported HIV/AIDS among older adults remains limited.

Objective: This study analyzed the spatial and temporal distribution of HIV/AIDS cases and its influencing factors among older adults in Eastern China from 2004 to 2021, with the goal of improving HIV/AIDS prevention and intervention.

Methods: We extracted data on newly reported HIV/AIDS cases between 2004 and 2021 from a case-reporting system and used a Joinpoint regression model and an age-period-cohort model to analyze the temporal trends in HIV/AIDS prevalence. Spatial autocorrelation and geographically weighted regression models were used for spatial aggregation and influence factor analysis.

Results: A total of 12,376 participants with HIV/AIDS were included in the study. The newly reported HIV infections among older adults increased from 0.13 cases per 100,000 people in 2004 to 7.00 cases per 100,000 people in 2021. The average annual percent change in newly reported HIV infections was 28.0% (95% CI –21.6% to 34.8%). The results of the age-period-cohort model showed that age, period, and cohort factors affected the newly reported HIV infections among older adults. The newly reported HIV/AIDS cases among men who have sex with men (MSM) had spatial clustering, and the hotspots were mainly concentrated in Hangzhou. The disposable income of urban residents, illiteracy rate among people aged 15 years or older, and number of hospital beds per 1000 residents showed a positive association with the newly reported HIV infections among older MSM in the Zhejiang province.

Conclusions: HIV/AIDS among older adults showed an increasing trend and was influenced by age, period, and cohort effects. Older MSM with HIV/AIDS showed regional clustering and was associated with factors such as the disposable income of urban residents, the illiteracy rate among people aged 15 years or older, and the number of hospital beds per 1000 people. Targeted prevention and control measures are needed to reduce HIV infection among those at higher risk.

(JMIR Public Health Surveill 2024;10:e51172) doi:10.2196/51172

KEYWORDS
HIV/AIDS; men who have sex with men; newly reported infections; older adults; spatiotemporal analysis
Introduction

The first case of AIDS from HIV was reported in 1981 [1]. Since then, over 84 million individuals have been infected with HIV with over 40 million reported deaths attributed to AIDS-related illnesses. According to the Joint United Nations Programme on HIV/AIDS, approximately 38.4 million individuals are living with HIV, with 1.5 million new infections in 2021 [2]. In China, the absolute number of new HIV diagnoses has increased annually since 2005, reaching nearly 131,671 in 2020. By the end of 2020, a total of approximately 1,053,000 cases had been reported in China [3].

Initially, the epidemic was thought to affect primarily individuals aged between 15 and 49 years; however, evidence suggests that the burden of HIV/AIDS is considerable among those aged 50 years or older [4]. Epidemiological studies in China have also reported an increasing trend in the number and proportion of HIV infections among older adults [5]. A study from the Henan Province reported that the proportion of newly reported cases aged 50 years or older increased gradually from 4.5% between 1995 and 2000 to 35.5% between 2016 and 2020 through sexual transmission [6]. In China, the sexual needs of older adults are often neglected by their partners and society, often leading to high-risk sexual behavior. Unprotected sexual contact is very common among older people, which in turn increases the risk of HIV infection [4].

The generation, transmission, and distribution of HIV/AIDS are closely associated with geospatial information. Spatiotemporal analysis is commonly used in HIV/AIDS research to assess long-term trends and geographic distribution patterns [7-9]. The Joinpoint regression model can estimate long-term trends in reported HIV/AIDS incidence or infections, whereas the age-period-cohort model can estimate the effects of age, period, and cohort on HIV/AIDS annual reported incidence or infections [7]. Spatial analysis provides insights into the spatial distribution of the HIV/AIDS epidemic and risk factors associated with the geographically weighted regression (GWR) model [8]. For example, using AIDS intensity ranking maps and spatial cluster analysis, Wang et al [9] found that AIDS incidence hotspots in China in 2008 and 2011 were mainly distributed in Yunnan, Guangxi, Guizhou, Chongqing, and Sichuan.

In terms of factors influencing the AIDS epidemic, economic development is strongly positively associated with the AIDS epidemic in China; the GWR results further indicate that the impact of health care and education on the AIDS epidemic varies across regions [9]. However, the long-term trends and spatial distribution of HIV/AIDS prevalence among older adults in Eastern China have not been adequately assessed. The design and evaluation of national HIV/AIDS planning often rely on comprehensive national data, which can mask local HIV/AIDS epidemics [10]. To better control the HIV/AIDS epidemic, a more comprehensive and detailed analysis of HIV/AIDS among older adults in Eastern China is needed.

Therefore, this study assessed the changing trend and spatial distribution of newly reported HIV infections among older adults in Eastern China and aimed to clarify the temporal trends, hotspot areas, and influencing factors of the HIV/AIDS epidemic. Our findings provide novel insights into the developmental trend of the HIV/AIDS epidemic among older adults in Eastern China, help optimize resource allocation, and facilitate HIV/AIDS prevention and treatment strategies.

Methods

Ethical Considerations

This study was approved by the ethics review committee of the Center for Disease Control and Prevention of Zhejiang Province (2018-032). All of the surveillance work was completed by the local Centers for Disease Control and Prevention (CDCs). We extracted data from the Zhejiang Province database of the National Data Information System for Comprehensive HIV/AIDS Control. We did not collect extra information or specimens for this study. Therefore, our study is based on data derived from daily work and was exempt from informed consent. During data analysis, we strictly followed the requirements of the ethical review committee of the Zhejiang Provincial CDC to protect the privacy of the participants, and all data used in this study were anonymous.

Study Area

The Zhejiang Province, located on the southern flank of the Yangtze River Delta along the southeast coast of China (latitude: N 27° 02'-31° 11’ and longitude: E 118° 01'-123° 10’; Multimedia Appendix 1), has a land area of 105,500 km² with approximately 65.77 million residents as of 2022 [11]. Currently, Zhejiang Province has 11 prefecture-level cities, 37 municipal districts, 20 county-level cities, and 33 counties.

Data Sources

Since 2004, China has used a direct network reporting system called the China Infectious Disease Reporting Information System. The database was established by the Chinese CDC. Patients aged 50 years or older with newly reported HIV infection between January 1, 2004, and December 31, 2021, were screened from the Zhejiang Province database of the National Data Information System for Comprehensive HIV/AIDS Control. There were no significant differences in the data information system during the study period. In this study, patients with HIV/AIDS were diagnosed according to the HIV/AIDS diagnostic principles and China’s National HIV/AIDS Testing Technical Specifications and Standards [12]. Data on sociodemographic characteristics, transmission routes, and addresses were collected by local CDC staff in face-to-face interviews using standardized forms. Vectorized county-level geographic maps of the Zhejiang Province were derived from the National Geographic Information Public Service Platform. The area codes were determined by the Disease Control Information System.

To explore the factors influencing the newly reported HIV/AIDS cases according to geographical distribution, we collected the following variables: disposable income of urban residents, gross domestic product (GDP) per county, proportion of the illiterate population, proportion of the unmarried population, area of residence per capita, car ownership rate, number of hospital
beds per 1000 people, number of doctors per 1000 people, urbanization rate, and proportion of the older population. Data were compiled using a Microsoft Excel (version 16.0; Microsoft Corporation) spreadsheet. Data on these influencing factors were obtained from the Statistical Yearbook of Zhejiang Province due to the presence of multicollinearity between the variables. The ordinary least squares (OLS) method was used to exclude redundant variables (variance inflation factor >7.5) and those that did not pass the significance test.

Statistical Analysis

Temporal Analysis

The temporal analysis used the crude rate of newly reported HIV/AIDS for each year. The new reporting rate is calculated by dividing the total number of incidents in the year by the number of people in the corresponding year and multiplied by 100,000.

The annual percent change (APC) and average APC (AAPC) of each component were estimated using a Joinpoint regression model. The infection of new reports ($y_i$) was set as the dependent variable, year was set as the independent variable ($x_i$), and sex and route of infection were set as subgroup variables [13]. Joinpoint regression analysis was performed using JPR software (version 4.9.1.0; Statistical Research and Applications Branch, National Cancer Institute).

The age-period-cohort model, a popular statistical tool for extracting information hidden in morbidities, has been used in long-term trend studies on social change, causes of disease, aging, demographic processes, and dynamic studies [7]. The effects of age, period, and cohort factors on the newly reported HIV infections were analyzed using an R language–based network analysis tool developed by the National Cancer Institute [14].

The model required consistent age, period, and cohort intervals; we divided the study period into four 5-year intervals, divided age into 7 groups, and calculated birth cohort by subtracting the period from age. The age, period, and cohort of the center were selected as controls [14].

Spatial Analysis

ArcGIS software (version 10.8; ESRI Inc) was used to map the spatial distribution of HIV/AIDS cases. The Global Moran I Index was applied to examine the clustering of HIV/AIDS among older adults at the county level in Zhejiang Province to determine whether the clustering elements within the region were statistically significant at $\alpha=.05$ test level [15].

Local spatial autocorrelation was used to analyze the correlation of HIV/AIDS case distribution at the county level. In the Local Indicators of Spatial Autocorrelation (LISA) map, we identified 4 clusters of spatial relationships for the analyzed variables. High-high and low-low clusters are those in which the high or low values of the study variable are surrounded by neighboring areas above or below the mean [16].

The OLS model requires the data to be independently distributed, and the results are global estimates of the parameters that do not reflect the pattern of change in the data with geographical location. The GWR model embeds the spatial location of the data into regression parameters and uses local weighted least squares to estimate the parameters, which is a local statistical model [17]. In the presence of spatial autocorrelation, the traditional OLS model is not applicable to data analysis. In the GWR model, the region-specific regression coefficient is no longer the same value estimated using global information but rather a variable coefficient that varies with geographical location. The $R^2$, adjusted $R^2$, and corrected Akaike information criterion values were used to compare the OLS and GWR models [18]. The GWR model is structured as follows:

$$
\hat{\mu}_{i} = \beta_{0} + \beta_{1}(x_{i}) + \beta_{2}(y_{i}) + \ldots + \beta_{k}(z_{i}) + \epsilon_{i}
$$

where $\mu$ is the coordinate of the geographic center of the $i$th sample space unit and $\beta_k(\mu, \nu)$ is the value of the continuous function $\beta_k(\mu, \nu)$ in the space of the $i$th sample unit.

Result

Demographic Characteristics

Between 2004 and 2021, a total of 12,376 new HIV/AIDS cases among older adults were reported in Zhejiang Province. The number of reported cases increased from 15 in 2004 to 1657 in 2021, and the number of reported infections increased from 0.13 per 100,000 persons in 2004 to 7.00 per 100,000 persons in 2021. In terms of geographical distribution, HIV/AIDS cases were reported in all counties and districts during the study period. In 2004, a total of 9 counties reported HIV/AIDS cases among older adults, while 86 newly reported HIV/AIDS cases were reported among older adults in 2021 (Figure 1). Of the 12,376 patients, 76.9% (9514/12,376) were male and 98.5% (12,196/12,376) of cases involved transmission through sexual contact (10,496/12,376, 84.8% heterosexual transmission, and 1700/12,376, 13.7% homosexual transmission; Table 1).
Figure 1. Distribution patterns of reported HIV/AIDS prevalence among older adults in the Zhejiang Province (2004-2021).
Table 1. Demographic characteristics of HIV/AIDS among older adults in Zhejiang Province (2004-2021).

<table>
<thead>
<tr>
<th>Year</th>
<th>Reported cases, n</th>
<th>Male, n</th>
<th>Heterosexual transmission</th>
<th>Homosexual transmission</th>
<th>Blood transmission</th>
<th>Other routes or unknown</th>
<th>Female, n</th>
<th>Heterosexual transmission</th>
<th>Blood transmission</th>
<th>Other routes or unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>15</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2005</td>
<td>11</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2006</td>
<td>39</td>
<td>28</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2007</td>
<td>51</td>
<td>31</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2008</td>
<td>86</td>
<td>54</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>15</td>
<td>2</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2009</td>
<td>164</td>
<td>107</td>
<td>15</td>
<td>1</td>
<td>5</td>
<td>32</td>
<td>1</td>
<td>3</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2010</td>
<td>264</td>
<td>165</td>
<td>30</td>
<td>9</td>
<td>3</td>
<td>52</td>
<td>2</td>
<td>3</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2011</td>
<td>318</td>
<td>196</td>
<td>48</td>
<td>2</td>
<td>4</td>
<td>62</td>
<td>3</td>
<td>3</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td>409</td>
<td>254</td>
<td>56</td>
<td>1</td>
<td>0</td>
<td>96</td>
<td>2</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td>561</td>
<td>364</td>
<td>85</td>
<td>1</td>
<td>3</td>
<td>107</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>729</td>
<td>470</td>
<td>107</td>
<td>0</td>
<td>2</td>
<td>149</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2015</td>
<td>838</td>
<td>538</td>
<td>133</td>
<td>1</td>
<td>0</td>
<td>166</td>
<td>0</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2016</td>
<td>1129</td>
<td>710</td>
<td>142</td>
<td>3</td>
<td>3</td>
<td>270</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2017</td>
<td>1454</td>
<td>965</td>
<td>167</td>
<td>2</td>
<td>9</td>
<td>307</td>
<td>0</td>
<td>4</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2018</td>
<td>1502</td>
<td>932</td>
<td>202</td>
<td>2</td>
<td>7</td>
<td>357</td>
<td>0</td>
<td>2</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2019</td>
<td>1685</td>
<td>1022</td>
<td>235</td>
<td>4</td>
<td>15</td>
<td>405</td>
<td>0</td>
<td>4</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2020</td>
<td>1464</td>
<td>866</td>
<td>215</td>
<td>0</td>
<td>10</td>
<td>367</td>
<td>1</td>
<td>5</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2021</td>
<td>1657</td>
<td>971</td>
<td>254</td>
<td>0</td>
<td>22</td>
<td>402</td>
<td>0</td>
<td>8</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Overall</td>
<td>12376</td>
<td>7686</td>
<td>1700</td>
<td>37</td>
<td>91</td>
<td>2810</td>
<td>16</td>
<td>36</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

**Temporal Trend and Joinpoint Regression Analysis**

The newly reported HIV infections in older male individuals, which peaked in 2017 and then fluctuated downward from 2018 to 2021, were significantly higher than those in older female individuals in Zhejiang Province from 2004 to 2021. Heterosexual and homosexual transmission were the predominant routes of HIV/AIDS transmission among older adults, with the number of heterosexual transmission cases growing rapidly over the study period, reaching its highest recorded level in 2019. In contrast, the number of men who have sex with men (MSM) and who are HIV positive grew slower over the study period (Figure 2).

The results of the Joinpoint regression model for the newly reported HIV infections showed that the optimal results of the model were all 2 nodes. There was a significant increase in newly reported HIV infections among older adults in Zhejiang Province from 2004 to 2021 (AAPC = 28.0%, 95% CI -21.6% to 34.8%; Table 2).

**Figure 2.** Trends in the reported HIV/AIDS infections among older adults by sex and transmission.
Table 2. Temporal trend analysis of newly reported HIV infections among older adults in Zhejiang Province (2004-2021).

<table>
<thead>
<tr>
<th>Variables and segment end points (lower-upper)</th>
<th>Annual percent change (95% CI)</th>
<th>Average annual percent change (95% CI)</th>
<th>Trend segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>28.0 (21.6 to 34.8)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2004-2010</td>
<td>61.1 (37.9 to 88.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2017</td>
<td>22.0 (17.2 to 27.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-2021</td>
<td>–1.3 (–5.8 to 3.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>27.6 (19.8 to 36.0)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2004-2010</td>
<td>62.0 (33.5 to 96.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2017</td>
<td>23.7 (18.0 to 29.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-2021</td>
<td>–5.8 (–10.8 to –0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>26.6 (18.5 to 35.2)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2004-2010</td>
<td>51.5 (24.5 to 84.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2018</td>
<td>20.1 (14.8 to 25.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018-2021</td>
<td>1.6 (–8.1 to 12.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Routes of transmission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual transmission</td>
<td>28.7 (20.8 to 35.2)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2004-2010</td>
<td>60.0 (34.8 to 89.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2017</td>
<td>23.0 (17.7 to 28.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-2021</td>
<td>–2.5 (–7.3 to 2.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homosexual transmission</td>
<td>42.4 (23.2 to 64.5)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2004-2010</td>
<td>111.3 (34.0 to 233.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2015</td>
<td>25.8 (12.7 to 40.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-2021</td>
<td>6.4 (2.7 to 10.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other routes or unknown</td>
<td>10.2 (–15.3 to 43.3)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2004-2010</td>
<td>29.9 (0.1 to 68.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2013</td>
<td>–42.0 (–88.0 to 179.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013-2021</td>
<td>23.9 (8.9 to 40.8)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From 2004 to 2017, the newly reported infections showed an increasing trend (2004-2010: APC=61.1%, 95% CI –37.9% to 88.2%; 2010-2017: APC=22.0%, 95% CI –17.2% to 27.1%). However, there was no significant difference in the trend in newly reported infections from 2017 to 2021 (2017-2021: APC=1.6%, 95% CI –8.1% to 12.4%). Similar changes were observed in older male and female individuals, with AAPCs of 27.6% (95% CI –19.8% to 36.0%) and 26.6% (95% CI –18.5% to 35.2%) for older male and female individuals, respectively. For the newly reported infections of different routes of transmission, both heterosexual and homosexual transmission showed an increasing trend, with AAPC rates of 28.7% (95% CI –20.8% to 35.2%) and 42.4% (95% CI –23.2% to 64.5%), respectively. However, the annual trend for the other transmission routes was not significant, with an AAPC of 10.2% (95% CI –15.3% to 43.3%).

**Age-Period-Cohort Model Analysis**

**Wald Chi-Square Test Results**

The net drift, all-age deviations, all-period deviations, all-period relative risk (RR), and all-cohort RR of the newly reported infections among older adults in Zhejiang Province from 2004 to 2021 were significant (all P<.001), indicating that changes in the newly reported infections were affected by age, period, and cohort factors (Table 3). As shown in Figure 3A, the net drift of the newly reported HIV infections was 31.501% (95% CI –25.510% to 73.780%). There was no significant change in local drift in the newly reported infections across age groups.
Table 3. Age-period-cohort analysis of the newly reported HIV infections among the older adults in Zhejiang Province.

<table>
<thead>
<tr>
<th></th>
<th>$\chi^2$ (df)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net drift=0</td>
<td>132.490 (1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>All age deviations=0</td>
<td>26.946 (5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>All period deviations=0</td>
<td>60.887 (2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>All cohort deviations=0</td>
<td>3.537 (8)</td>
<td>.90</td>
</tr>
<tr>
<td>All period RR$^a$=1</td>
<td>161.170 (3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>All cohort RR=1</td>
<td>234.936 (9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>All local drifts=net drift</td>
<td>3.485 (7)</td>
<td>.84</td>
</tr>
</tbody>
</table>

$^a$RR: relative risk.

Figure 3. Age-period-cohort effect of reported HIV/AIDS infections among older adults in Zhejiang Province (2004-2021). (A) Net and local drifts, (B) Age effect, (C) Period effect, (D) Cohort effect. RR: relative risk.

**Age Effect**

Figure 3B shows the newly reported HIV infections showed a monotonic increasing trend from 0.01 per 100,000 persons in the age group of 50-54 years to 16.71 per 100,000 persons in the age group of ≥80 years, and the rate of increase was significantly accelerated after the age of 70 years.

**Period Effect**

Using the 2007-2011 period group as a control group (RR=1), an overall upward trend was observed in the period effect of HIV/AIDS risk reported by older adults, with the RR increasing from 0.099 (95% CI –0.052 to 0.185) to 6.23 (95% CI –4.269 to 9.092; Figure 3C).
Cohort Effect
Using the 1940-1944 birth cohort as a control group (RR=1), the cohort effect on the risk of HIV/AIDS among older adults born before 1940 in Zhejiang Province did not change significantly. Since 1944, the birth cohort effect has increased monotonically, with the RR increasing from 4.134 (95% CI –2.999 to 5.700) in the 1945-1949 birth cohort to 1064.394 (95% CI –418.537 to 2706.892) in the 1965-1969 birth cohort (Figure 3D).

Spatial Analysis
Global Spatial Autocorrelation
Global spatial autocorrelation analysis of the newly reported infections for each year of the study period was performed and grouped according to different routes of infection (Multimedia Appendix 2). The results showed no significant differences in the global geographic autocorrelation of newly reported infections for the total population and the heterosexual transmission population. However, the global Moran’s I values for the newly reported rates of older MSM after 2012 were all greater than 0 (P<.05), indicating significant clustering of newly reported infections among older MSM in the Zhejiang Province. Therefore, further spatial clustering analyses are required in this population.

Local Spatial Autocorrelation
LISA cluster distribution maps were used to identify HIV/AIDS hotspots and outlier sites among older MSM in the Zhejiang Province (Figure 4). In 2012, the high-high cluster areas of newly reported HIV infections among older MSM were concentrated in the urban areas of Hangzhou, including Shangcheng, Xiacheng, Gongshu, and Jianggan districts, and there were no low-low cluster areas. In the following years, the high-high cluster areas were also concentrated in the main urban area of Hangzhou, while the low-low cluster areas were small and unstable, scattered in Chun’an and Jiande counties as well as Cangnan and Taishun areas.

Spatial Regression Analysis
In the GWR, the independent variables for the best model in 2020 include urban disposable income, the illiteracy rate of residents aged 15 years or older, and the number of hospital beds per 1000 people. Multimedia Appendix 3 shows that the GWR model is better than the OLS model and has no residual autocorrelation. Figure 5A-C shows the spatial distribution of the independent variables, including the disposable income of urban residents, the illiteracy rate of residents aged 15 years or older, and the number of hospital beds per 1000 people, respectively. Figure 5D-F shows the β coefficients determined using GWR regression, and Figure 5D indicates that the disposable income of urban residents in the Zhejiang Province is positively associated with the newly reported HIV infections among older MSM. The newly reported HIV infections were higher in cities with high urban disposable income in southeastern Zhejiang Province. In Zhejiang Province, the illiteracy rate of residents aged 15 years or older was positively associated with newly reported HIV infections in the population (Figure 5E). Figure 5F shows the newly reported HIV infections associated with the number of hospital beds per 1000 people; respectively.
an increase in reported HIV infections in cities with a high number of hospital beds per 1000 people in northwestern Zhejiang Province was observed.

Figure 5. Spatial distribution of independent variables and regression coefficients.

Discussion

As HIV/AIDS is initially prevalent among individuals aged between 15 and 49 years, there are fewer reports of HIV infections in older adults [4]. The number and proportion of HIV/AIDS cases among older adults have increased significantly in recent years due to aging and the development of the HIV/AIDS epidemic. Yuan et al [8] found that the proportion of older adult cases of all reported HIV/AIDS cases in Sichuan Province, China, increased from 4.1% in 2008 to 59.2% in 2019. In the central and western regions of China, 98.2% of older adults were infected through heterosexual transmission [8], while in Zhejiang Province, 13.7% were infected through...
homosexual transmission. This suggests a slightly different pattern of HIV/AIDS prevalence in the eastern coastal region than in central and western China. However, few studies have used spatial and temporal distributions to analyze the prevalence of HIV/AIDS among older adults in the eastern, developed coastal areas of China. This study summarized the spatial and temporal distribution of the pandemic and its influencing factors, providing a rationale for designing prevention programs and allocating resources more efficiently.

Overall, the Joinpoint regression results showed that the proportion of newly reported HIV/AIDS cases increased significantly from 2004 to 2021. This phenomenon could be illustrated by several factors. First, the Chinese government adopted a policy of expanding testing in recent years, which has significantly improved the detection rate of the virus. This indicates that the expanded testing policy was widely implemented, leading to more HIV/AIDS being detected [19]. Second, at the beginning of the pandemic, HIV/AIDS mainly affected young individuals [20]; therefore, most education catered to that age group, while older individuals were overlooked, which increases their risk of infection in that age group. Third, the main route of HIV/AIDS transmission among older adults was heterosexual transmission (10,496/12,376, 84.8%). Low-cost commercial establishments are a major source of infection in older adults [21]. Simultaneously, older male individuals tend not to use condoms when engaging in heterosexual contact, which also increases their risk of HIV/AIDS transmission [4]. Finally, transportation, communication, and sexual activity have changed over the past decade with advances in technology and economic growth [8]. Access to transportation and communication has enhanced the sex work industry; with economic development, the older population has the financial means to pay for these services. This change also poses significant challenges to traditional HIV/AIDS prevention and control strategies [8]. Male individuals have significantly higher rates of HIV/AIDS than female individuals. Possible reasons for this are the lack of physiological decline in those older than 50 years of age and the high level of physical condition and sexual needs of older male individuals [22]. In summary, the current prevention services are inadequate for addressing the HIV/AIDS epidemic among older adults, and further research and public health interventions are required. Along with expanding testing, sex education for older adults should be strengthened to reduce the incidence of high-risk sexual behavior.

Age, period, and cohort had significant effects on newly reported HIV infections among older adults in Zhejiang Province. Age is an important factor influencing HIV/AIDS. Our findings suggest that, within the same birth cohort, older age correlates with a higher risk of reporting HIV/AIDS, which is consistent with previous results [23]. Lu et al [24] also found that HIV incidence among male individuals in Zhejiang Province peaked between the ages of 20 and 35 years, then declined and peaked again between the ages of 60 and 70 years. In this study, homosexual and heterosexual sexual contact were the main modes of HIV/AIDS transmission among older adults. Contrary to the common belief that sexual activity among older adults decreases with age, older adults still engage in sexual activity [25]. One study found that most older patients found sex with an HIV-positive partner to be acceptable [26]. However, unprotected sexual contact is common among older adults due to lower level of education, which increases their risk of HIV infection [4].

The period effect of newly reported HIV infections among older adults in the Zhejiang Province from 2004 to 2021 was consistent with the change in the Joinpoint regression. The monotonic upward trend in the period effect of newly reported infections was associated with factors such as policies to expand testing and unsafe sex among older adults [8,19,21]. Period effects are an important factor influencing HIV/AIDS transmission among older adults and is linked to medical advances and policy development. As economies and societies advance, multidisciplinary approaches to curb growing HIV/AIDS-related problems must be considered. For example, the Chinese CDC launched a web-based AIDS information system and integrated response information management system in 2006 [27]. This has implemented health promotion strategies within social networks and strengthened sexual health services and partner services.

Cohort effects included the risk of morbidity or mortality in individuals in the same birth cohort with the same exposure to a disease risk factor [24]. In this study, the birth cohort effect on newly reported HIV infections among older adults in Zhejiang Province increased from 2004 to 2021. Cohorts born before 1950 had fewer reported infections, whereas those born after 1950 showed a gradual increase in reported infections with increasing birth year. This may be attributed to the fact that the cohort born after 1950 is in the sexually active age group. A previous study reported higher rates of sexual transmission among people born later in life, which may indicate that they are more sexually active or risky [23].

Spatial analysis revealed a geographic expansion of the newly reported HIV infections among older adults in Zhejiang Province. HIV-positive MSM showed a spatial correlation during the study period. The spatial autocorrelation results showed high clustering of HIV-positive MSM in Hangzhou City. Factors influencing the aggregation of HIV infections include the local HIV burden [28]. The high number of chronically infected individuals in Hangzhou has hindered HIV/AIDS control. Sexual transmission has become a major mode of HIV transmission in Zhejiang Province; however, homosexuality remains highly stigmatized and discriminated against in China [29]. As a result, Chinese MSM tend to hide their identity and engage in sexual behavior in other areas. Most MSM are concentrated in metropolitan areas, where self-identification and sexual partners are available. Hangzhou is the capital city of the Zhejiang Province, with a rapidly growing economy that has attracted migrants, including MSM [30]. Meanwhile, the provincial transmission networks found that Hangzhou plays a central role in the cross-regional transmission of HIV/AIDS among MSM in Zhejiang Province [31]. Studies have found that stigma can also lead to risky behaviors, such as lack of testing and avoidance of treatment. Various forms of stigma, such as substance abuse, depression, and traumatic stress, interact to increase the vulnerability of MSM and increase the risk of HIV infection [32]. The stigma
associated with MSM, coming out, accepting sexual partners, and being perceived as “older” in the gay community can lead to risky behaviors such as avoiding sex education [33]. This study provides clues to the epidemic clustering characteristics of older MSM in Zhejiang Province, which will facilitate the design and implementation of evidence-based interventions. Eliminating stigmatization and discrimination against MSM populations is critical to controlling epidemic aggregation and reducing HIV infection, as stigma and discrimination are barriers to MSM seeking HIV/AIDS services [32].

Overall, the GWR analysis showed that urban disposable income, the illiteracy rate of residents aged 15 years or older, and the number of hospital beds per 1000 people were associated with the reported infections of HIV-positive MSM among older adults in the Zhejiang Province. The economy is an important factor influencing the pattern of HIV-reported infections. People with a higher income and lower awareness of HIV/AIDS were more likely to be HIV-positive [18], and their high disposable income allowed them to have more sexual partnerships. Moreover, illiteracy was positively associated with the reported HIV infections among older MSM in Zhejiang Province. Higher educational attainment is associated with a lower risk of HIV infection; targeted educational programs and attitude changes among people living with HIV/AIDS can help advance voluntary counseling and testing [34]. This study demonstrates the need to provide health education for older adults to further promote sexual health in older adults. Finally, the number of hospital beds per 1000 people was strongly associated with the reported infections among older MSM in Zhejiang Province. Health care is an important determinant of HIV/AIDS prevention. Qin et al [35] found a significant positive association between the number of health care facilities and the number of reported cases of HIV/AIDS. Detection bias is a possible cause, and better health care resources are associated with higher detection rates; therefore, areas with abundant health care resources generally have more diagnosed cases. The provision of good health care is strongly associated with early diagnosis of HIV/AIDS, which can help control the spread of the epidemic.

This study has some limitations. First, since this was an ecological study, it was difficult to determine the causal relationships between the outcomes and variables. Second, the number of newly reported cases differs from the incidence, and the number of reported HIV/AIDS cases among older adults is influenced by the testing intensity, coverage, and efficiency. The number of HIV/AIDS tests decreased by 41% worldwide due to inadequate HIV/AIDS testing in health services as a result of the COVID-19 pandemic [36]. In Zhejiang Province, where the number of reported cases among older adults began to decline for the first time in 2019, future research must assess the impact of the COVID-19 pandemic on HIV/AIDS testing. Third, spatial autocorrelation is influenced by partition effects. In this study, we used county districts as the spatial units of analysis. Different spatial statistical values would have been obtained had the analysis been conducted at the municipal or street levels. Further studies should consider using smaller spatial units for analysis, which may provide more locational information and make the results more comprehensive. Fourth, the GWR model used indicators obtained from the local bureau of statistics, which did not have specific indicators for older people. However, we have attempted to use these indicators to reflect the overall level of development in the area. Despite these limitations, we identified an increasing trend of HIV/AIDS among older adults in the Zhejiang Province, along with spatial clustering in the epidemic distribution.

In summary, the overall number of newly reported HIV infections among older adults in Zhejiang Province has increased in recent years, and older adults have become a key population in the HIV/AIDS epidemic. The majority of HIV/AIDS among older adults is transmitted through heterosexual transmission, but cases among MSM are showing aggregation in some counties. The combination of economic growth and low cognition among older adults are positively correlated with reported HIV infections, so prevention and control strategies should be more inclined to focus on the lower literacy level of older populations in better-off areas. Finally, areas with better health care resources can increase detection; therefore, it is important to comprehensively evaluate testing systems in different regions and improve access to testing services.

Acknowledgments

We would like to thank the physicians and staff at the HIV/AIDS surveillance sites across Zhejiang Province for their dedication to completing the countless data forms that made this work possible.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors’ Contributions

CC and Jianmin Jiang conceived and designed the study. GH and WC analyzed the data. YX, JY, XZ, Jun Jiang, and XP were responsible for reagents, materials, and analytical tools. GH and WC drafted the paper. CC and Jianmin Jiang reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.
Multimedia Appendix 1
Location of the study area, Zhejiang Province, China.
[PDF File, 2060 KB - publichealth_v10i1e51172_app1.png]

Multimedia Appendix 2
[DOCX File, 21 KB - publichealth_v10i1e51172_app2.docx]

Multimedia Appendix 3
Results of the selected explanatory model in ordinary least squares and geographically weighted regression.
[DOCX File, 15 KB - publichealth_v10i1e51172_app3.docx]

References


23. Huang et al. JMIR PUBLIC HEALTH AND SURVEILLANCE


Abbreviations

AAPC: average annual percent change
APC: annual percent change
CDC: Center for Disease Control and Prevention
GDP: gross domestic product
GWR: geographically weighted regression
LISA: Local Indicators of Spatial Autocorrelation
MSM: men who have sex with men
OLS: ordinary least squares
RR: relative risk

Please cite as:
Spatiotemporal Pattern and Its Determinants for Newly Reported HIV/AIDS Among Older Adults in Eastern China From 2004 to 2021: Retrospective Analysis Study
JMIR Public Health Surveill 2024;10:e51172
URL: https://publichealth.jmir.org/2024/1/e51172
doi:10.2196/51172
PMID:38349727

©Gang Huang, Wei Cheng, Yun Xu, Jiezhe Yang, Jun Jiang, Xiaohong Pan, Xin Zhou, Jianmin Jiang, Chengliang Chai. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 13.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.

Lovely Jain1, MPH; Sreya Pradhan2, MPH, PhD; Arun Aggarwal1, MD; Bijaya Kumar Padhi1, MPH, PhD; Ramaiah Itumalla3, PhD; Mahalaqua Nazli Khatib4, MD; Shilpa Gaidhane5, MD; Quazi Syed Zahiruddin6, MD; Celso Augusto Guimarães Santos7, PhD; Khalid AL-Mugheedi8, PhD; Tahani Alrahbeni9, PhD; Neelima Kukreti10, PhD; Prakasini Satapathy11,12, MPH, PhD; Sarvesh Rustagi13, PhD; Petra Heidler14, PhD; Roy Rillera Marzo15, MD

1Department of Community Medicine and School of Public Health, Postgraduate Institute of Medical Education and Research, Chandigarh, India
2IPE Global House, Delhi, India
3School of Management, The Apollo University, Chittoor, India
4Division of Evidence Synthesis, Global Consortium of Public Health and Research, Datta Meghe Institute of Higher Education, Wardha, India
5One Health Centre, Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education, Wardha, India
6Global Health Academy, Division of Evidence Synthesis, School of Epidemiology and Public Health and Research, Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education and Research, Wardha, India
7Department of Civil and Environmental Engineering, Federal University of Paraíba, João Pessoa, Brazil
8Adult Health Nursing and Critical Care, Riyadh Elm University, Riyadh, Saudi Arabia
9Molecular Toxicology and Genetics, Riyadh Elm University, Riyadh, Saudi Arabia
10School of Pharmacy, Graphic Era Hill University, Dehradun, India
11Center for Global Health Research, Saveetha Medical College and Hospital, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India
12Medical Laboratories Techniques Department, AL-Mustaqbal University, Hillah, Babil, Iraq
13School of Applied and Life Sciences, Uttaranchal University, Dehradun, India
14Institute International Trade and Sustainable Economy, IMC Krems University of Applied Sciences, Krems, Austria
15Faculty of Humanities and Health Sciences, Curtin University, Miri Sarawak, Malaysia

Corresponding Author:
Bijaya Kumar Padhi, MPH, PhD
Department of Community Medicine and School of Public Health
Postgraduate Institute of Medical Education and Research
Sector 12
Chandigarh, 160036
India
Phone: 91 8763766300
Email: bkpadhi@gmail.com

Abstract

Background: Undernutrition among children younger than 5 years is a subtle indicator of a country’s health and economic status. Despite substantial macroeconomic progress in India, undernutrition remains a significant burden with geographical variations, compounded by poor access to water, sanitation, and hygiene services.

Objective: This study aimed to explore the spatial trends of child growth failure (CGF) indicators and their association with household sanitation practices in India.

Methods: We used data from the Indian Demographic and Health Surveys spanning 1998-2021. District-level CGF indicators (stunting, wasting, and underweight) were cross-referenced with sanitation and sociodemographic characteristics. Global Moran I and Local Indicator of Spatial Association were used to detect spatial clustering of the indicators. Spatial regression models were used to evaluate the significant determinants of CGF indicators.

Results: Our study showed a decreasing trend in stunting (44.9%-38.4%) and underweight (46.7%-35.7%) but an increasing prevalence of wasting (15.7%-21.0%) over 15 years. The positive values of Moran I between 1998 and 2021 indicate the presence
of spatial autocorrelation. Geographic clustering was consistently observed in the states of Madhya Pradesh, Jharkhand, Odisha, Uttar Pradesh, Chhattisgarh, West Bengal, Rajasthan, Bihar, and Gujarat. Improved sanitation facilities, a higher wealth index, and advanced maternal education status showed a significant association in reducing stunting. Relative risk maps identified hotspots of CGF health outcomes, which could be targeted for future interventions.

Conclusions: Despite numerous policies and programs, malnutrition remains a concern. Its multifaceted causes demand coordinated and sustained interventions that go above and beyond the usual. Identifying hotspot locations will aid in developing control methods for achieving objectives in target areas.

(JMIR Public Health Surveill 2024;10:e41567) doi:10.2196/41567

KEYWORDS
undernutrition; malnutrition; stunting; wasting; underweight; sanitation; WaSH; LISA; NFHS; DHS; spatial epidemiology; children; child; India; intervention

Introduction

Child malnutrition continues to pose a significant challenge in low- and middle-income countries. Around 60% of deaths among children younger than 5 years in this countries are attributable to malnutrition [1]. Much of this burden is seen in regions of South Asia, where 2 of every 5 stunted children globally are found [2]. An alarming irony lies in the fact that regions in South Asia such as India, Pakistan, and Bangladesh report child malnutrition rates that are higher than those in sub-Saharan Africa [3]. From the earliest Demographic and Health Surveys (DHS) in the 1990s to the most recent surveys in 2021, all countries have experienced a reduction in the prevalence of stunting. However, India and Pakistan recorded the lowest reductions—2.3 and 0.6 percentage points per year, respectively—compared with Bangladesh and Nepal, which recorded the largest reductions—2.9 and 4.1 percentage points per year, respectively. The Global Hunger Index, a comprehensive instrument used for measuring and monitoring hunger on a global, regional, and national scale, in 2022 ranked India 107th out of the 121 countries; additionally, India has recorded the highest rate of child wasting and has performed comparatively worse than its neighboring countries, with Pakistan scoring 99, Bangladesh scoring 84, Nepal scoring 81, and Sri Lanka scoring 64 [4]. Malnutrition is the leading cause of significant health and developmental issues among children globally, accounting for 3.5 million deaths and 35% of morbidities among children younger than 5 years [5]. For those who survive, malnutrition can cause lasting damage, increase the likelihood of child morbidity, contribute to poor cognitive development in childhood, result in short stature in adulthood, decrease adult productivity, increase the risk of perinatal and neonatal death for women, and heighten the risk for chronic disease if accompanied by overweight in later childhood [6]. Malnutrition bears noteworthy economic ramifications, as India experiences an annual loss of 4% of its gross domestic product due to malnutrition [7].

Childhood malnutrition is a complex issue. Merely increasing household income will not suffice to alleviate the childhood malnutrition if children lack food security, health care, education, water, and sanitation access [8]. According to a recent study, 3 key drivers of malnutrition include poor diet among children in their first years of life, dietary inadequacy among women before and during the pregnancy, and poor sanitation practices at both household and community levels [6,9]. A recent study based on large-scale national representative survey data of India found that the likelihood of stunting increased with decreased regional water availability, unimproved sanitation, and unsanitary toilet disposal methods. Wasting was also exacerbated by a decline in regional water supplies and inadequate sanitation [7]. Household water, sanitation, and hygiene facilities and practices were reported as key mediating factors in the reduction of stunting in the Democratic Republic of Congo [9]. Children with better toilet facilities had a 37% less chance of being wasted [10]. Conversely, poor sanitation facilities could put the child at risk for a variety of infections and health problems [11]. Despite these issues being primarily associated with inadequate feeding patterns [12], addressing them remains challenging, particularly in India, where malnutrition affects both well-nourished and malnourished individuals [13]. According to the India State-Level Disease Burden Initiative, child growth failure (CGF) is responsible for almost one-fifth of all under-5 mortality in India, with a prevalence rate for stunting ranging from 21.3% to 49%, for wasting from 6.3% to 19.3%, and for underweight from 16.5% to 42.2% in 2017, showing a widespread geographical variance [14]. This disparity among Indian states makes achieving the Sustainable Development Goal 2 more challenging [15].

India has several nutrition-specific (food fortification and deworming) and nutrition-sensitive (Public Distribution System, Midday Meal Scheme, and POSHAN Abhiyaan) policies and programs in place to address recognized causes of malnutrition and child development. However, these programs encounter a variety of implementation challenges [16] across states, including poor targeting, leakage [17], inadequate infrastructure [18], corruption, and delayed payments [19]. If these large-scale social safety programs aim to address the fundamental drivers of hunger, it is necessary to bridge these implementation gaps. The provision of safe water and basic sanitation facilities in India has increased, although there remain disparities as indicated by the data reported by Water.org. According to recent statistics, a notable proportion of the Indian populace, specifically 6%, are deprived of access to potable water, whereas a significant majority (54%) lack access to properly managed household sanitation amenities. The score of India in terms of the availability of clean water and sanitation also reduced from 88 (2019) to 83 (2020) [20,21]. To gain a more nuanced understanding of the regional distribution of these indicators in
India, a spatial-regional approach is needed. Additionally, to identify the determining factors holistically and particularly across spatial-regional contexts, this study used a spatial regression model to explain the determinants of the unequal distribution of stunting, wasting, and underweight. Pursuant to this recommendation, our research aims to present the spatiotemporal trends of CGF indicators (wasting, stunting, and underweight) among Indian children younger than 5 years from 1992 to 2020, as well as their correlation with household sanitation practices.

Methods

Overview
This study used data from Indian Demographic and Health Surveys, also known as the National Family Health Surveys (NFHS). The NFHS is a large-scale, multiround survey conducted in a representative sample of households across India. It used a stratified random sampling method at multiple stages, distinguishing between urban and rural areas as natural strata. The survey captures national and state-level data on fertility, infant and child mortality, the practice of family planning, maternal and child health, reproductive health, nutritional status, and use and quality of health and family planning services [22].

A total of 33,026, 51,555, 259,627, and 232,920 children younger than 5 years who participated in NFHS-2 (1998), NFHS-3 (2005), NFHS-4 (2015), and NFHS-5 (2021), respectively, were included in the analysis.

Ethical Considerations
The study is based on a secondary data set from an NFHS survey that contains no personally identifying information about the survey participants. These NFHS data sets can be downloaded from the DHS program website after registration [23]. The NFHS survey had received ethical clearance from the International Institute for Population Science’s ethical review board before the survey. Additionally, the study protocol was reviewed and approved by the institutional research ethics committee of the Post Graduate Institute of Medical Education and Research, Chandigarh, India (vide letter number: NK/7176/MPH/188, July 27, 2021).

Outcome Variables
The nutritional status of children served as the outcome variable in this study. Children with a height-for-age $z$ score of less than $-2$ SD from the reference population median are considered stunted. Those with a height-for-age $z$ score of less than $-3$ SD from the reference population median are classified as severely stunted [24]. Children with a weight-for-height $z$ score of less than $-2$ SD from the reference population median are considered thin (wasted) or acutely undernourished. Those with a weight-for-height $z$ score of less than $-3$ SD from the reference population median are classified as severely wasted [24]. Underweight children have a weight-for-age $z$ score of less than $-2$ SD from the reference population median. Children are considered severely underweight if their weight-for-age $z$ score is less than $-3$ SD from the median [24]. Height and age SD, weight and height SD, and weight and age SD data were provided for these indicators. Anthropometric $z$ scores were calculated based on the World Health Organization (WHO) growth standards of children, as guided by the DHS-7 Guide to DHS Statistics. Children with height-for-age, weight-for-age, or weight-for-height $z$ scores falling outside of the acceptable range were identified as having invalid data.

Exposure Variable
We explored various characteristics that influence the outcome variables, as reported by previous studies [25,26]. The independent variables, along with potential confounding factors, included household sanitation facilities and sources of drinking water. Personal characteristics of children (eg, age and gender), maternal-related factors (eg, mother’s education), and household-related factors (eg, wealth index and ethnicity [caste]) were among these influences. Water supply and sanitation facilities, as defined by the DHS, were divided into 2 categories: improved and unimproved. Households with improved drinking water sources included those with piped water to yard, plot, or dwelling; tube well or borehole; public tap or standpipe; rainwater; protected well; protected spring; boiled water; and reverse osmosis plant. Households with unimproved water sources included those using surface water (dam, river, lake, canal, stream, or pond and irrigation), unprotected springs, unprotected wells, tanker trucks, or carts with a small tank. Improved sanitation facilities comprised flush or pour-flush to piped sewer systems, septic tanks, pit latrines, twin pits, pit latrines with slabs, ventilated improved pit or biogas latrines, and composting toilets. Unimproved sanitation facilities included any type of unimproved toilet facility, shared facilities, and open defecation (no facility, bush, or field).

Data Extraction
Figure 1 illustrates the framework used in the data extraction and analysis for this study. The data set, presented in the Stata(.dta) file format, was extracted from the DHS survey data set files. The extracted data were cleaned for the variables of interest using Stata (StataCorp). Data from NFHS-2, NFHS-3, NFHS-4, and NFHS-5 were used for the trend analysis.
Figure 1. Framework deployed in data extraction and analysis in this study. NFHS, National Family Health Surveys; UT, union territory.

Data Analysis

Spatial Analysis

We used India’s 2011 Census shapefile, which includes 640 districts for spatial analysis using the district-level data from NFHS-2, NFHS-4, and NFHS-5 (Figure 2).

Figure 2. Regions of India.
**Exploratory Spatial Analysis**

The spatial autocorrelation statistic (Global Moran I) was used to assess the clustering or randomness of data, that is, whether the CGF patterns are randomly distributed, scattered, or clustered. If the Moran I value is closest to –1, the prevalence of wasting, stunting, and underweight is distributed; if the Moran I value is closest to +1, these variables are clustered in the study region. A Moran I value of zero, however, indicates a random distribution of CGF indications. Once it was confirmed that the distribution of stunting, wasting, and underweight is nonrandom, local Moran I was used to study the local-level cluster locations of CGF indication in India using the LISA (Local Indicator of Spatial Association) cluster map [27].

Hotspot clusters and cold spot clusters were identified by the local Moran I. Hotspots (High-High) and cold spots (Low-Low) were identified. It also accounted for outliers where high values were mostly surrounded by low values (High-Low) and outliers where the low values were primarily surrounded by high values (Low-High) [28]. The bivariate LISA was used to calculate the relationship between independent variables and CGF indicators.

The LISA significance map was used for hotspot analysis, computing the $P$ value for the estimation of the statistical significance of the clustering area. The $P$ value associated with 99%, 95%, and 90% confidence levels was used in this study to determine the presence of substantial clustering. The software used in this study included GeoDa 1.18 (Spatial Analysis Laboratory of the University of Illinois at Urbana-Champaign) [29] and QGIS 2.18.16 (QGIS Development Team) [30].

**Confirmatory Spatial Analysis**

The regression technique was used to investigate the relationship between the percentage of stunting, wasting, and underweight and background characteristics (improved and unimproved drinking water supply and toilet facilities, maternal education, ethnicity [caste], and wealth index). The choice of the regression technique was determined by the presence or absence of multicollinearity, homoscedasticity (constant variance), and normal distribution in the data. To check if the explanatory variables correspond, the multicollinearity condition number was used. Homoscedasticity was determined using the Breusch-Pagan test. The Jarque-Bera test was used to assess the normality of the error distribution. If the independent variables were found to be uncorrelated, followed a normal distribution, and had a constant variance (homoscedasticity), the ordinary least-squares regression, a nonspatial regression technique, would be used. If this assumption was broken, the use of spatial regression would be justified. Spatial regression is used for predicting the value of the dependent variable based on the collection of values from independent factors while considering spatial dependence. Spatial lag and spatial error are 2 types of spatial dependence. The spatial lag model assumes that the result of the outcome variable is affected by neighboring areas, whereas the spatial error model evaluates the impact of variables that were not included in the regression model but have an impact on the dependent variable [31]. The key difference between the 2 models is the inclusion of error terms of spatial dependence in the spatial error model. The analysis was conducted with the assistance of the RStudio [32] program.

**Results**

**Characteristics**

A total of 24,395, 45,377, 215,511, and 206,112 observations were available from the survey rounds 1998-1999, 2005-2006, 2015-2016, and 2019-2021, respectively. Table 1 describes the sociodemographic characteristics of the samples from NFHS-2, NFHS-3, NFHS-4, and NFHS-5 that were analyzed. Most of the sociodemographic distribution in the sample remained almost the same across NFHS-3, NFHS-4, and NFHS-5. However, in NFHS-2, there was higher representativeness of children from 0 to 24 months, particularly in middle wealth quintile; around 50% of mothers of children were uneducated in both NFHS-2 and NFHS-3, 30% in NFHS-4, and 21% in NFHS-5. There were 22%, 13%, 7.5%, and 4% unimproved sources of water in NFHS-2, NFHS-3, NFHS-4, and NFHS-5, respectively. The percentage of unimproved type of sanitation facility was 71% in NFHS-2, which reduced to 24% in NFHS-5. Table 2 indicates that the prevalence of stunting was 44.5% for NFHS-2, 48% for NFHS-3, 38.4% for NFHS-4, and 35.5% for NFHS-5. Similarly, for wasting, the prevalence was 16%, 20%, 21%, and 19%, respectively, and for underweight, the prevalence was 47%, 43%, 36%, and 32%, respectively. The observed prevalence is adjusted by survey weights according to the DHS guideline [33], and trends are shown at the regional level in Figures 3-5. The results were consistent with the prevalence reported in the DHS reports.
<table>
<thead>
<tr>
<th>Sociodemographic characteristic</th>
<th>NFHS-2 (1998-1999), n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>NFHS-3 (2005-2006), n (%)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>NFHS-4 (2015-2016), n (%)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>NFHS-5 (2019-2021), n (%)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (months)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-24</td>
<td>19,567 (69.9)</td>
<td>21,522 (41.1)</td>
<td>96,928 (41.1)</td>
<td>91,269 (41.2)</td>
</tr>
<tr>
<td>25-59</td>
<td>8402 (30.0)</td>
<td>30,763 (58.8)</td>
<td>138,783 (58.8)</td>
<td>129,994 (58.7)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16,996 (51.8)</td>
<td>29,415 (52.1)</td>
<td>130,572 (52.2)</td>
<td>119,960 (51.6)</td>
</tr>
<tr>
<td>Female</td>
<td>15,768 (48.1)</td>
<td>27,022 (47.8)</td>
<td>119,394 (47.7)</td>
<td>110,910 (48.0)</td>
</tr>
<tr>
<td><strong>Household wealth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>11,619 (35.2)</td>
<td>27,030 (47.8)</td>
<td>118,332 (47.3)</td>
<td>117,869 (50.6)</td>
</tr>
<tr>
<td>Middle</td>
<td>10,404 (31.5)</td>
<td>11,180 (19.8)</td>
<td>49,577 (19.8)</td>
<td>45,083 (19.4)</td>
</tr>
<tr>
<td>Rich</td>
<td>10,990 (33.3)</td>
<td>18,226 (32.3)</td>
<td>82,056 (32.8)</td>
<td>69,968 (30.4)</td>
</tr>
<tr>
<td><strong>Type of residence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>25,498 (77.8)</td>
<td>42,134 (74.6)</td>
<td>198,248 (76.3)</td>
<td>169,342 (73.4)</td>
</tr>
<tr>
<td>Urban</td>
<td>7266 (22.1)</td>
<td>14,303 (25.3)</td>
<td>61,379 (24.1)</td>
<td>61,527 (27.1)</td>
</tr>
<tr>
<td><strong>Caste or tribe</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheduled caste</td>
<td>6556 (20.2)</td>
<td>11,693 (21.2)</td>
<td>53,851 (22.3)</td>
<td>53,756 (24.4)</td>
</tr>
<tr>
<td>scheduled tribe</td>
<td>3121 (9.6)</td>
<td>5442 (9.8)</td>
<td>26,350 (10.9)</td>
<td>23,140 (10.4)</td>
</tr>
<tr>
<td>Other backward classes</td>
<td>10,514 (32.4)</td>
<td>22,716 (41.2)</td>
<td>110,399 (45.7)</td>
<td>100,407 (45.5)</td>
</tr>
<tr>
<td>Don’t know/other</td>
<td>12,189 (37.6)</td>
<td>15,286 (27.7)</td>
<td>50,687 (21.0)</td>
<td>43,256 (20.0)</td>
</tr>
<tr>
<td><strong>Mothers’ education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No education</td>
<td>17,786 (54.3)</td>
<td>28,237 (50.0)</td>
<td>75,140 (30.1)</td>
<td>49,305 (21.4)</td>
</tr>
<tr>
<td>Primary</td>
<td>5159 (15.7)</td>
<td>7919 (14.0)</td>
<td>35,119 (14.1)</td>
<td>28,434 (12.3)</td>
</tr>
<tr>
<td>Secondary</td>
<td>7230 (22.0)</td>
<td>17,463 (31.0)</td>
<td>113,519 (45.4)</td>
<td>117,031 (51.1)</td>
</tr>
<tr>
<td>Higher</td>
<td>2578 (8.0)</td>
<td>2816 (5.0)</td>
<td>26,187 (10.4)</td>
<td>36,098 (15.2)</td>
</tr>
<tr>
<td><strong>Source of drinking water</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved source</td>
<td>25,263 (77.1)</td>
<td>44,603 (86.6)</td>
<td>216,438 (92.4)</td>
<td>208,654 (90.4)</td>
</tr>
<tr>
<td>Unimproved source</td>
<td>7495 (22.8)</td>
<td>6884 (13.4)</td>
<td>17,755 (7.5)</td>
<td>8359 (4.0)</td>
</tr>
<tr>
<td><strong>Type of sanitation facility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved</td>
<td>9562 (29.2)</td>
<td>22,846 (40.4)</td>
<td>135,644 (54.3)</td>
<td>162,339 (70.3)</td>
</tr>
<tr>
<td>Unimproved (open defecation)</td>
<td>23,199 (70.8)</td>
<td>33,526 (59.4)</td>
<td>114,122 (45.6)</td>
<td>54,810 (23.7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>n (%): number of observations (proportion of observations).
Table 2. Prevalence of stunting, wasting, and underweight among children during different rounds of National Family and Health Survey (NFHS), India (N=577,128).

<table>
<thead>
<tr>
<th>Variable</th>
<th>NFHS-2 (1998-1999), n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>NFHS-3 (2005-2006), n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>NFHS-4 (2015-2016), n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>NFHS-5 (2019-2021), n (%)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total observations, n</td>
<td>33,026 (79.2)</td>
<td>51,555 (88.0)</td>
<td>259,627 (83.8)</td>
<td>232,920 (88.5)</td>
</tr>
<tr>
<td>Total student for height and weight</td>
<td>24,395 (73.8)</td>
<td>45,377 (80.8)</td>
<td>215,511 (83.0)</td>
<td>206,112 (88.5)</td>
</tr>
<tr>
<td><strong>Height for age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe stunted</td>
<td>5510 (22.5)</td>
<td>10,758 (23.7)</td>
<td>34,983 (16.2)</td>
<td>30,439 (15.1)</td>
</tr>
<tr>
<td>Stunted</td>
<td>10,958 (44.9)</td>
<td>21,798 (48.0)</td>
<td>82,698 (38.4)</td>
<td>71,566 (35.5)</td>
</tr>
<tr>
<td><strong>Weight for age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe underweight</td>
<td>4322 (17.7)</td>
<td>7185 (15.8)</td>
<td>24,232 (11.0)</td>
<td>21,885 (10.6)</td>
</tr>
<tr>
<td>Underweight</td>
<td>11,391 (46.7)</td>
<td>19,283 (42.5)</td>
<td>77,005 (35.7)</td>
<td>66,118 (32.1)</td>
</tr>
<tr>
<td><strong>Weight for height</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severely wasted</td>
<td>694 (2.8)</td>
<td>2916 (6.4)</td>
<td>16,012 (7.5)</td>
<td>15,144 (7.6)</td>
</tr>
<tr>
<td>Wasted</td>
<td>3860 (15.7)</td>
<td>8991 (19.8)</td>
<td>45,338 (21.0)</td>
<td>38,024 (19.2)</td>
</tr>
</tbody>
</table>

<sup>a</sup>n (%): number of observations (proportion of observations).

**Figure 3.** Trends of stunting by region; data from Indian Demographic and Health Surveys 1998, 2005, 2015, and 2021. NFHS: National Family Health Survey.
Trends of Stunting, Wasting, and Underweight
From 1998-1999 to 2005-2006, the prevalence of stunting increased in most of the regions, including Central, South, West, and Northeast, and started decreasing from 2005 onward (Figure 3). The prevalence of wasting increased from 1998 to 2015 and decreased in the subsequent 5 years (Figure 4). The prevalence of underweight decreased from NFHS-2 to NFHS-5 in all Indian regions (Figure 5).

Spatial Autocorrelation of Stunting, Wasting, and Underweight Among Children
During the survey rounds NFHS-2 to NFHS-5, the spatial patterns of stunting, wasting, and underweight among children were found to be nonrandom. For stunting, wasting, and underweight, the value of Global Moran I varied from 0.293 to 0.419, 0.253 to 0.033, and 0.404 to 0.584, respectively, indicating significant clustering of these indicators across the country (Table 3) in NFHS-2, NFHS-4, and NFHS-5. In NFHS-4, the Moran I value was the highest, indicating considerable positive spatial autocorrelation, that is, districts with comparable prevalence are clustered together. Similarly, the spatial patterns for the source of drinking water, sanitation, wealth index, maternal education, and caste were found to be nonrandom during the survey rounds NFHS-2 to NFHS-5. The values of Global Moran I varied from 0.344 to 0.352 for the drinking water source, 0.443 to 0.459 for sanitation facility, 0.473 to 0.417 for maternal education, 0.361 to 0.472 for ethnicity, and 0.524 to 0.637 for wealth index (Table 3). The clustering trends were highly significant (>90%) during this study period. Table 4 represents the bivariate Moran I value of the outcome and explanatory variables. The highest bivariate
Moran I (CGF vs explanatory variable) value was for stunting versus maternal education (−0.444), followed by stunting versus improved sanitation facility (−0.439), stunting versus wealth index (−0.426), and underweight versus improved sanitation facility (−0.487) during the NFHS-4 round.

**Table 3.** Univariate Moran I value for the outcome and explanatory variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate Moran I (P value), NFHS-2</th>
<th>Univariate Moran I (P value), NFHS-4</th>
<th>Univariate Moran I (P value), NFHS-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stunting</td>
<td>0.293 (&lt;.001)</td>
<td>0.555 (&lt;.001)</td>
<td>0.419 (&lt;.001)</td>
</tr>
<tr>
<td>Wasting</td>
<td>0.251 (&lt;.001)</td>
<td>0.401 (&lt;.001)</td>
<td>0.033 (.10)</td>
</tr>
<tr>
<td>Underweight</td>
<td>0.404 (&lt;.001)</td>
<td>0.659 (&lt;.001)</td>
<td>0.584 (&lt;.001)</td>
</tr>
<tr>
<td>Improved source of drinking water</td>
<td>0.344 (.22)</td>
<td>0.414 (&lt;.001)</td>
<td>0.352 (&lt;.001)</td>
</tr>
<tr>
<td>Improved sanitation facility</td>
<td>0.444 (.97)</td>
<td>0.621 (&lt;.001)</td>
<td>0.459 (&lt;.001)</td>
</tr>
<tr>
<td>Maternal education</td>
<td>0.459 (&lt;.001)</td>
<td>0.537 (&lt;.001)</td>
<td>0.417 (&lt;.001)</td>
</tr>
<tr>
<td>Caste</td>
<td>0.345 (&lt;.001)</td>
<td>0.438 (&lt;.001)</td>
<td>0.472 (&lt;.001)</td>
</tr>
<tr>
<td>Wealth index</td>
<td>0.502 (&lt;.001)</td>
<td>0.640 (&lt;.001)</td>
<td>0.637 (&lt;.001)</td>
</tr>
</tbody>
</table>

*NFHS: National Family Health Survey.

**Table 4.** Bivariate Moran I values for the outcome and explanatory variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Bivariate Moran I (P value), NFHS-2</th>
<th>Bivariate Moran I (P value), NFHS-4</th>
<th>Bivariate Moran I (P value), NFHS-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved source of drinking water</td>
<td>0.103 (.001)</td>
<td>0.051 (.01)</td>
<td>−0.016 (.01)</td>
</tr>
<tr>
<td>Improved sanitation facility</td>
<td>−0.246 (.001)</td>
<td>−0.439 (.01)</td>
<td>0.301 (.01)</td>
</tr>
<tr>
<td>Maternal education</td>
<td>0.028 (.002)</td>
<td>−0.444 (.01)</td>
<td>0.271 (.01)</td>
</tr>
<tr>
<td>Caste</td>
<td>0.000 (&lt;.001)</td>
<td>−0.151 (.01)</td>
<td>0.125 (.01)</td>
</tr>
<tr>
<td>Wealth index</td>
<td>−0.198 (.001)</td>
<td>−0.426 (.01)</td>
<td>0.346 (.01)</td>
</tr>
</tbody>
</table>

Wasting

| Improved source of drinking water | −0.059 (.001)                        | 0.055 (.001)                        | 0.003 (.36)                          |
| Improved sanitation facility    | −0.180 (.001)                       | −0.246 (.001)                       | 0.100 (.01)                          |
| Maternal education             | 0.065 (<.001)                       | −0.093 (.001)                       | 0.066 (.01)                          |
| Caste                         | −0.108 (<.001)                      | −0.137 (.001)                       | 0.039 (.01)                          |
| Wealth index                  | −0.083 (<.001)                      | −0.098 (.001)                       | 0.097 (.01)                          |

Underweight

| Improved source of drinking water | 0.103 (.001)                        | 0.084 (.001)                        | −0.009 (.100)                        |
| Improved sanitation facility    | −0.356 (.001)                       | −0.487 (.001)                       | 0.360 (.01)                          |
| Maternal education             | −0.259 (.001)                       | −0.368 (.001)                       | 0.252 (.01)                          |
| Caste                         | −0.043 (.001)                       | −0.205 (.001)                       | 0.144 (.01)                          |
| Wealth index                  | −0.169 (.001)                       | −0.367 (.001)                       | 0.331 (.01)                          |

*NFHS, National Family and Health Survey.

**Hotspot Detection**

The data associated with the shapefile, along with the spatial relationship for each district-level observation, are visualized using Moron’s scatter plot. Figure 6 represents the univariate LISA maps for stunting, wasting, and underweight. The hotspot areas (enumeration areas with a high risk of stunting) were identified in districts of Rajasthan, Madhya Pradesh, Jharkhand, Bihar, Uttar Pradesh, Gujarat, West Bengal, and Chhattisgarh. Conversely, Karnataka, Arunachal Pradesh, Assam, Kerala, and Tamil Nadu were identified as the cold spots (areas with a low risk for stunting) in all the 3 surveys (NFHS-2, NFHS-4, and NFHS-5). Meanwhile, Haryana, Punjab, Himachal Pradesh, Manipur, Mizoram, Tripura, Odisha, Andhra Pradesh, Nagaland, Uttarakhand, and Maharashtra were identified as the cold spot regions in the fourth and fifth surveys. The observed blank spots on the Indian map result from a lack of data for the corresponding districts. Similarly, the hotspots for wasting were identified...
found in the districts of Maharashtra, Madhya Pradesh, Bihar, Odisha, Jharkhand, Gujarat, Andhra Pradesh, Rajasthan, Karnataka, West Bengal, and Chhattisgarh. The cold spot exists in Uttar Pradesh, Haryana, Punjab, Assam, Arunachal Pradesh, Uttarakhand, Manipur, Kerala, Nagaland, Mizoram, and Himachal Pradesh. The hotspot areas for underweight were in Uttar Pradesh, Bihar, Maharashtra, Madhya Pradesh, Odisha, Jharkhand, Rajasthan, West Bengal, Andhra Pradesh, Gujarat, Karnataka, and Chhattisgarh. The cold spots were located in Kerala, Manipur, Himachal Pradesh, Punjab, Arunachal Pradesh, Tamil Nadu, Assam, Nagaland, Uttarakhand, Kerala, Mizoram, Haryana, Karnataka, and Punjab.

Figure 6. Univariate LISA (Local Indicator of Spatial Association) maps for the prevalence of (A) stunting, (B) wasting, and (C) underweight. NFHS: National Family Health Survey.

LISA Modeling

Figures 7-9 represent the bivariate LISA maps showing associations of stunting, wasting, and underweight with sanitation, respectively. For stunting and unimproved sanitation facilities, the hotspots are in Uttar Pradesh, Gujarat, Madhya Pradesh, Chhattisgarh, West Bengal, and Meghalaya. Regarding wasting and sanitation, Andhra Pradesh, Maharashtra, Madhya Pradesh, Gujarat, Rajasthan, West Bengal, and Chhattisgarh are the hotspot areas. Similarly, for underweight, Maharashtra, West Bengal, Jharkhand, Uttar Pradesh, Madhya Pradesh, Andhra Pradesh, Gujarat, Chhattisgarh, and Rajasthan are high-risk areas.

Hotspots for stunting and unimproved drinking water sources are in the districts of Rajasthan, Gujarat, Madhya Pradesh, Bihar, West Bengal, Uttar Pradesh, and Maharashtra. For wasting, hotspots are in Gujarat, Maharashtra, Rajasthan, Andhra Pradesh, Karnataka, Madhya Pradesh, Uttar Pradesh, West Bengal, Chhattisgarh, Bihar, and Odisha. The common hotspots for underweight and unimproved drinking water sources are the districts of Uttar Pradesh, Bihar, Chhattisgarh, Andhra Pradesh, Jharkhand, Maharashtra, Gujarat, Karnataka, Madhya Pradesh, Odisha, West Bengal, and Rajasthan. The hotspots for maternal education and CGF indicators are in Jharkhand, Gujarat, Uttar Pradesh, Chhattisgarh, Maharashtra, Madhya Pradesh, West Bengal, and Rajasthan for stunting. For wasting, these areas include Maharashtra, West Bengal, Karnataka, Chhattisgarh, Gujarat, Jharkhand, Odisha, Madhya Pradesh, and Rajasthan. The hotspots for underweight are in Gujarat, Maharashtra, Uttar Pradesh, Odisha, West Bengal, Chhattisgarh, Madhya Pradesh, Rajasthan, and Jharkhand. The hotspots for poor wealth index and CGF indicators are in Bihar, Uttar Pradesh, Gujarat, West Bengal, Madhya Pradesh, and Maharashtra for stunting. The hotspots for wasting are in Andhra Pradesh, Maharashtra, Madhya Pradesh, Rajasthan, Gujarat, and Karnataka. For underweight, Maharashtra, Rajasthan, Madhya Pradesh, and Chhattisgarh are consistently high-risk areas from 1998 to 2006. Uttar Pradesh and Gujarat are the hotspots added during the fourth and fifth rounds only. The hotspot areas for stunting and ethnic groups are Madhya Pradesh, Rajasthan, Jharkhand, West Bengal, Gujarat, Uttar Pradesh, Bihar, and Maharashtra. Similarly, for wasting, the hotspot areas are Uttar Pradesh, Gujarat, Mizoram, Rajasthan, Maharashtra, Kerala, Uttarakhand, Punjab, Andhra Pradesh, Jammu & Kashmir, Odisha, Assam, Tamil Nadu, Manipur, Meghalaya, Karnataka, Bihar, and West Bengal. The underweight hotspots are Uttar Pradesh, Rajasthan,
Bihar, Jharkhand, Gujarat, West Bengal, Andhra Pradesh, Karnataka, Odisha, Maharashtra, and Madhya Pradesh.

The map’s red areas (hotspots) are concerning areas for stunting, wasting, and underweight. Uttar Pradesh, Odisha, Bihar, Rajasthan, Chhattisgarh, Madhya Pradesh, West Bengal, Gujarat, and Jharkhand are some of the hotspot areas for wasting, stunting, and underweight and their background characteristics.

Figure 7. LISA (Local Indicator of Spatial Association) map showing the association between stunting and household sanitation facility between 1998 and 2021.

Figure 8. LISA (Local Indicator of Spatial Association) map showing the association between wasting and household sanitation facility between 1998 and 2021.

Figure 9. LISA (Local Indicator of Spatial Association) map showing the association between underweight and household sanitation facility between 1998 and 2021.
Spatial Regression

The test for multicollinearity suggests that there are unnecessary dependences between the explanatory variables in the data from NFHS-2 and NFHS-4. The Breusch-Pagan test indicates the presence of heteroskedasticity ($P<$0.001), and the Jarque-Bera test reveals a violation of the assumption of normal distribution ($P<0.001$). The rejection of these assumptions suggests that the data are spatially dependent, which necessitates the use of spatial regression. Compared with the lag model, the spatial error model demonstrated a lower Akaike information criterion score for NFHS-2, NFHS-4, and NFHS-5, indicating a better fit for the data from these surveys.

Stunting

Table 5 presents the regression coefficients for the association between stunting and sanitation facility, obtained from both ordinary least-squares (OLS) and the spatial error model. According to the models for NFHS-2, NFHS-4, and NFHS-5, with each unit increment in sanitation facility (%), wealth quintile (%), and mothers’ education (%), the prevalence of stunting is expected to decrease by 0.109-0.145, 0.049-0.111, and 0.058-0.274 units, respectively.

### Table 5. Regression coefficients for stunting obtained from the ordinary least-squares (OLS) and the spatial error model (SEM)$^a$.

<table>
<thead>
<tr>
<th>Year/Variable</th>
<th>OLS model</th>
<th>SEM</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>SE</td>
<td>$P$ value</td>
<td>Coefficient</td>
<td>SE</td>
<td>$P$ value</td>
</tr>
<tr>
<td><strong>1998-1999</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>65.778</td>
<td>2.447</td>
<td>.001</td>
<td>66.425</td>
<td>2.665</td>
</tr>
<tr>
<td>Water facility</td>
<td>0.002</td>
<td>0.025</td>
<td>.10</td>
<td>−0.002</td>
<td>0.026</td>
</tr>
<tr>
<td>Sanitation</td>
<td>−0.117</td>
<td>0.026</td>
<td>.001</td>
<td>−0.109</td>
<td>0.029</td>
</tr>
<tr>
<td>Wealth</td>
<td>−0.069</td>
<td>0.026</td>
<td>.01</td>
<td>−0.069</td>
<td>0.029</td>
</tr>
<tr>
<td>Education</td>
<td>−0.271</td>
<td>0.033</td>
<td>.001</td>
<td>−0.274</td>
<td>0.036</td>
</tr>
<tr>
<td>Caste</td>
<td>0.006</td>
<td>0.024</td>
<td>.10</td>
<td>−0.009</td>
<td>0.026</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.353</td>
<td>0.056</td>
</tr>
<tr>
<td><strong>2015-2016</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>51.114</td>
<td>2.378</td>
<td>.001</td>
<td>47.165</td>
<td>2.538</td>
</tr>
<tr>
<td>Water facility</td>
<td>0.101</td>
<td>0.022</td>
<td>.001</td>
<td>0.087</td>
<td>0.024</td>
</tr>
<tr>
<td>Sanitation</td>
<td>−0.110</td>
<td>0.017</td>
<td>.001</td>
<td>−0.145</td>
<td>0.021</td>
</tr>
<tr>
<td>Wealth</td>
<td>−0.071</td>
<td>0.018</td>
<td>.001</td>
<td>−0.049</td>
<td>0.022</td>
</tr>
<tr>
<td>Education</td>
<td>−0.178</td>
<td>0.020</td>
<td>.001</td>
<td>−0.101</td>
<td>0.023</td>
</tr>
<tr>
<td>Caste</td>
<td>−0.040</td>
<td>0.013</td>
<td>.01</td>
<td>−0.035</td>
<td>0.015</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.643</td>
<td>0.037</td>
</tr>
<tr>
<td><strong>2019-2021</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>45.742</td>
<td>3.629</td>
<td>.001</td>
<td>45.348</td>
<td>3.707</td>
</tr>
<tr>
<td>Water facility</td>
<td>0.132</td>
<td>0.033</td>
<td>.001</td>
<td>0.086</td>
<td>0.035</td>
</tr>
<tr>
<td>Sanitation</td>
<td>−0.138</td>
<td>0.023</td>
<td>.001</td>
<td>−0.117</td>
<td>0.025</td>
</tr>
<tr>
<td>Wealth</td>
<td>−0.103</td>
<td>0.014</td>
<td>.001</td>
<td>−0.111</td>
<td>0.017</td>
</tr>
<tr>
<td>Education</td>
<td>−0.099</td>
<td>0.020</td>
<td>.001</td>
<td>−0.058</td>
<td>0.021</td>
</tr>
<tr>
<td>Caste</td>
<td>−0.034</td>
<td>0.014</td>
<td>.02</td>
<td>−0.030</td>
<td>0.017</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.484</td>
<td>0.046</td>
</tr>
</tbody>
</table>

$^a$Akaike information criterion of spatial lag model (SLM)>SEM: National Family and Health Survey (NFHS)-2: 3160.3 (SLM), 3145.7 (SEM); NFHS-4: 3986.2 (SLM), 3950.1 (SEM); and NFHS-5: 4207.1 (SLM), 4207.1 (SEM).

$^b$P<0.05 denotes a significant impact on stunting prevalence.

$^c$Not available.
Wasting

Table 6 represents the regression coefficients for the association between wasting and sanitation facility, as obtained from OLS and the spatial error model. In all 3 models, a unit increase in sanitation facility (%) results in an expected decrease in the prevalence of wasting by 0.073 to 0.266 units.

Table 6. Regression coefficients for wasting obtained from the ordinary least-squares (OLS) and the spatial error model (SEM)a.

<table>
<thead>
<tr>
<th>Year/Variable</th>
<th>OLS model</th>
<th>SEM</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
<td>P value</td>
</tr>
<tr>
<td>1998-1999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>65.778</td>
<td>2.447</td>
<td>.001</td>
</tr>
<tr>
<td>Water facility</td>
<td>0.002</td>
<td>0.025</td>
<td>.100</td>
</tr>
<tr>
<td>Sanitation</td>
<td>-0.117</td>
<td>0.026</td>
<td>.001</td>
</tr>
<tr>
<td>Wealth</td>
<td>-0.069</td>
<td>0.026</td>
<td>.01</td>
</tr>
<tr>
<td>Education</td>
<td>-0.271</td>
<td>0.033</td>
<td>.001</td>
</tr>
<tr>
<td>Caste</td>
<td>0.006</td>
<td>0.024</td>
<td>.10</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2015-2016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water facility</td>
<td>-0.020</td>
<td>0.024</td>
<td>.10</td>
</tr>
<tr>
<td>Sanitation</td>
<td>-0.138</td>
<td>0.018</td>
<td>.001</td>
</tr>
<tr>
<td>Wealth</td>
<td>0.050</td>
<td>0.019</td>
<td>.001</td>
</tr>
<tr>
<td>Education</td>
<td>-0.002</td>
<td>0.022</td>
<td>.10</td>
</tr>
<tr>
<td>Caste</td>
<td>-0.026</td>
<td>0.014</td>
<td>.05</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2019-2021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>25.631</td>
<td>11.391</td>
<td>.02^b</td>
</tr>
<tr>
<td>Water facility</td>
<td>0.079</td>
<td>0.105</td>
<td>.45</td>
</tr>
<tr>
<td>Sanitation</td>
<td>-0.266</td>
<td>0.074</td>
<td>&lt;.001^b</td>
</tr>
<tr>
<td>Wealth</td>
<td>-0.051</td>
<td>0.045</td>
<td>.26</td>
</tr>
<tr>
<td>Education</td>
<td>0.099</td>
<td>0.065</td>
<td>.13</td>
</tr>
<tr>
<td>Caste</td>
<td>0.019</td>
<td>0.046</td>
<td>.67</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

^aAkaike information criterion of spatial lag model (SLM)>SEM: National Family and Health Survey (NFHS)-2: 2979 (SLM), 2978 (SEM); NFHS-4: 4029 (SLM), 4021 (SEM); and NFHS-5: 5661.7 (SLM), 5661.6 (SEM).

^bP<.05 denotes a significant impact on stunting prevalence.

^cNot available.

Underweight

Table 7 represents the regression coefficients for the association between underweight and sanitation facility, as obtained from both OLS and the spatial error model. For each unit increase in sanitation facility (%), the decrease of underweight is expected to decrease by 0.124 to 0.157 units.
Table 7. Regression coefficients for underweight obtained from the ordinary least-squares (OLS) and the spatial error model (SEM).^a

<table>
<thead>
<tr>
<th>Year/Variable</th>
<th>OLS model</th>
<th>SEM</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
<td>P value</td>
<td>Coefficient</td>
<td>SE</td>
<td>P value</td>
</tr>
<tr>
<td>1998-1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>63.229</td>
<td>2.502</td>
<td>&lt;.001</td>
<td>64.817</td>
<td>2.734</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Water facility</td>
<td>0.006</td>
<td>0.026</td>
<td>.10</td>
<td>0.003</td>
<td>0.026</td>
<td>.88</td>
</tr>
<tr>
<td>Sanitation</td>
<td>–0.209</td>
<td>0.027</td>
<td>&lt;.001</td>
<td>–0.124</td>
<td>0.030</td>
<td>.001^b</td>
</tr>
<tr>
<td>Wealth</td>
<td>–0.058</td>
<td>0.026</td>
<td>.05</td>
<td>–0.061</td>
<td>0.031</td>
<td>.05^b</td>
</tr>
<tr>
<td>Education</td>
<td>–0.160</td>
<td>0.034</td>
<td>&lt;.001</td>
<td>–0.223</td>
<td>0.036</td>
<td>.001^b</td>
</tr>
<tr>
<td>caste</td>
<td>–0.022</td>
<td>0.024</td>
<td>.10</td>
<td>–0.053</td>
<td>0.026</td>
<td>.04^b</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.462</td>
<td>0.042</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

2015-2016

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SE</th>
<th>P value</th>
<th>Coefficient</th>
<th>SE</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>27.756</td>
<td>2.514</td>
<td>&lt;.001</td>
<td>48.083</td>
<td>2.728</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Water facility</td>
<td>–0.020</td>
<td>0.024</td>
<td>.10</td>
<td>0.011</td>
<td>0.025</td>
<td>.65</td>
</tr>
<tr>
<td>Sanitation</td>
<td>–0.138</td>
<td>0.018</td>
<td>.001</td>
<td>–0.157</td>
<td>0.023</td>
<td>&lt;.001^b</td>
</tr>
<tr>
<td>Wealth</td>
<td>0.050</td>
<td>0.019</td>
<td>.001</td>
<td>–0.025</td>
<td>0.023</td>
<td>.28</td>
</tr>
<tr>
<td>Education</td>
<td>–0.002</td>
<td>0.022</td>
<td>.10</td>
<td>–0.095</td>
<td>0.024</td>
<td>&lt;.001^b</td>
</tr>
<tr>
<td>caste</td>
<td>–0.026</td>
<td>0.014</td>
<td>.05</td>
<td>0.000</td>
<td>0.016</td>
<td>.96</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.788</td>
<td>0.027</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

2019-2021

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SE</th>
<th>P value</th>
<th>Coefficient</th>
<th>SE</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>48.481</td>
<td>4.237</td>
<td>&lt;.001</td>
<td>43.869</td>
<td>3.551</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Water facility</td>
<td>0.103</td>
<td>0.039</td>
<td>.001</td>
<td>0.006</td>
<td>0.033</td>
<td>.85</td>
</tr>
<tr>
<td>Sanitation</td>
<td>–0.273</td>
<td>0.027</td>
<td>.001</td>
<td>–0.137</td>
<td>0.025</td>
<td>.001^b</td>
</tr>
<tr>
<td>Wealth</td>
<td>–0.081</td>
<td>0.016</td>
<td>.001</td>
<td>–0.104</td>
<td>0.018</td>
<td>.001^b</td>
</tr>
<tr>
<td>Education</td>
<td>–0.044</td>
<td>0.024</td>
<td>.10</td>
<td>0.010</td>
<td>0.020</td>
<td>.61</td>
</tr>
<tr>
<td>caste</td>
<td>–0.040</td>
<td>0.017</td>
<td>.01</td>
<td>–0.011</td>
<td>0.018</td>
<td>.56</td>
</tr>
<tr>
<td>Lambda</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.759</td>
<td>0.029</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

^aAkaike information criterion of spatial lag model (SLM) > SEM; National Family and Health Survey (NFHS)-2: 3153.1 (SLM), 3127.6 (SEM); NFHS-4: 4027.9 (SLM), 3988.9 (SEM); and NFHS-5: 4093 (SLM), 4052.8 (SEM).

^bP < .05 denotes a significant impact on stunting prevalence.

^cNot available.

Discussion

Principal Findings

This study used LISA modeling to establish a significant association between improved sanitation facilities and the reduction of growth failure health outcomes in children younger than 5 years in India. The relative risk maps identified the hotspots of CGF health cases, which could be targeted for future interventions. Nutritional status is critical for monitoring a country's development. Moreover, undernutrition is a global concern that poses numerous health, welfare, and well-being challenges [5]. Consequently, it becomes a pressing matter in the context of Sustainable Development Goals. Although global efforts have intensified over the past decade, governments have also strategized to address undernutrition. In addition, research has sought to unravel potential impediments that could lead to health and economic problems in developing countries. Although these issues were primarily associated with inadequate feeding patterns [12], this remains insufficient to address the problem, especially in India where undernutrition affects both the well-nourished and malnourished [13]. Researchers have striven to examine various factors that retard children’s growth, but clear evidence linking it to sanitation, particularly unimproved drinking water sources facilities (open defecation), is yet to emerge.

According to this study’s findings, more than a third of children younger than 5 years are malnourished. A study comparing the nutritional health of Indian children using WHO charts and an Indian synthetic chart also discovered a higher prevalence of undernutrition among this age group [34]. Neighboring countries

https://publichealth.jmir.org/2024/1/e41567 JMIR Public Health Surveill 2024 | vol. 10 | e41567 | p.1729 (page number not for citation purposes)
such as Nepal (36.0%) [35] and Pakistan (37.6%) [36] have nearly identical stunting rates, whereas Bangladesh (18.2%) [37] has comparable wasting and underweight rates. Our study shows a decreasing trend in stunting (44.9% to 38.4%) and underweight (46.7% to 35.7%) but an increasing prevalence of wasting (15.7% to 21.0%) over 15 years. Since 1990, despite a decrease in the prevalence of undernutrition in India, it has been a significant concern. The Global Burden of Disease Study (1990-2017) stated that malnutrition remained the leading risk factor for disease burden in India and that much faster progress is needed to meet India’s 2022 and global 2030 targets [14].

The spatial findings show that the CGF and water, sanitation, and hygiene (WaSH) indices are nonrandom at the national and district levels. Significant clusters for these indicators were found in several states. However, during the NFHS-4 survey, indicators improved for both factors in states such as Haryana, Punjab, Himachal Pradesh, Manipur, Mizoram, Karnataka, Uttarakhand, and Tripura. Many factors, ranging from demographic to socioeconomic, contribute to the increasing prevalence of undernutrition, which can be attributed to regional preferences in food consumption and disparities in the availability of health care facilities [38]. Other factors could be inadequate sanitation and lack of clean water, which elevate the prevalence of soil-transmitted diseases, such as diarrhea, contributing to undernutrition. This finding aligns with a study that found significant geographical variations across child malnutrition across Indian states, with particularly high rates in the central region [39].

The measure of spatial dependence was suggested by Moran I statistics. Stunting, underweight, and toilet facility have higher statistical values, confirming the regional gradient of malnutrition and sanitation in India. The results of the spatial error model show that increased sanitation and the level of mothers’ education are strongly and significantly associated with child malnutrition. As previously noted, poor sanitation can lead to diarrhea and other intestinal infections in children. Furthermore, the positive correlation between child nutrition and a mother’s education status aligns a previous study [40] that found a connection between maternal education and improved nutrition [41]. The findings of this study concerning the spatial relationship and correlation with WaSH factors are comparable to those of an Indian study using NFHS-4 data. Similar to the findings of this study, that study found a clear spatial dependence of the CGF indicator and increased clustering in the districts of Madhya Pradesh, Rajasthan, Bihar, Jharkhand, and Uttar Pradesh. There was also a statistically significant link between malnutrition and maternal education, improved sanitation, and wealth [42]. However, our findings reveal a strong link to WaSH variables, which, according to UNICEF, are important contributors to malnutrition [43]. Several studies have found that a lack of sanitary facilities increases the risk of nutritional deficiencies due to anorexia, fluids loss, and other complications that can lead to undernutrition [44]. Similarly, a study conducted in Ethiopia found that children living in households where open pit or unsanitary pit latrines were used were more likely to have malnutrition [45].

In 2014, the Indian government launched a cleanliness initiative called “Swachh Bharat Abhiyan” to enhance waste management and sanitation quality across the country, aligning with the sanitation aspect of child nutrition [46]. The multifaceted causes of undernutrition necessitate coordinated and consistent interventions that go beyond the norm. Exposure to fecal pathogens poses a significant threat to children’s growth, making improved WaSH conditions essential in all settings [47]. This could be achieved by embracing various aspects of WaSH practices, particularly incorporating sanitation into the nutrition program, strengthening the collaboration between WaSH and nutrition factors, and making WaSH initiatives more nutrition sensitive.

Strength and Limitations

Given the widespread socioeconomic and regional disparities among people in India, this study provides estimates of the prevalence of stunting, wasting, and underweight at the district level. The study’s regional spatial approach will assist policy makers in developing policies that ensure district-specific interventions for improving malnutrition. Moreover, the study’s results can be well generalized by using a nationally representative sample from a well-known large-scale survey in India. However, the study also has some limitations. First, because of the cross-sectional nature of the data, we could not establish a causal relationship among the variables. Second, because of the unavailability of NFHS-3 district-level data, a spatial analysis for NFHS-3 could not be conducted. Third, as the study relied on secondary data, we were unable to investigate community-level factors such as the availability of improved drinking water sources and sanitation facilities that might influence sanitation practice behavior among women. These factors could be investigated more deeply in future qualitative studies.

Conclusions

This study underscores that the malnutrition and undernutrition remain concerns in India. The prevalence of stunting does not seem to be decreasing or increasing consistently. However, wasting and underweight show increasing and decreasing trends, respectively. This study demonstrates that the presence of improved sanitation facilities and the educational status of mothers have a substantial positive correlation with the risk of CGF indicators. Policy makers can incorporate these factors when devising specific interventions to improve malnutrition indicators. The implementation of WaSH interventions should aim to save women’s time; if the time spent fetching water can be reduced, women could devote more time to childcare. Moreover, lack of access to WaSH facilities impacts the academic success of school-aged children, resulting in a decreased likelihood of securing employment, household food insecurity, and perpetuated poverty—all primary factors of child malnutrition. On the other hand, improved sanitation facilities have positive socioeconomic outcomes, such as supporting women’s dignity and safety, increasing girls’ school attendance, and facilitating nutrient recovery.
Acknowledgments
The authors acknowledge the Demographic Health Surveys (DHS) Program for providing the raw data for this study.

Conflicts of Interest
None declared.

References

20. SDG India index and dashboard 2021-22. NITI Aayog. URL: https://sdgindiaindex.niti.gov.in/urban/#/ [accessed 2024-02-14]
23. The DHS Program. Demographic and Health Survey. URL: https://www.dhsprogram.com/data/available-datasets.cfm [accessed 2024-02-14]


28. Anselin L. GeoDa: An introduction to spatial data science. URL: https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html#introduction [accessed 2024-02-14]


Abbreviations

CGF: child growth failure
DHS: Demographic and Health Survey
LISA: Local Indicator of Spatial Association
NFHS: National Family Health Survey
OLS: ordinary least squares
WaSH: water, sanitation, and hygiene
WHO: World Health Organization

Please cite as:
JMIR Public Health Surveill 2024;10:e41567
URL: https://publichealth.jmir.org/2024/1/e41567
doi:10.2196/41567
PMID:38787607

©Lovely Jain, Sreya Pradhan, Arun Aggarwal, Bijaya Kumar Padhi, Ramaiah Itumalla, Mahalaqua Nazli Khatib, Shilpa Gaidhane, Quazi Syed Zahiruddin, Celso Augusto Guimarães Santos, Khalid AL-Mugheed, Tahani Alrahbeni, Neelima Kukreti, Prakasini Satapathy, Sarvesh Rustagi, Petra Heidler, Roy Rillera Marzo. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 24.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Spatiotemporal Analysis of HIV/AIDS Incidence in China From 2009 to 2019 and Its Association With Socioeconomic Factors: Geospatial Study

Ziyi Xie1*, MEd; Bowen Chen1*, MEd; Zhizhuang Duan2, PhD

1Faculty of Humanities and Social Science, Macao Polytechnic University, Macao, China
2Xingzhi College, Zhejiang Normal University, Jinhua, China
* these authors contributed equally

Corresponding Author:
Zhizhuang Duan, PhD
Xingzhi College
Zhejiang Normal University
No. 3388 Yingbin Avenue
Lanxi City, Zhejiang Province
Jinhua, 321000
China
Phone: 86 18911026906
Email: 105009342@qq.com

Abstract

Background: The Joint United Nations Program on HIV/AIDS (UNAIDS) has set the “95-95-95” targets to ensure that 95% of all people living with HIV will know their HIV status, 95% of all people living with HIV will receive sustained antiretroviral therapy (ART), and 95% of all people receiving ART will achieve viral suppression (<1000 copies/mL). However, few countries have currently achieved these targets, posing challenges to the realization of the UNAIDS goal to eliminate the global HIV/AIDS epidemic by 2030. The Chinese government has implemented corresponding policies for HIV/AIDS prevention and control; however, it still faces the challenge of a large number of HIV/AIDS cases. Existing research predominantly focuses on the study of a particular region or population in China, and there is relatively limited research on the macro-level analysis of the spatiotemporal distribution of HIV/AIDS across China and its association with socioeconomic factors.

Objective: This study seeks to identify the impact of these factors on the spatiotemporal distribution of HIV/AIDS incidence in China, aiming to provide scientific recommendations for future policy development.

Methods: This study employed ArcGIS 10.2 (Esri) for spatial analysis, encompassing measures such as the imbalance index, geographical concentration index, spatial autocorrelation analysis (Moran I), and hot spot analysis (Getis-Ord Gi*). These methods were used to unveil the spatiotemporal distribution characteristics of HIV/AIDS incidence in 31 provinces of China from 2009 to 2019. Geographical Detector was used for ecological detection, risk area detection, factor detection, and interaction detection. The analysis focused on 9 selected socioeconomic indicators to further investigate the influence of socioeconomic factors on HIV/AIDS incidence.

Results: The spatiotemporal distribution analysis of HIV/AIDS incidence in China from 2009 to 2019 revealed distinct patterns. The spatial distribution type of HIV/AIDS incidence in China was random in 2009-2010. However, from 2011 to 2019, the distribution pattern evolved toward a clustered arrangement, with the degree of clustering increasing each year. Notably, from 2012 onwards, there was a significant and rapid growth in the aggregation of cold and hot spot clusters of HIV/AIDS incidence in China, stabilizing only by the year 2016. An analysis of the impact of socioeconomic factors on HIV/AIDS incidence in China highlighted the “urbanization rate” and “urban basic medical insurance fund expenditure” as the primary factors influencing the spatial distribution of HIV/AIDS incidence. Additionally, among social factors, indicators related to medical resources exerted a crucial influence on HIV/AIDS incidence.

Conclusions: From 2009 to 2019, HIV/AIDS incidence in China was influenced by various socioeconomic factors. In the future, it is imperative to optimize the combination of different socioeconomic indicators based on regional incidence patterns. This optimization will facilitate the formulation of corresponding policies to address the challenges posed by the HIV/AIDS epidemic.
KEYWORDS
HIV/AIDS; spatiotemporal distribution; cluster analysis; socioeconomic factors; China

Introduction

The Joint United Nations Program on HIV/AIDS (UNAIDS) has set the “95-95-95” targets to be achieved by 2025 [1], aiming for 95% of all people living with HIV/AIDS to be aware of their HIV status, 95% of people living with HIV/AIDS to receive sustained antiretroviral therapy (ART), and 95% of those receiving ART to achieve viral suppression (<1000 copies/mL) [2]. The global goal is to eliminate the HIV/AIDS epidemic worldwide by 2030, which is a public health threat [3]. However, a recent report by the UNAIDS titled “The path that ends AIDS: UNAIDS Global AIDS Update 2023” revealed that only 5 countries (Botswana, Eswatini, Rwanda, the United Republic of Tanzania, and Zimbabwe) have achieved the “95-95-95” targets ahead of schedule [4]. China, which is a middle-income country with a relatively high number of people living with HIV/AIDS [3], plays a crucial role in advancing the comprehensive realization of the “95-95-95” targets. Since the report of the first AIDS-related death in China in 1985 [5], the Chinese government has implemented a series of policies for the prevention and control of the HIV/AIDS epidemic, achieving certain successes [6]. However, due to its vast geographical expanse and large population, the HIV/AIDS epidemic in China, despite having an overall low prevalence rate [7], still presents a significant challenge for public health governance due to the substantial number of reported cases [8-10]. An epidemiological analysis of HIV/AIDS in China from 2004 to 2016 also confirmed the escalating severity of the epidemic and the significant regional disparities in HIV/AIDS infection rates [11]. Therefore, for HIV/AIDS prevention and control in China, it is particularly crucial to accurately identify high-risk areas within its vast territory and tailor HIV/AIDS prevention and control policies accordingly. The widespread application of geographic information technology in the field of epidemiological surveillance offers new avenues for addressing this challenge [12,13].

Existing studies have indicated a close correlation between the prevention and control of HIV/AIDS and both geographical spatial factors [14] and socioeconomic elements [15]. For instance, a study using gridded demographic data sets for spatial microsimulation to map HIV/AIDS indicators in low- and middle-income countries emphasized the need for appropriate socioeconomic categorization of data to explain social spatial disparities [16]. Another study conducted geographical clustering analysis of social capital distribution data and HIV/AIDS incidence, revealing that the spatial distribution of social capital would significantly impact HIV/AIDS care and prevention efforts [17]. Additionally, an analysis of the geographical distribution disparities in HIV mortality rates in South Africa identified more pronounced local spatial clustering characteristics in periurban areas and around national highways, and this finding was attributed to these areas being population clusters but lacking sufficient economic support for people living with HIV/AIDS to seek medical assistance [18]. Furthermore, 2 studies on the clustering analysis of HIV mortality rates among children and pregnant women further confirmed this viewpoint [19,20]. From this, it can be seen that studies in other countries have already linked these factors to the prevention and control of HIV/AIDS. However, there is currently a scarcity of research in academia that analyzes the HIV epidemic in China based on these 2 categories of factors, and the limited research conducted often focuses on the spatiotemporal analysis of data from specific regions or populations. For instance, in terms of geographic selection, the majority of current spatiotemporal analyses of HIV/AIDS in China targeted specific regions, with little detailed discussion on the correlation between spatiotemporal factors and socioeconomic factors [21-23]. While some studies have conducted spatiotemporal analyses of HIV/AIDS data across China, they have not linked the findings to socioeconomic factors [24,25]. Additionally, in some spatiotemporal analyses based on the entire HIV/AIDS population in China, characteristics of specific groups, such as youth and men who have sex with men, have been studied. Although these findings have been connected to macroscopic socioeconomic factors, the degree of the influence of corresponding microscopic indicators under these factors on spatiotemporal distribution has not been thoroughly explored [26-28]. This provides research space for our study to further investigate the correlation between the spatiotemporal distribution characteristics of HIV/AIDS incidence across China and the subdivision indicators under socioeconomic factors.

Additionally, it is worth emphasizing that, while the COVID-19 pandemic in China in 2020 posed a threat to HIV/AIDS treatment and health care services [29], the stringent lockdown measures may have influenced the incidence of HIV/AIDS differently [30]. Research in the context of this sudden public health event needs to be carried out separately to ensure scientific rigor. Therefore, this study specifically selected data on the incidence of HIV/AIDS in China before the occurrence of the COVID-19 pandemic. It is important to note that the standard system for assessing the HIV/AIDS epidemic includes infection and mortality data. The reason for choosing the incidence of HIV/AIDS as the core criterion in this study instead of the incidence rate is that the infection rate involves a considerable long incubation period for HIV and issues related to voluntary or mandatory testing [31]. Meanwhile, the mortality rate is, in a sense, an extension of the incidence rate. In terms of selecting socioeconomic indicators, a study on the spatiotemporal distribution and influencing factors of HIV/AIDS incidence in China found that the level of economic development (gross domestic product [GDP]) is the most closely related factor [32]. Another study further demonstrated that, in addition to overall economic conditions, factors, such as the urbanization rate and population density, are closely correlated with HIV/AIDS incidence [33]. Furthermore, several other studies on the access of people living with HIV/AIDS to health care...
services have shown that the accessibility of medical services [34] and individual-level economic conditions [35,36] also impact HIV/AIDS incidence, and this situation can be improved if the government continues to increase public health expenditure [37]. Building upon the aforementioned existing research, we combined statistical data published by the China National Bureau of Statistics (CNBS) with the actual situation of HIV/AIDS prevention and control in China, ultimately selecting 9 representative socioeconomic indicators as variables.

Taking into consideration the aforementioned factors, this study initially employed the imbalance index, geographical concentration index, spatial autocorrelation analysis (Moran I), and hot spot analysis (Getis-Ord Gi*) in ArcGIS 10.2 (Esri). These analyses aim to reveal the spatiotemporal distribution characteristics of HIV/AIDS incidence (both the number of cases and incidence rate per 100,000 people) in 31 provinces of China from 2009 to 2019. Subsequently, Geographical Detector was employed to conduct ecological detection, risk area detection, factor detection, and interaction detection. This analysis explored the impact of 9 indicators related to China’s social and economic factors on the spatiotemporal distribution of HIV/AIDS incidence. The aim was to provide policy recommendations for China to further address the HIV/AIDS epidemic effectively.

Methods

Data Collection

This study integrated HIV/AIDS incidence data from 2009 to 2019 in 31 provinces of China (excluding Hong Kong, Macao, and Taiwan) with socioeconomic indicators to facilitate a comprehensive analysis and causal discussion in both temporal and spatial dimensions. Specifically, social indicators included the number of health care institutions, number of beds in health care institutions, number of health care personnel, urbanization rate, and residential population density. Economic indicators included GDP, expenditure on health care in local finances, per capita disposable income of residents, and urban basic medical insurance fund expenditure. The socioeconomic data were sourced from the CNBS, while the HIV/AIDS incidence data were obtained from the Chinese Public Health Science Data Center (CPHSD). Specifically, the HIV/AIDS incidence data from CPHSD are compiled based on the Chinese Infectious Disease Reporting System. Local hospitals and Centers for Disease Control and Prevention (CDC) regularly report newly diagnosed HIV/AIDS cases through this network system and calculate HIV/AIDS incidence, which is then used by the Chinese Center for Disease Control and Prevention (CCDC) to compile national HIV/AIDS data [38]. Spatial administrative boundary vector data are derived from the China basic geographic information data released by the National Geomatics Center of China (NGCC).

Furthermore, to eliminate differences in dimensional units among various indicators and meet the requirements of the Geographical Detector for operational data, the Natural Breaks method in ArcGIS 10.2 software was employed to categorize the independent variables. Each socioeconomic indicator was divided into 5 levels, represented by X1 to X9 (Table 1).

### Table 1. Socioeconomic indicators (factors) influencing the incidence of HIV/AIDS in China.

<table>
<thead>
<tr>
<th>Influencing Indicator</th>
<th>Code</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross domestic product</td>
<td>X1</td>
<td>5</td>
</tr>
<tr>
<td>Expenditure on health care in local finances</td>
<td>X2</td>
<td>5</td>
</tr>
<tr>
<td>Per capita disposable income of residents</td>
<td>X3</td>
<td>5</td>
</tr>
<tr>
<td>Urban basic medical insurance fund expenditure</td>
<td>X4</td>
<td>5</td>
</tr>
<tr>
<td>Number of health care institutions</td>
<td>X5</td>
<td>5</td>
</tr>
<tr>
<td>Number of beds in health care institutions</td>
<td>X6</td>
<td>5</td>
</tr>
<tr>
<td>Number of health care personnel</td>
<td>X7</td>
<td>5</td>
</tr>
<tr>
<td>Urbanization rate</td>
<td>X8</td>
<td>5</td>
</tr>
<tr>
<td>Residential population density</td>
<td>X9</td>
<td>5</td>
</tr>
</tbody>
</table>

Ethical Considerations

The study used original data sourced from the CPHSD, an official government institution, which publicly releases data. According to the “Ethical Review Measures for Human Life Science and Medical Research” issued by the National Health Commission of China and the “Scientific Data Management Measures” issued by the General Office of the State Council of the People’s Republic of China, researchers are permitted to conduct secondary analysis of publicly available anonymized data for scientific research purposes without requiring Institutional Review Board (IRB) scrutiny and approval. These regulations exempt research involving pre-existing data that are publicly available or that are recorded in a manner that subjects cannot be identified directly or through identifiers linked to the subjects. Given that the data used in our study were both anonymized and publicly available, IRB review was not necessary for the study. Additionally, the authors have obtained permission to access and use the data, ensuring adherence to data usage policies and ethical standards.
Statistical Methods

Imbalance Index
The imbalance index is a crucial indicator that illustrates the degree of distribution balance of the research subjects in different regions [39]. The imbalance index (S) takes values between 0 and 1, with a larger S value indicating a greater imbalance in the distribution of HIV/AIDS incidence within the study area. Its calculation formula is as follows:

Geographical Concentration Index
The geographical concentration index is a crucial indicator that indicates the degree of geographical concentration of the research object [40]. The geographical concentration index (G) takes values between 0 and 100. A higher G value indicates a more concentrated spatial distribution of HIV/AIDS incidence in China, while a lower value suggests a more dispersed spatial distribution. Its calculation formula is as follows:

Spatial Autocorrelation Analysis (Moran I)
The Moran I index is a significant indicator for analyzing the spatial distribution relationships among units in the study area [41]. It can be used to identify and measure the distribution status and clustering degree of HIV/AIDS incidence in the 31 provinces of China. An I value >0 indicates positive spatial autocorrelation among the units, while a value <0 indicates negative spatial autocorrelation. A larger I value signifies a more significant spatial correlation. The significance of the Moran I index is generally tested using the Z-score formula. When the Z-score exceeds the critical value of 1.96, it indicates the statistical significance of spatial clustering, with the probability of randomly generating this clustering pattern being less than 5%. The calculation formula is as follows:

Hot Spot Analysis (Getis-Ord Gi*)
Hot spot analysis is an important analytical method that explores the similarity and heterogeneity between the attribute values of spatial geographic units and their adjacent units [42]. This method compares the local sum of a particular feature and its adjacent features with the total sum of all features. When the local sum of an area significantly differs from the expected local sum to the extent that it cannot be attributed to random chance, the area is identified as a high- or low-value cluster (hot or cold spot) based on the attribute values. This facilitates the visualization of the spatial distribution of HIV/AIDS incidence. The calculation formula is as follows:

Geographical Detector
The Geographical Detector encompasses 4 detection modes (ecological detection, risk area detection, factor detection, and interaction detection) used to uncover the driving factors behind the spatial distribution of a particular phenomenon [43]. It is applied for detecting and assessing HIV/AIDS incidence. Ecological detection, in this context, aims to determine if there is a significant difference in the mean HIV/AIDS incidence between 2 factors, X_1 and X_2. The calculation formula is as follows:

Among these, risk area detection is used to assess whether there is a significant difference in the mean HIV/AIDS incidence between 2 subregions. The calculation formula is as follows:

Factor detection is a method used to assess the extent to which the 9 socioeconomic indicators can explain the spatial variation of HIV/AIDS incidence. The calculation formula is as follows:

Interaction detection is primarily employed to identify the interactions among the 9 socioeconomic indicators (Xs). It assesses whether the joint effect of any 2 socioeconomic indicators increases or decreases the explanatory power regarding HIV/AIDS incidence. Additionally, it examines whether the impact of these indicators on HIV/AIDS incidence is mutually independent. The results of the interaction can be categorized into the following 5 types: nonlinear attenuation, si-factor nonlinear attenuation, bi-factor authentication, mutually independent, and nonlinear enhancement (Table 2).

Table 2. Types of interactions among socioeconomic indicators (factors) influencing the incidence of HIV/AIDS in China.

<table>
<thead>
<tr>
<th>Basis for judgment</th>
<th>Type of interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>q(X_1 ∩ X_2) &lt; Min(q(X_1),q(X_2))</td>
<td>Nonlinear attenuation</td>
</tr>
<tr>
<td>Min(q(X_1),q(X_2)) &lt; q(X_1 ∩ X_2) &lt; Max(q(X_1),q(X_2))</td>
<td>Si-factor nonlinear attenuation</td>
</tr>
<tr>
<td>q(X_1 ∩ X_2) &gt; Max(q(X_1),q(X_2))</td>
<td>Bi-factor authentication</td>
</tr>
<tr>
<td>q(X_1 ∩ X_2) = q(X_1) + q(X_2)</td>
<td>Mutually independent</td>
</tr>
<tr>
<td>q(X_1 ∩ X_2) &gt; q(X_1) + q(X_2)</td>
<td>Nonlinear enhancement</td>
</tr>
</tbody>
</table>
Results

Spatiotemporal Distribution Characteristics of HIV/AIDS Incidence in China

This study initially used the imbalance index and geographical concentration index to measure the equilibrium and concentration of the spatiotemporal distribution characteristics of HIV/AIDS incidence in 31 provinces of China from 2009 to 2019. Subsequently, spatial autocorrelation analysis (Moran I) was introduced to account for the spatial proximity of the study objects, further specifying the spatial distribution types of HIV/AIDS incidence from 2009 to 2019. Finally, hot spot analysis was employed to visually depict the evolutionary characteristics of the spatiotemporal distribution of HIV/AIDS incidence in China from 2009 to 2019.

Based on the calculation results of the imbalance index and geographical concentration index (Table 3), there was a slight downward trend in both indices from 2013 to 2018, but the values remained stable within a certain range. Overall, the imbalance index (S) stabilized in the range of 0.4-0.5, and the geographical concentration index (G) stabilized in the range of 24-34, indicating that the spatiotemporal distribution of HIV/AIDS incidence in China from 2009 to 2019 exhibited a certain degree of imbalance and concentration. Moreover, this distribution state demonstrated a certain level of persistence and stability over time.

Table 3. Imbalance index and geographical concentration index of HIV/AIDS incidence in China from 2009 to 2019.

<table>
<thead>
<tr>
<th>Year</th>
<th>HIV incidence rate (/100,000 people)</th>
<th>Imbalance index (S)</th>
<th>Geographical concentration index (G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>1.00</td>
<td>0.58</td>
<td>34.68</td>
</tr>
<tr>
<td>2010</td>
<td>1.20</td>
<td>0.54</td>
<td>32.09</td>
</tr>
<tr>
<td>2011</td>
<td>1.53</td>
<td>0.53</td>
<td>31.77</td>
</tr>
<tr>
<td>2012</td>
<td>3.11</td>
<td>0.54</td>
<td>30.25</td>
</tr>
<tr>
<td>2013</td>
<td>3.12</td>
<td>0.49</td>
<td>27.53</td>
</tr>
<tr>
<td>2014</td>
<td>3.33</td>
<td>0.46</td>
<td>26.28</td>
</tr>
<tr>
<td>2015</td>
<td>3.69</td>
<td>0.42</td>
<td>24.72</td>
</tr>
<tr>
<td>2016</td>
<td>3.97</td>
<td>0.40</td>
<td>24.13</td>
</tr>
<tr>
<td>2017</td>
<td>4.14</td>
<td>0.41</td>
<td>24.30</td>
</tr>
<tr>
<td>2018</td>
<td>4.62</td>
<td>0.41</td>
<td>24.81</td>
</tr>
<tr>
<td>2019</td>
<td>5.10</td>
<td>0.44</td>
<td>25.89</td>
</tr>
</tbody>
</table>

Using the spatial autocorrelation analysis tool in ArcGIS 10.2, the Moran I value for the spatial distribution of HIV/AIDS incidence in China from 2009 to 2019 was obtained (Table 4). The results indicated that the Moran I value was only 0.05 in 2009-2010, suggesting a random spatial distribution pattern for HIV/AIDS incidence in China during that period. In 2011, the Moran I value increased to 0.07 and passed the significance test at the 5% level, indicating a transition to a clustered distribution pattern. Subsequently, the Moran I value continued to rise, with an increase from 0.07 to 0.25, demonstrating strong positive spatial autocorrelation, and it consistently passed the significance test at the 5% level. This confirms that the distribution pattern of HIV/AIDS incidence in China evolved from a random distribution to a clustered distribution from 2011 to 2019, with the degree of clustering increasing annually, demonstrating strong spatial aggregation and dependence.
Table 4. Moran I values and spatial distribution types of HIV/AIDS incidence in China from 2009 to 2019.

<table>
<thead>
<tr>
<th>Year</th>
<th>Moran I</th>
<th>Z-score</th>
<th>P value</th>
<th>Distribution type^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.05</td>
<td>1.44</td>
<td>.14</td>
<td>Random</td>
</tr>
<tr>
<td>2010</td>
<td>0.05</td>
<td>1.43</td>
<td>.15</td>
<td>Random</td>
</tr>
<tr>
<td>2011</td>
<td>0.07</td>
<td>1.71</td>
<td>.09</td>
<td>Clustered</td>
</tr>
<tr>
<td>2012</td>
<td>0.11</td>
<td>2.05</td>
<td>.04</td>
<td>Clustered</td>
</tr>
<tr>
<td>2013</td>
<td>0.14</td>
<td>2.48</td>
<td>.01</td>
<td>Clustered</td>
</tr>
<tr>
<td>2014</td>
<td>0.16</td>
<td>2.67</td>
<td>.01</td>
<td>Clustered</td>
</tr>
<tr>
<td>2015</td>
<td>0.19</td>
<td>2.97</td>
<td>&lt;.001</td>
<td>Clustered</td>
</tr>
<tr>
<td>2016</td>
<td>0.22</td>
<td>3.32</td>
<td>&lt;.001</td>
<td>Clustered</td>
</tr>
<tr>
<td>2017</td>
<td>0.23</td>
<td>3.51</td>
<td>&lt;.001</td>
<td>Clustered</td>
</tr>
<tr>
<td>2018</td>
<td>0.23</td>
<td>3.55</td>
<td>&lt;.001</td>
<td>Clustered</td>
</tr>
<tr>
<td>2019</td>
<td>0.25</td>
<td>3.77</td>
<td>&lt;.001</td>
<td>Clustered</td>
</tr>
</tbody>
</table>

^aEvolution from a random distribution to a clustered distribution is indicated.

For the study of spatial distribution characteristics, areas with higher or lower attribute values are often the first to attract attention. However, the high or low values of attributes do not necessarily represent statistically significant hot or cold spots [40]. Therefore, areas that can be considered hot or cold spots should have features with high values (low values) and be surrounded by other features that also have high values (low values). To more precisely illustrate these spatial distribution relationships, this study used the Hot Spot Analysis tool in ArcGIS 10.2 to conduct Getis-Ord Gi* statistical analysis. The analysis identified hot and cold spots in the spatial clustering distribution of HIV incidence in China from 2009 to 2019 (Figure 1). It should be noted that during the data analysis process, we found that the distribution of hot and cold spots did not change from 2016 to 2019. Therefore, the distribution status of hot and cold spots from 2016 to 2019 has been combined and presented as 1 map in Figure 1H.

Figure 1. Spatial distribution map of hot and cold spots of HIV/AIDS incidence in China from 2009 to 2019. (A) 2009, (B) 2010, (C) 2011, (D) 2012, (E) 2013, (F) 2014, (G) 2015, (H) 2016-2019.

The number of cold spot clusters of HIV incidence in China started from 0 in 2009-2010 and experienced significant and rapid growth in 2012. Although there was a slight decline in the number of cold spots in 2014, it did not prevent the overall stability of cold spot clusters at 6-7 from 2012 to 2019. In general, from 2009 to 2019, the spatial extent of low-value cold spot areas expanded significantly, spreading from Shandong Province to the central and northern provinces of China. Ultimately, these cold spot clusters stabilized and concentrated in the central and northern provinces of China (Shandong, Beijing, Tianjin, Hebei, Shanxi, Liaoning, and Inner Mongolia).

In terms of hot spot clusters, the number of hot spot clusters of HIV/AIDS incidence in China showed a trend of first decreasing and then increasing, but the magnitude of the increase and decrease was not significant. It mainly stabilized in 5 southwestern provinces of China (Yunnan, Guangxi, Sichuan,
Guizhou, and Chongqing), with an improvement in both aggregation and significance. Overall, compared to cold spot areas, hot spot areas exhibited a smaller fluctuation range and stronger spatial distribution stability.

**Impact of Socioeconomic Factors on the Spatiotemporal Distribution of HIV/AIDS Incidence in China**

HIV/AIDS, as an infectious disease, is profoundly influenced by socioeconomic factors in its spatiotemporal distribution [44]. This study selected 9 socioeconomic indicators related to HIV/AIDS incidence and employed the Geographical Detector for quantitative exploration, aiming to analyze the extent of the impact of these indicators on the spatial distribution of HIV/AIDS incidence in China.

The results of risk detection and ecological detection indicated, at a significance level of 0.05, that there were no significant differences in the mean values of attributes between subregions of different levels. Among the socioeconomic indicators, only the urbanization rate, per capita disposable income, and population density exhibited significantly different impact capacities. There was no significant difference in the impact of other indicators on the spatial distribution of HIV/AIDS incidence in China. This suggests a certain rationality in the selection and grading of indicators.

The results of factor detection showed the strength of the impact of various factors on the spatial distribution of HIV/AIDS incidence, with the magnitude of q values as follows (Table 5): urbanization rate ($X_8=0.49$) > urban basic medical insurance fund expenditure ($X_4=0.31$) > the number of beds in health care institutions ($X_6=0.28$) > local financial expenditure on health care ($X_7=0.26$) > the number of health care personnel ($X_7=0.26$) > the number of health care institutions ($X_5=0.19$) > provincial GDP ($X_1=0.12$) > nonsignificant impacts for per capita disposable income and population density ($X_3=0.05$ and $X_9=0.06$, respectively).

<table>
<thead>
<tr>
<th>Indicator</th>
<th>q statistic</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>0.12</td>
<td>.60</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.26</td>
<td>.19</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0.05</td>
<td>.85</td>
</tr>
<tr>
<td>$X_4$</td>
<td>0.31</td>
<td>.17</td>
</tr>
<tr>
<td>$X_5$</td>
<td>0.19</td>
<td>.35</td>
</tr>
<tr>
<td>$X_6$</td>
<td>0.28</td>
<td>.15</td>
</tr>
<tr>
<td>$X_7$</td>
<td>0.26</td>
<td>.19</td>
</tr>
<tr>
<td>$X_8$</td>
<td>0.49</td>
<td>.01</td>
</tr>
<tr>
<td>$X_9$</td>
<td>0.06</td>
<td>.80</td>
</tr>
</tbody>
</table>

*aThe indicators are clarified in Table 1.*

Based on the above detection results, it can be concluded that the urbanization rate and the expenditure on urban basic medical insurance were the most influential factors explaining the spatial distribution of HIV/AIDS incidence in China. Provinces with low urbanization rates and minimal spending on urban basic medical insurance were more likely to face a severe situation of HIV/AIDS incidence. The number of beds in health care institutions, local financial expenditure on health care, number of health care personnel, and number of health care institutions were secondary factors influencing the distribution. These 4 indicators belong to health care resource factors within social indicators, indicating that a better health care development situation has a certain inhibitory effect on the distribution of HIV/AIDS incidence in China. GDP had a certain explanatory power for the distribution of the incidence but with a low P value, suggesting lower confidence in its explanatory power and indicating a correlation (not significant) with the spatial distribution of HIV/AIDS incidence in China. Additionally, per capita disposable income and population density had a relatively weak impact on the spatial distribution pattern of HIV/AIDS incidence in China.

The results of interactive detection are shown in Table 6. After the interaction of any 2 of the 9 socioeconomic indicators, all q-values exhibited varying degrees of increase, showing nonlinear enhancement or bivariate enhancement states, with no independent or weakening relationships. This suggests that, compared to single indicators, the coupling effect of any 2 factors can enhance the explanatory power of the spatiotemporal distribution pattern of HIV/AIDS incidence in China, indicating that the spatial distribution differences of HIV/AIDS incidence in China are the result of the interaction of multiple influencing factors. According to the analysis results of ecological detection, the differences in the effects of each indicator were not statistically significant, which indicates the significant importance of the comprehensive synergistic effect of multiple factors in exploring HIV/AIDS incidence.
Furthermore, among all combinations, the enhancement effect of the urbanization rate with other indicators was the most significant. The output results of ecological detection also indicated that only the urbanization rate exhibited a significant difference with other variables. For instance, GDP alone had a weak explanatory effect on the distribution of the disease, but its explanatory effect was strengthened when coupled with the urbanization rate or local financial health expenditure. This indicates that among the 9 indicators, the urbanization rate should be the primary factor of concern. Additionally, after interaction, the factor combinations that achieved an explanatory degree with a q-value above 0.7 included residential population density ($X_9$), per capita disposable income of residents ($X_3$), expenditure on health care in local finances ($X_4$), and urban basic medical insurance fund expenditure ($X_5$). Although residential population density had a low individual impact on disease distribution, its explanatory effect significantly improved when coupled with economic indicators related to health care. This suggests that for HIV/AIDS incidence, residential population density should not be considered in isolation but should be simultaneously considered in conjunction with the synergistic effects of economic factors.

In terms of the type of enhanced interaction, apart from the nonlinear enhancement dominating the interactions of GDP ($X_1$), per capita disposable income of residents ($X_3$), and residential population density ($X_9$), the interactions of the remaining 6 factors were predominantly bivariate enhancement. Furthermore, the detection results of factors $X_1$, $X_3$, and $X_9$ were the weakest. Therefore, it is once again emphasized that GDP, per capita disposable income of residents, and residential population density exhibited relatively weak spatial distribution effects on HIV/AIDS incidence. However, it is important to note that these 3 factors are still relevant factors affecting HIV/AIDS incidence in China, albeit to a lesser extent compared to other factors.

**Discussion**

This study conducted a spatiotemporal distribution analysis of HIV/AIDS incidence in 31 provinces of China from 2009 to 2019. Furthermore, it integrated the analysis with 9 economic and social indicators to explore the extent of the impact of socioeconomic factors on the spatiotemporal distribution of HIV/AIDS incidence in China. The research revealed that the occurrence of HIV/AIDS in China is significantly influenced by a combination of factors, including socioeconomic indicators, beyond the medical origins of the disease.

Before delving into various discussions, it is important to note that from the late 20th century to the early 21st century, China experienced localized outbreaks of HIV/AIDS in certain regions due to specific reasons. For instance, in Henan Province, from the late 1980s to the early 1990s, against the backdrop of official blood donation quotas, market demand from pharmaceutical companies, and the emergence of a blood-selling industry in rural areas, widespread transmission of HIV/AIDS occurred through blood products from rural blood stations due to technical and management flaws in blood collection. In August 2001, the Chinese Ministry of Health first publicized the situation of an “HIV/AIDS village” in Wenlou Village, Shangcai County, Henan Province, where around 43% of the approximately 3710 blood sellers were reported to be HIV positive [45]. In 2004, for the first time, Henan Province conducted a comprehensive HIV/AIDS census across the province. Out of the 280,000 individuals with a history of paid blood donation, 25,000 were found to be people living with HIV/AIDS [46]. Similarly, in Yunnan and Sichuan provinces, which serve as important land routes from southwestern China to Southeast Asia and South Asia, there was a relatively serious drug trafficking problem [47], leading to intravenous drug use becoming the main transmission route for HIV in these areas [48,49]. Existing research data have shown that before 2001, 74.8% of people living with HIV/AIDS identified in Sichuan Province were drug users, while 62.8% of those identified in Yunnan Province from 1989 to 2000 were drug users [50].

After the early 21st century, although the Chinese government increased its attention to the AIDS issue and gradually implemented corresponding prevention and control efforts (such as the “Four Frees and One Care” policy in 2003 and the promulgation of the HIV/AIDS Prevention and Control
Regulation in 2006), the phenomenon of previous HIV/AIDS outbreaks gradually subsided. However, due to the large number of reported HIV/AIDS cases in China prior to these policies taking effect, coupled with a certain lag in policy implementation, HIV/AIDS prevention and control did not immediately curb the nationwide spread of HIV/AIDS. This also explains why the spatial distribution of HIV/AIDS incidence in China was random in 2009-2010 but gradually evolved from random distribution to clustered distribution from 2011 to 2019, as found in this study.

First, as a manageable chronic disease [51], the spatial distribution of HIV/AIDS is influenced by socioeconomic factors. For instance, the government’s investment in public health is one of the influencing factors. Since the end of 2003, China has implemented the “Four Frees and One Care” policy [10]. All people living with HIV/AIDS diagnosed at the CCDC can receive antiretroviral drugs for free. However, it was not until 2012, when important antiretroviral drugs like lamivudine (3TC) and efavirenz (EFV) gradually entered domestic production, that the government could progressively reduce the cost of obtaining drugs [52], thereby enhancing the practical implementation of the “Four Frees and One Care” policy in China. From the analysis of spatial distribution characteristics in this study, it is evident that although the incidence of HIV/AIDS in China from 2009 to 2019 exhibited continuous features of imbalance and concentration in spatial distribution, there was a significant turning point around 2012. This may be related to the continuous expansion of the coverage of the “Four Frees and One Care” policy. Subsequently, the results obtained from our application of the Geographical Detector to analyze the impact of medical factors once again indirectly corroborated this viewpoint. In this regard, it is necessary to acknowledge that some economically underdeveloped regions benefit more because the government can ensure the transfer and distribution of basic antiretroviral drugs (resulting in an increase in cold spots). However, due to differences in economic development and the proportion of public health expenditure among regions, there is a direct impact on the accessibility of new antiretroviral drugs [53]. This has a strong correlation with the spatiotemporal distribution differences in the incidence of HIV/AIDS. In simpler terms, in regions where the incidence of HIV/AIDS was relatively mild but economic development levels and public health expenditure were low, the comprehensive implementation of the “Four Frees and One Care” policy in 2012 ensured the right of people living with HIV/AIDS to receive free antiretroviral drugs. Although subsequent public health challenges in HIV/AIDS prevention and control still exist in these areas, considering the initially low disease burden, the overall incidence is expected to further decrease. In regions where the HIV/AIDS epidemic was already severe, after 2012, they faced more acute local financial difficulties. The lack of momentum in public health expenditure directly results in a large number of people living with HIV/AIDS who are unable to access quality medical resources and new antiretroviral drugs. Consequently, the incidence in these areas has not reached the desired level of control. This finding suggests that, in future reforms of HIV/AIDS prevention and control policies, the Chinese government should pay attention to balancing the relationship between policy and local economic conditions.

Furthermore, parallel to the “Four Frees and One Care” policy, there is an ongoing reform of China’s medical insurance system. In March 2009, China issued the “Opinions on Deepening the Reform of the Medical and Health System” [54], marking a period of improvement in China’s medical security system. This is one of the reasons why this study chose 2009 as the starting point for data observation. In 2012, several government departments in China jointly issued a document requiring the comprehensive implementation of total control over medical insurance payments [55]. This led to a significant increase in the pressure on local public health expenditure at the regional level. This situation is intertwined with the regional disparities in providing free antiretroviral drugs under the “Four Frees and One Care” policy. However, it is crucial to note that there is no absolute proportional relationship between local GDP and the intensity of investment in public health [56]. In other words, a higher GDP in a certain region does not necessarily mean a synchronous increase in investment in public health. This is also clearly reflected in the previous data of this study. However, in 2016, the Chinese Ministry of Human Resources and Social Security, along with the Ministry of Finance, officially initiated the direct settlement of inpatient medical expenses across regions [57], facilitating the access of the public to medical care across different areas. This development is consistent with the earlier findings of our analysis, indicating that from 2016 to 2019, the spatial and temporal distributions of HIV/AIDS incidence in China stabilized and the impact of factors, such as population density and per capita disposable income, on the HIV/AIDS incidence rate decreased. However, despite the potential enhancement of shared health care resources across regions, the actual economic burden borne by individuals still exhibits significant regional economic disparities. For instance, the contribution base for medical insurance, which is closely tied to local fiscal investment in public health, continues to vary markedly across regions [58]. This may explain why, after 2016, the disparity in HIV/AIDS incidence rates among different regions in China did not widen or narrow. This finding suggests a direction for the Chinese government to prioritize and strengthen its focus on certain indicators in the ongoing efforts to combat the HIV/AIDS epidemic.

Looking further, how do these indicators specifically and directly relate to the incidence of HIV/AIDS among individuals, as mentioned previously? First, in terms of economic indicators, the current “Four Frees and One Care” policy provides a combination of free antiretroviral drugs that are categorized into first-line (tenofovir disoproxil fumarate [TDF]/azidothymidine [AZT]/abacavir [ABC]+3TC+nevirapine [NVP]/EFV/rtipivirine [RPV]) and second-line (TDF+3TC+dolutegravir [DTG]/lopinavir combined with ritonavir [LPV/r]; AZT+TDF+3TC+DTG/LPV/r) treatments [52]. Typically, first-line drugs are centrally purchased by the CCDC and systematically distributed to various regions. As for second-line drugs, their availability primarily relies on the level of financial investment by local governments in public health care. This directly results in situations where drugs like LPV/r, previously categorized as a second-line treatment, are not widely accessible in all regions. In comparison with numerous antiretroviral drugs, DTG, which is considered to have fewer side effects and lower susceptibility to resistance, is more accessible in all regions. In this way, the spatial and temporal distributions of HIV/AIDS incidence in China stabilized and the impact of factors, such as population density and per capita disposable income, on the HIV/AIDS incidence rate decreased. However, despite the potential enhancement of shared health care resources across regions, the actual economic burden borne by individuals still exhibits significant regional economic disparities. For instance, the contribution base for medical insurance, which is closely tied to local fiscal investment in public health, continues to vary markedly across regions [58]. This may explain why, after 2016, the disparity in HIV/AIDS incidence rates among different regions in China did not widen or narrow. This finding suggests a direction for the Chinese government to prioritize and strengthen its focus on certain indicators in the ongoing efforts to combat the HIV/AIDS epidemic.
representative [59]. Currently, access to DTG for individuals in different regions can be categorized in 3 ways. First, it is directly obtained for free through the “Four Frees and One Care” policy in certain regions (such as cold spot areas like Beijing). Second, in areas where DTG is included in medical insurance, infected individuals with insurance coverage can purchase DTG and be reimbursed through insurance (such as Zhejiang) [60]. Third, in regions where DTG is not covered by medical insurance, infected individuals have to self-fund the purchase, and in some cases, they have to resort to ordering it through mail from other regions (such as Sichuan and other hot spot areas) [61]. It is evident that economic indicators in different regions influence whether people living with HIV/AIDS can access antiretroviral drugs for free or acquire updated and more effective medications at lower costs, thereby affecting the incidence of HIV/AIDS among individuals.

Certainly, another crucial factor in preventing the onset of the disease is the timely detection of treatment failure (such as drug resistance) in people living with HIV/AIDS. This is closely related to the social indicators mentioned in this study. In our research, we observed that regions with higher levels of health care resources are more effective at suppressing the incidence of HIV/AIDS. In different regions, there are significant variations in whether people living with HIV/AIDS need routine examinations and whether the costs of such checks, which are conducted each time they receive free antiretroviral drugs under the “Four Frees and One Care” policy (as stipulated by the policy, approximately once every 3 months), need to be self-funded by the individuals. In some regions with more developed public health services, there are even provisions to offer 1 to 2 free drug resistance testing services annually for individuals [52]. In other words, in regions with more abundant public health care resources, individuals are more likely to timely detect their drug-resistance situations, enabling targeted changes in antiretroviral drug combinations and thus preventing the onset of the disease. In addition, it is worth noting that China has implemented the “designated hospital” system for HIV/AIDS medical care, as outlined in the 2013 “Notice on Submitting the List of Designated Hospitals for AIDS by the Medical Administration Bureau of the Ministry of Health” [62]. While this system aims to reduce urban-rural disparities in implementing the “Four Frees and One Care” policy, its effectiveness has been a subject of mixed opinions and has, to some extent, led to instances of nondesignated hospitals refusing to provide medical services. A region that is more densely populated, economically developed, and well-endowed with public medical resources usually has more numerous and specialized designated hospitals, with a high degree of professionalism and comprehensive capabilities. For instance, hospitals like You’an Hospital and Ditan Hospital in Beijing (a cold spot region) have become the “most renowned” designated hospitals for people living with HIV/AIDS in the country. In contrast, designated hospitals in regions with relatively scarce public medical resources mainly rely on infectious disease specialist hospitals as they lack comprehensive departmental setups. This directly results in medical challenges for individuals living with HIV/AIDS in those areas. In fact, before the onset of the disease, the medical needs of people living with HIV/AIDS are not limited to the field of infectious diseases, similar to the general population [63]. However, given the vulnerability of the immune system [64], if people living with HIV/AIDS do not receive timely and effective medical services, they are more susceptible to overall health deterioration due to other diseases. This also demonstrates the significant impact of medical accessibility on the onset of the disease. In the future, the Chinese government should continue to optimize medical resource allocation for this specific group of people living with HIV/AIDS to address the public health challenges posed by the HIV/AIDS epidemic.

Finally, this study not only revealed the spatiotemporal distribution of HIV/AIDS incidence in China as a result of the comprehensive impact of multiple influencing factors through the interactive detection of geographic detectors but also discovered that the dual-factor coupling of 9 socioeconomic indicators can have different effects on the spatiotemporal distribution of HIV/AIDS incidence in China. This finding can provide insights for the Chinese government to focus on the combined optimization of socioeconomic indicators when formulating HIV/AIDS prevention and control policies. Specifically, when 2 indicators show a significantly improved coupling effect, the government can enact policies to promote the development of the optimal indicator combination to further reduce the incidence of HIV/AIDS. For example, in the case where the coupling effect of the urbanization rate with other indicators is most noticeable, the government can actively formulate policies to leverage the synergistic effects between this indicator and others while continuing to promote urbanization.

This study has certain limitations. First, due to the lack of authoritative data statistics on the incidence of HIV/AIDS in Hong Kong, Macao, and Taiwan by official institutions, this study used relevant data from 31 other provinces in China. Second, the socioeconomic indicators used in this study are macro and common statistical indicators, but the causes of HIV/AIDS are complex and diverse, and these indicators may not necessarily cover all influencing factors. To overcome these limitations, future efforts should focus on collecting more detailed data and interpreting them in the context of emerging research. However, as one of the few studies addressing this topic, our research covers the majority of China’s regions, and the analyzed content and recommendations have a relatively high degree of universality.

This study conducted a spatiotemporal distribution analysis of HIV/AIDS incidence in China from 2009 to 2019. The findings revealed that, starting from 2012, there was an enhancement in the clustering pattern of HIV/AIDS incidence in China, with cold and hot spots becoming more pronounced. From 2016 to 2019, the clustering of cold and hot spots began to stabilize. Additionally, we investigated the impact of HIV/AIDS incidence from 2009 to 2019 in relation to 9 common socioeconomic indicators. The study findings suggest that optimizing the combination of different socioeconomic indicators can further suppress the incidence of HIV/AIDS. These findings provide insights for the government to adjust relevant policies based on the specific impact of different indicators on HIV/AIDS incidence in the future.
Acknowledgments
This study did not use generative AI during the writing process. This research project did not receive any external financial support or funding.

Data Availability
The data sets analyzed during this study are available in the Chinese Public Health Science Data Center repository [65].

Authors' Contributions
ZX and BC, as co-first authors, made significant contributions to the conception and design of the work and actively participated in the acquisition, analysis, and interpretation of the data. They collaborated in drafting and substantively revising the manuscript. ZD provided valuable insights into the study design and critically revised the manuscript. All authors approved the submitted version and any substantial modifications. Each author agrees to be personally accountable for the contributions and commits to addressing questions regarding the accuracy or integrity of the work, ensuring appropriate investigation, resolution, and documentation in the literature.

Conflicts of Interest
None declared.

References
1. 2025 AIDS Targets. UNAIDS. URL: https://aidstargets2025.unaids.org/ [accessed 2023-11-20]


65. Study data sets. Chinese Public Health Science Data Center repository. URL: https://www.phsciencedata.cn/Share/kysjml.jsp?id=%22c2ca694e-3995-4c7f-9078-3ed0af14556%22 [accessed 2024-05-30]

Abbreviations

3TC: lamivudine
ART: antiretroviral therapy
AZT: azidothymidine
CCDC: Chinese Center for Disease Control and Prevention
CNBS: China National Bureau of Statistics
CPHSD: Chinese Public Health Science Data Center
DTG: dolutegravir
EFV: efavirenz
GDP: gross domestic product
IRB: Institutional Review Board
LPV/r: lopinavir combined with ritonavir
NVP: nevirapine
RPV: rilpivirine
TDF: tenofovir disoproxil fumarate
UNAIDS: Joint United Nations Program on HIV/AIDS
Pulmonary Tuberculosis Notification Rate Within Shenzhen, China, 2010-2019: Spatial-Temporal Analysis

Peixuan Lai¹, BD; Weicong Cai¹, MD; Lin Qu², BD; Chuangyue Hong¹, MD; Kaihao Lin¹, MD; Weiguo Tan¹, PhD; Zhiguang Zhao¹, PhD

¹Shenzhen Center for Chronic Disease Control, Shenzhen, China
²School of Public Health, Sun Yat-sen University, Guangzhou, China

Abstract

Background: Pulmonary tuberculosis (PTB) is a chronic communicable disease of major public health and social concern. Although spatial-temporal analysis has been widely used to describe distribution characteristics and transmission patterns, few studies have revealed the changes in the small-scale clustering of PTB at the street level.

Objective: The aim of this study was to analyze the temporal and spatial distribution characteristics and clusters of PTB at the street level in the Shenzhen municipality of China to provide a reference for PTB prevention and control.

Methods: Data of reported PTB cases in Shenzhen from January 2010 to December 2019 were extracted from the China Information System for Disease Control and Prevention to describe the epidemiological characteristics. Time-series, spatial-autocorrelation, and spatial-temporal scanning analyses were performed to identify the spatial and temporal patterns and high-risk areas at the street level.

Results: A total of 58,122 PTB cases from 2010 to 2019 were notified in Shenzhen. The annual notification rate of PTB decreased significantly from 64.97 per 100,000 population in 2010 to 43.43 per 100,000 population in 2019. PTB cases exhibited seasonal variations with peaks in late spring and summer each year. The PTB notification rate was nonrandomly distributed and spatially clustered with a Moran $I$ value of 0.134 ($P=.02$). One most-likely cluster and 10 secondary clusters were detected, and the most-likely clustering area was centered at Nanshan Street of Nanshan District covering 6 streets, with the clustering time spanning from January 2010 to November 2012.

Conclusions: This study identified seasonal patterns and spatial-temporal clusters of PTB cases at the street level in the Shenzhen municipality of China. Resources should be prioritized to the identified high-risk areas for PTB prevention and control.

(Keywords: tuberculosis; spatial analysis; spatial-temporal cluster; Shenzhen; China)

Introduction

Tuberculosis (TB), caused by the bacillus Mycobacterium tuberculosis, is a chronic communicable disease that is a major public health and social problem and one of the leading causes of death from a single bacterial pathogen worldwide [1]. According to the World Health Organization (WHO) 2023 report, there were 7.5 million new TB cases and 1.3 million deaths caused by TB estimated in 2022 globally [1]. The 30 high-TB-burden countries accounted for 87% of all estimated incident cases worldwide [1]. China had the third highest TB burden in the world in 2022, with an estimated 530,000 new cases of TB, accounting for 7.1% of the global total [1]. To reduce the magnitude and burden of TB, China has implemented several effective interventions such as the adoption and implementation of the End-TB...
Strategy, expansion of directly observed therapy strategies (DOTS), and expansion of multidrug-resistant TB diagnosis and treatment centers. As a result, China has made great progress in reducing the annual TB incidence by 3.4% since 1990 [2]. However, TB remains a major public health problem in China.

Over the past few decades, China’s urbanization process, led by nationwide economic reforms, has been accompanied by the rapid growth of cities, including Shenzhen, due to the migration of the younger population from rural areas owing to more attractive employment opportunities [3], contributing to large-scale domestic migrations. Rural labor migrants might change TB transmission patterns in the cities; moreover, the health problems of migrants, including TB infection and active disease development, might change under the new circumstance of living in the city [4]. As a special economic zone, Shenzhen has experienced remarkable development in terms of both the economic scale and population growth. However, the massive influx of migrants to Shenzhen has brought serious challenges for the prevention and control of TB [3].

Identifying areas where TB is geographically concentrated is particularly essential for the planning, implementation, monitoring, and evaluation of TB control programs and to inform the allocation of resources for targeted and effective interventions [5]. In recent years, spatial-temporal analysis has been widely used to describe the distribution characteristics and transmission patterns of TB, demonstrating that TB exhibits highly dynamic and spatially heterogeneous patterns at provincial, national, and international levels during certain periods of time [6,7]. However, few studies have been performed to reveal the changes in the small-scale clustering of TB cases at the street level. In this study, we performed a spatial-temporal cluster analysis to characterize the distribution and patterns of pulmonary tuberculosis (PTB) cases at the street level in the Shenzhen municipality of China from 2010 to 2019.

**Methods**

**Study Setting**

Shenzhen is a city located in southern China at a longitude of 113°43’-114°38’ East and a latitude of 22°24’-22°52’ North (see Multimedia Appendix 1), with 74 streets among its 10 districts. Shenzhen has greatly benefited from the implementation of reforms and opening-up policies of China in the 1980s and developed quickly from a small fishery village of less than 1 million inhabitants to a modern metropolis with over 10 million inhabitants, making it one of the most developed cities in China [8]. Notably, over 70% of the residents in Shenzhen are migrants and most of these migrants are young, with those 65 years and older accounting for only 3.22% of the population [8]. The gross domestic product (GDP) of Shenzhen has increased by more than 110,000 times in the past 40 years, from less than 200 million yuan in 1979 (~US $28 million) to approximately 2.69 trillion yuan in 2019 (~US $370 billion) [8].

**Data Source**

Data of reported PTB cases in Shenzhen from January 2010 to December 2019 were extracted from the China Information System for Disease Control and Prevention, which is established and operated by the Chinese Center for Disease Control and Prevention (CCDC). Each reported case contains demographic and medical information and is identified by a unique identity card number to avoid duplicate reports.

The annual population data of each administrative street for each year were extracted from Shenzhen Statistical Yearbooks (2010-2019) of the Shenzhen Municipality Bureau of Statistics [8]. Basic map data were obtained from the Shenzhen Geographical Information Public Service Platform. The administrative number of the street was used as a reference at a 1:500,000 scale to correlate and match the data of reported PTB cases on each street to establish the database.

As one of the most serious infectious diseases in China, each diagnosed PTB case must be reported online within 24 hours after diagnosis. The Diagnostic Criteria for PTB (WS288-2017) issued by the Ministry of Health (the former National Health Commission) of the People’s Republic of China in 2017 were used in this study, in which cases of PTB were diagnosed using radiography, pathogen detection, and pathologic diagnosis [9]. All forms of PTB were included in this study, including bacteriologically confirmed and clinically diagnosed PTB, previously treated and new PTB, and childhood and adult PTB.

**Ethical Considerations**

This study was approved by the Ethics Committee of Shenzhen Center for Chronic Disease Control (SZCCC-2023-003-01-PJ). Personal sensitive information such as the name and phone number associated with each case were blocked and excluded in this study to protect the patients’ privacy. A secondary analysis based on reported data was conducted and thus informed consent from individuals was not required. This study was conducted strictly in accordance with Norms for the Management of Information Reporting on Infectious Diseases in China [10].

**Data Analysis**

**Descriptive and Time-Series Analyses**

The epidemiological characteristics of PTB cases were analyzed. PTB cases reported from 2010 to 2019 were aggregated annually by age, sex, occupation, permanent residents and migrant population, and date of onset for PTB notification rate analysis. Comparison between different groups was performed using the $\chi^2$ test or Fisher exact test as appropriate. All statistical analyses were performed in SPSS version 26 and a two-tailed $P$ value less than .05 was considered statistically significant.

In addition, time-series seasonal decomposition analysis was used to identify the seasonality of the PTB notification rate in Shenzhen municipality. The time series of reported PTB cases were decomposed into seasonal variation, long-term trend, and random effect to explore the temporal patterns. The temporal patterns were determined according to the onset date of the monthly registration of all reported PTB cases. The time series included 120 months from January 2010 to December 2019 and were examined using Excel software.
Spatial Autocorrelation Analysis

Moran $I$ was selected as the global index of spatial autocorrelation to detect the spatial distribution pattern of PTB cases in Shenzhen municipality, China. The expression of the global Moran $I$ statistic is as follows:

$$I = \frac{\sum (x_i - \bar{x})(x_j - \bar{x})w_{ij}}{\sum (x_i - \bar{x})^2}$$

where $n$ is the number of units and $x_i$ and $x_j$ are the attribute values of unit $i$ and unit $j$, respectively. The Moran $I$ value varies between $-1$ and $1$; when $|I|$ is larger, the autocorrelation is higher, and $I=0$ indicates no autocorrelation [11]. Both the Z score and $P$ value calculated by GeoDa 1.20 (Arizona State University) were used to test the significance of the Global Moran $I$ value.

Spatial-Temporal Scan Analysis

Kulldorff spatiotemporal scan statistical analysis was performed to explore the spatial, temporal, and spatial-temporal clusters of PTB across different streets geographically and at different time periods [12]. The SaTScan 10.1 software (Kulldorff) was used for spatial-temporal scanning analysis, which is based on a moving-column scanning window that contains geographical information with the height corresponding to time [13]. According to a previous study [6], the maximum radius of the spatial-scanning window and the maximum length of the temporal-scanning window was set to 11% of the population at risk and 30% of the whole study period, respectively. The log likelihood ratio (LLR) of different circle centers and various radii was calculated as follows to compare the notification rate of PTB inside and outside the scanning window [14]:

$$LLR = \log \left( \frac{\sum n_z}{\sum n_g} \right)$$

Where $n_z$ signifies the observed count of events within spatiotemporal window $z$, $n_g$ denotes the total aggregate of events across the entirety of the study area, $u_z$ represents the expected count of events within the particular spatiotemporal window $z$, and $\mu_g$ indicates the overall anticipated count of events within the study area. A larger LLR value indicates a more likely cluster. The Monte Carlo simulation test was used to evaluate whether the likelihood of the cluster was statistically significant, and the relative risk (RR) was calculated as the estimated risk within the cluster divided by that outside the cluster [15,16]. For each possible spatiotemporal cluster, when the $P$ value is less than .05, a larger LLR value indicates that the area covered by the dynamic scanning window is more likely to be a cluster region. The window with the largest LLR value selected from all the scanning windows is considered the strongest cluster and the secondary clusters are the other windows with statistically significant LLR values. ArcMap software (Esri) was used to visualize the scanning results.

Results

Descriptive Analysis of PTB Cases

A total of 58,122 PTB cases were notified in Shenzhen from 2010 to 2019, including 38,358 (66.00%) cases from male patients. Over 60% of cases were from the migrant population; 20,795 cases (35.78%) were detected in workers and 20,629 (35.49%) were from individuals who were unemployed. The mean age of the patients diagnosed with PTB was 34.73 (SD 14.13) years, with a range from 0 to 97 years (Figure 1).
Temporal Patterns of PTB Cases

The annual average notification rate of PTB for the 10-year study period was 51.62 per 100,000 population. As shown in Figure 2, the total notification rate of PTB showed a significantly volatile downward trend over time from the highest rate of 64.97 per 100,000 population in 2010 to the lowest rate of 43.43 per 100,000 population in 2019 ($P < .001$).
The monthly notification rates of PTB showed a trend of volatility and decline over the study period in Shenzhen municipality. PTB cases have shown seasonal variations with the highest number of PTB cases notified in May and July each year, followed by a steady decreasing trend after July and a nadir in January and February (Figure 3).

Spatial Patterns of PTB Cases
As displayed in Table 1, the global spatial autocorrelation analysis showed that the Moran I value of the PTB notification rate in Shenzhen municipality from 2010 to 2019 was positive (0.134) and the P value was consistently less than .05, indicating that the notification rate of PTB in Shenzhen municipality was nonrandomly distributed. Annually, except for the years 2012 and 2014, there was significant global spatial autocorrelation in the PTB notification rates every year with the Moran I value ranging from 0.130 to 0.353. Therefore, further spatial-temporal scan analysis of PTB was needed.
Table 1. Global spatial autocorrelation analysis of the pulmonary tuberculosis notification rate from 2010 to 2019 in Shenzhen municipality, China.

<table>
<thead>
<tr>
<th>Year</th>
<th>Moran I value</th>
<th>Z score</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>0.353</td>
<td>5.320</td>
<td>.001</td>
</tr>
<tr>
<td>2011</td>
<td>0.179</td>
<td>2.748</td>
<td>.002</td>
</tr>
<tr>
<td>2012</td>
<td>0.095</td>
<td>1.398</td>
<td>.07</td>
</tr>
<tr>
<td>2013</td>
<td>0.143</td>
<td>2.103</td>
<td>.02</td>
</tr>
<tr>
<td>2014</td>
<td>0.047</td>
<td>0.894</td>
<td>.19</td>
</tr>
<tr>
<td>2015</td>
<td>0.130</td>
<td>1.851</td>
<td>.03</td>
</tr>
<tr>
<td>2016</td>
<td>0.190</td>
<td>2.786</td>
<td>.006</td>
</tr>
<tr>
<td>2017</td>
<td>0.217</td>
<td>3.065</td>
<td>.002</td>
</tr>
<tr>
<td>2018</td>
<td>0.201</td>
<td>2.852</td>
<td>.002</td>
</tr>
<tr>
<td>2019</td>
<td>0.155</td>
<td>2.361</td>
<td>.02</td>
</tr>
<tr>
<td>2010-2019</td>
<td>0.134</td>
<td>1.894</td>
<td>.02</td>
</tr>
</tbody>
</table>

aThe autocorrelation is considered significant at $P<.05$ (two-tailed).

Spatial Clustering of PTB Cases

Spatial clustering analysis of notified PTB cases every year from 2010 to 2019 showed that the most likely clusters have changed dynamically over time. The most likely clusters of PTB in Shenzhen municipality from 2010 to 2014 were concentrated in the southwest regions, covering the streets of Nanshan District and Baoan District. From 2015 to 2019, the most likely clusters were mainly distributed in the central and northern streets of Shenzhen municipality, including Longhua District, Longgang District, and Guangming District (Figure 4).
Temporal Clustering of PTB Cases

As presented in Table 2, the temporal clustering analysis in each year showed that PTB notification rates were concentrated in late spring and summer annually, ranging from March to August. The greatest aggregated time for PTB across the whole study period was observed from March 2010 to September 2010. During this period, there was a total of 4562 notified PTB cases and the risk of PTB-related incidents was 53% (RR 1.53, P<.001) higher than that in other time periods.
Table 2. Temporal clustering of notified pulmonary tuberculosis cases from 2010 to 2019 in Shenzhen municipality, China.

<table>
<thead>
<tr>
<th>Year</th>
<th>Cluster time frame</th>
<th>Observed cases</th>
<th>Expected cases</th>
<th>RR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>LLR&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>March 1 to May 31</td>
<td>2136</td>
<td>1696.58</td>
<td>1.38</td>
<td>72.30</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2011</td>
<td>July 1 to September 30</td>
<td>1642</td>
<td>1469.98</td>
<td>1.16</td>
<td>13.13</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2012</td>
<td>February 1 to April 30</td>
<td>1500</td>
<td>1286.07</td>
<td>1.23</td>
<td>22.79</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2013</td>
<td>April 1 to June 30</td>
<td>1430</td>
<td>1185.74</td>
<td>1.29</td>
<td>32.14</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2014</td>
<td>June 1 to August 31</td>
<td>1600</td>
<td>1344.21</td>
<td>1.27</td>
<td>31.31</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2015</td>
<td>April 1 to June 30</td>
<td>1714</td>
<td>1428.82</td>
<td>1.28</td>
<td>36.40</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2016</td>
<td>April 1 to June 30</td>
<td>1693</td>
<td>1418.70</td>
<td>1.27</td>
<td>33.92</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2017</td>
<td>July 1 to August 31</td>
<td>1458</td>
<td>1186.66</td>
<td>1.29</td>
<td>35.34</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2018</td>
<td>March 1 to May 31</td>
<td>1766</td>
<td>1517.62</td>
<td>1.23</td>
<td>26.28</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2019</td>
<td>June 1 to August 31</td>
<td>1698</td>
<td>1460.91</td>
<td>1.23</td>
<td>24.88</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2010-2019</td>
<td>March 1 to September 30, 2010</td>
<td>4562</td>
<td>3072.58</td>
<td>1.53</td>
<td>334.01</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>RR: relative risk.
<sup>b</sup>LLR: log likelihood ratio.
<sup>c</sup>Considered significant at P<.05 (two-tailed).

Spatial-Temporal Clustering of PTB Cases

The results of spatial-temporal clustering analysis for notified PTB cases in Shenzhen municipality from 2010 to 2019 are shown in Table 3 and Figure 5, indicating that the PTB notification rates were spatiotemporally clustered. A total of 1 most-likely cluster and 10 secondary clusters were detected in this study. The most-likely cluster area was distributed in the southwestern region of Shenzhen municipality and the clustering time was from January 2010 to November 2012 (RR 1.96, P<.001) with a total of 3000 PTB cases notified during this period. In addition, the area centered at Nanshan Street of Nanshan District (22.53 N, 113.94 E) with a radius of 6.39 kilometers covered 6 streets, similar to the most likely cluster identified in the purely spatial clustering analysis. The other 10 secondary clusters were mainly located in the central and northwestern regions of Shenzhen municipality, covering streets in Baoan District, Nanshan District, Longgang District, Guangming District, and Longhua District, along with several relatively small areas (clusters 1, 4, and 8) in Futian District and Luohu District. The main clustering time ranged from January 2010 to March 2013 except for the secondary clusters 2, 5, 6, and 9.
<table>
<thead>
<tr>
<th>Cluster type</th>
<th>Cluster time frame</th>
<th>Coordinates/radius (km)</th>
<th>Prefec-tures, n</th>
<th>Expected cases</th>
<th>Observed cases</th>
<th>RR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>LLR&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most likely cluster</td>
<td>January 1, 2010, to November 30, 2012</td>
<td>22.53 N, 113.94 E/6.39</td>
<td>6</td>
<td>3000</td>
<td>1573.33</td>
<td>1.96</td>
<td>527.74</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 1</td>
<td>April 1, 2010, to February 28, 2013</td>
<td>22.58 N, 114.09 E/0</td>
<td>1</td>
<td>441</td>
<td>143.58</td>
<td>3.09</td>
<td>198.21</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 2</td>
<td>February 1, 2017, to November 30, 2018</td>
<td>22.72 N, 114.02 E/8.98</td>
<td>7</td>
<td>1943</td>
<td>1258.93</td>
<td>1.56</td>
<td>163.27</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 3</td>
<td>January 1, 2010, to September 30, 2012</td>
<td>22.58 N, 113.99 E/4.49</td>
<td>2</td>
<td>888</td>
<td>458.81</td>
<td>1.95</td>
<td>158.79</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 4</td>
<td>March 1, 2010, to January 31, 2013</td>
<td>22.52 N, 114.07 E/0</td>
<td>1</td>
<td>728</td>
<td>366.75</td>
<td>2.00</td>
<td>139.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 5</td>
<td>March 1, 2015, to November 30, 2017</td>
<td>22.77 N, 114.31 E/6.33</td>
<td>2</td>
<td>772</td>
<td>416.77</td>
<td>1.86</td>
<td>121.77</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 6</td>
<td>March 1, 2013, to November 30, 2015</td>
<td>22.78 N, 113.90 E/0</td>
<td>1</td>
<td>373</td>
<td>147.19</td>
<td>2.54</td>
<td>121.47</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 7</td>
<td>March 1, 2010, to June 30, 2012</td>
<td>22.63 N, 113.84 E/5.28</td>
<td>3</td>
<td>1516</td>
<td>1006.46</td>
<td>1.52</td>
<td>113.75</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 8</td>
<td>March 1, 2010, to September 30, 2011</td>
<td>22.55 N, 114.14 E/3.72</td>
<td>6</td>
<td>706</td>
<td>447.87</td>
<td>1.58</td>
<td>63.76</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 9</td>
<td>July 1, 2017, to September 30, 2017</td>
<td>22.63 N, 114.14 E/5.94</td>
<td>7</td>
<td>261</td>
<td>164.74</td>
<td>1.59</td>
<td>23.92</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Secondary cluster 10</td>
<td>March 1 to July 31, 2010</td>
<td>22.73 N, 113.81 E/5.13</td>
<td>4</td>
<td>317</td>
<td>219.60</td>
<td>1.45</td>
<td>19.05</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>RR: relative risk.

<sup>b</sup>LLR: log likelihood ratio.

<sup>c</sup>Statistical significance is indicated at P<.05 (two-tailed).
Discussion

Principal Findings

To the best of our knowledge, this is the first study to present a spatial-temporal cluster analysis of PTB at the street level in Shenzhen municipality using routinely collected PTB surveillance data, which identified seasonal patterns and spatial-temporal clusters of PTB in Shenzhen.

The notification rate of PTB in Shenzhen municipality declined steadily during the 10-year study period, decreasing from the peak of 64.97 cases per 100,000 population in 2010 to the lowest of 43.43 cases per 100,000 population in 2019. This downward trend is in line with other municipal, provincial, and national studies [6,17-20]. This achievement is due in large part to the great importance attached by the local government and health administrative department in Shenzhen to PTB control and prevention in recent years. First, financial support received from the government has increased continuously from 28.2 million yuan (~US $3.9 million) in the Twelfth Five-Year Plan (2011-2015) to approximately 197.64 million yuan (~US $27 million) in the Thirteenth Five-Year Plan (2016-2020) for TB control and prevention in Shenzhen. Subsequently, an integrated system under collaboration of the CCDC, tuberculosis-designated hospitals, and community health service centers was established and developed among institutions with a clear division of labor and coordination [21]. In addition, a batch of molecular biological testing equipment for PTB diagnosis and treatment was purchased and distributed to designated PTB medical institutions in various streets and districts of the city [21]. As a result, the diagnosis time of PTB and the possibility of pretreatment transmission have declined, despite an increase in the overall percentage of cases detected and confirmed bacteriologically. At the same time, PTB screening programs were added to the routine physical examination for the older population aged ≥ 65 years and for primary and secondary school students, and rifampicin resistance testing was further strengthened for patients with PTB [22]. In addition to the DOTS recommended by the WHO since 2000, Shenzhen has implemented a series of measures, including providing free drug treatment and a certain degree of reimbursement for examination costs, which enable early diagnosis and timely treatment for patients with PTB to effectively curb the spread of disease. With the implementation of these targeted and effective measures, the PTB epidemic situation has improved substantially, although it remains a great public health challenge in Shenzhen municipality.

According to the time-series analysis, seasonal variations and cyclical trends of PTB cases were observed, with apparent peaks
in late spring and summer, especially in May and July. This is consistent with the findings of neighboring Guangzhou city [23]. Shenzhen has a subtropical monsoon humid climate, a long summer and short winter, abundant sunshine, and abundant rainfall, with an average annual temperature of 20 °C and relative humidity of 80%. This is aligned with the transmission of respiratory infectious diseases and may be related to the spread of PTB. Previous studies indicated that lack of exposure to ultraviolet rays from sunlight and poor ventilation in the indoor environment may increase the risk of PTB infection [24-26]. Consistent with other reports [18,19], we found the lowest number of notified PTB cases in January and February. On the one hand, this may be closely related to the massive public transportation and population flow during the Chinese Spring Festival, which usually occurs in late January or February [23,27], when many migrants leave Shenzhen to return to their hometowns. On the other hand, people are busy celebrating the Lunar New Year at this time, coupled with frequent visits to relatives and friends, and thereby may avoid seeking medical care [18,19]. These multiple factors would result in a significant decline in PTB cases reported but an increased risk of transmission among the infectious source and close contacts during the holiday, concentrating in late spring and summer after an incubation period of several months to half a year, including a 2-month interval from symptom appearance to a medical diagnosis [25]. This may be one of the contributors for the peak months observed in Shenzhen. In addition, the return of migrants and delayed diagnosis should also be considered. Therefore, the detection and tracking of patients with PTB should be strengthened during the Spring Festival, and suspected or confirmed cases should be transferred to designated medical institutions for further diagnosis and treatment as soon as possible to shorten the delay time of treatment and reduce the risk of transmission. Health education should also be emphasized and effective measures such as actively wearing masks and opening windows for ventilation should be promoted, especially in crowded places.

Many studies have confirmed a spatial-clustering distribution of PTB in certain areas [6,18,19,28]. In this study, the global spatial autocorrelation analysis indicated that the notification rate of PTB in Shenzhen municipality displayed an obvious spatial-clustering distribution between 2010 and 2019. The results of advanced local spatial autocorrelation showed that the clusters of PTB at the street level had a dynamic change over time from the southwest to the central and northern part of the city. The hot spots of PTB observed in this study are largely consistent with the PTB notification rates in Shenzhen. PTB has long been known as “the disease of the poor,” and poverty has been considered one of the causes of disease clustering [28-31]. However, a previous study indicated that the incidence of PTB is higher in areas with a better economic situation than in less developed areas because the former areas are more attractive to population inflows [32]. Many migrants are at increased risk of ill health because of the adverse conditions through which they travel and then work and live [33]. Migrant workers are usually not entitled to the social welfare and health resources that are available to local permanent residents, which poses a challenge for this specific population in accessing health care services [34,35]. In this study, most of the notified PTB cases were detected in workers or unemployed individuals, shifting in parallel with the economic development of local areas of the city. The regional economic development in the past 10 years started from Nanshan District and Baoan District in southwest Shenzhen, where the GDP ranked in the top three during the study period. With the change of urban planning, the GDP in central and northern Shenzhen, including Longhua District, Longgang District, and Guangming District, has grown rapidly recently, which has brought about a more obvious population agglomeration effect in these areas [36]. Similar dynamics can also be found in Beijing [37] and Shanghai [38]. Understanding the interaction of TB transmission, population migration, and social development, albeit highly complex and dynamic, is essential for the control and prevention of TB [32].

To consider the role of time when evaluating the geographical distribution of PTB, spatial-temporal scanning analysis was used to supplement the local spatial analysis. The results from the spatial-temporal clustering analysis of the PTB cases from 2010 to 2019 showed that the most likely cluster was concentrated in southwestern Shenzhen, covering 6 streets of 2 districts, and the clustering period was from January 2010 to November 2012. In addition, most of the secondary clustering areas were also found in this period, indicating that this period represents a peak of PTB transmission in Shenzhen. The clusters identified by the spatial-temporal scanning analysis and the simple spatial clustering analysis were similar, which may indicate the robustness of our results. More importantly, these similar findings indicate that more effective and targeted measures should be urgently developed and implemented in the high-risk areas for PTB control and prevention in Shenzhen municipality.

Compared with the Guangdong provincial and China national findings in recent years [39,40], Shenzhen has entered a period with a low epidemic level of PTB. However, there is still a long way to go to reach the WHO’s goal of eliminating PTB. Achieving this ambitious goal requires stronger, more tailored and effective responses. First, the clustering results can serve as a guide to develop more accurate and effective interventions, with a focus on areas where PTB is concentrated, and to strengthen the deployment and implementation of corresponding actions to prevent and discover the outbreak and epidemic status of PTB at the street level. PTB diagnosis, treatment, and care for migrants should be integrated into the general health services, while special efforts may be needed to reach migrants for improving the availability, accessibility, and quality of comprehensive medical services [41]. Screening for PTB contacts and selected high-risk groups should be linked to follow-up, strategies for preventive treatment, or referral to the treatment program [41]. In addition, systematic and extensive health educational campaigns are needed for the provision of more accessible information to raise public awareness of PTB and improve public access to relevant health care services. Third, increasing the speed of referral of newly detected patients with PTB to designated clinics is also crucial to minimize the risk of disease transmission and infections in the community.
Limitations

This is the first study to analyze the spatial-temporal clustering characteristics of PTB at the street level and identify the high-risk areas of PTB in Shenzhen municipality, which provides valuable information for future strategies and measures of PTB prevention and control. However, this study was subject to several limitations. First, our analysis was based on data extracted from the National Surveillance System and we were unable to preclude the possibility of missing cases. This might cause an underestimation of the PTB epidemic in Shenzhen. Second, this study only focused on the spatial and temporal patterns and clusters of PTB cases. Potential risk factors associated with PTB incidence, such as individual habits, socioeconomic status, living conditions, and environmental pollutants, were not evaluated. Further research should take these limitations into consideration.

Conclusions

This study identified spatial and temporal patterns and spatial-temporal clusters of PTB cases at the street level in Shenzhen municipality from 2010 to 2019. A volatile downward trend of PTB incidence over the study period was observed in Shenzhen municipality. The most likely clustering areas changed from the southwest to the central and northern part of the city, and the most likely clustering time was late spring and summer. Resources should be prioritized to high-risk areas for PTB prevention and control.

Acknowledgments

This study was supported by the Chinese Center for Disease Control and Prevention. We appreciate the contributions made by colleagues of the tuberculosis program for their assistance in data collection. This study was supported by the Shenzhen Health Economics Association Research Fund (202369), Shenzhen Medical Key Discipline Construction Fund (SZXK065), National Natural Science Foundation of China (82373641), and Guangdong Scientific and Technological Foundation (2020B111170014). The funding bodies had no role in the study design, data collection, analysis, or manuscript preparation.

Data Availability

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors’ Contributions

PL and ZZ conceptualized the study and established the research framework. PL and WC were responsible for the methodology. PL contributed to software and analysis. WT and ZZ performed validation. PL and WC performed the formal analysis. ZZ contributed resources. PL, WC, LQ, CH, and KL wrote the original draft of the manuscript. WT and ZZ contributed to manuscript editing and review. PL was responsible for data visualization. PL and ZZ undertook project administration. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Location of Shenzhen municipality in China.

References


11. Kulldorff M. SatScan user guide for version 9. URL: [https://www.satscan.org/techdoc.html] [accessed 2024-03-03]


Abbreviations

CCDC: Chinese Center for Disease Control and Prevention
DOTS: directly observed therapy strategy
GDP: gross domestic product
LLR: log likelihood ratio
PTB: pulmonary tuberculosis
RR: relative risk
TB: tuberculosis
WHO: World Health Organization

©Peixuan Lai, Weicong Cai, Lin Qu, Chuangyue Hong, Kaihao Lin, Weiguo Tan, Zhiguang Zhao. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 14.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: The World Health Organization declared mpox an international public health emergency. Since January 1, 2022, China has been ranked among the top 10 countries most affected by the mpox outbreak globally. However, there is a lack of spatial epidemiological studies on mpox, which are crucial for accurately mapping the spatial distribution and clustering of the disease.

Objective: This study aims to provide geographically accurate visual evidence to determine priority areas for mpox prevention and control.

Methods: Locally confirmed mpox cases were collected between June and November 2023 from 31 provinces of mainland China excluding Taiwan, Macao, and Hong Kong. Spatiotemporal epidemiological analyses, including spatial autocorrelation and regression analyses, were conducted to identify the spatiotemporal characteristics and clustering patterns of mpox attack rate and its spatial relationship with sociodemographic and socioeconomic factors.

Results: From June to November 2023, a total of 1610 locally confirmed mpox cases were reported in 30 provinces in mainland China, resulting in an attack rate of 11.40 per 10 million people. Global spatial autocorrelation analysis showed that in July (Moran I = 0.0938; P = .08), August (Moran I = 0.1276; P = .08), and September (Moran I = 0.0934; P = .07), the attack rates of mpox exhibited a clustered pattern and positive spatial autocorrelation. The Getis-Ord Gi* statistics identified hot spots of mpox attack rates in Beijing, Tianjin, Shanghai, Jiangsu, and Hainan. Beijing and Tianjin were consistent hot spots from June to October. No cold spots with low mpox attack rates were detected by the Getis-Ord Gi* statistics. Local Moran I statistics identified a high-high (HH) clustering of mpox attack rates in Guangdong, Beijing, and Tianjin. Guangdong province consistently exhibited HH clustering from June to November, while Beijing and Tianjin were identified as HH clusters from July to September. Low-low clusters were mainly located in Inner Mongolia, Xinjiang, Xizang, Qinghai, and Gansu. Ordinary least squares regression models showed that the cumulative mpox attack rates were significantly and positively associated with the proportion of the urban population (t0.05/2,1 = 2.4041; P = .02), per capita gross domestic product (t0.05/2,1 = 2.6955; P = .01), per capita disposable income (t0.05/2,1 = 2.8303; P = .008), per capita consumption expenditure (PCCE; t0.05/2,1 = 2.7452; P = .01), and PCCE for health care (t0.05/2,1 = 2.5924; P = .01).

The geographically weighted regression models indicated a positive association and spatial heterogeneity between cumulative mpox attack rates and the proportion of the urban population, per capita gross domestic product, per capita disposable income, and PCCE, with high R² values in north and northeast China.

Conclusions: Hot spots and HH clustering of mpox attack rates identified by local spatial autocorrelation analysis should be considered key areas for precision prevention and control of mpox. Specifically, Guangdong, Beijing, and Tianjin provinces should be prioritized for mpox prevention and control. These findings provide geographically precise and visualized evidence to assist in identifying key areas for targeted prevention and control.
Introduction

Mpox is a zoonotic viral disease caused by the mpox virus, primarily circulating in animals but also transmissible to humans [1,2]. The main reservoir of the mpox virus is believed to be rodents and primates such as squirrels, kangaroos, dormice, monkeys, and apes [3,4]. Mpox can be transmitted to humans through direct contact with infected animals, their bodily fluids, or contaminated materials [4]. While human-to-human transmission of mpox is less common, it can occur through respiratory droplets, close physical contact, or contact with skin lesions or bodily fluids of an infected individual [4]. An ongoing outbreak of mpox since January 1, 2022, has primarily affected men who have sex with men in countries outside of West and Central Africa, with cases reported to the World Health Organization (WHO) [5].

Historically, mpox has been mainly observed in Central and West Africa, with mortality rates ranging from 1% to 10% [6,7]. Since 2022, person-to-person transmission of mpox has been significantly increasing, thereby resulting in its widespread occurrence in other countries [8-10]. As of the end of February 2024, 94,707 confirmed cases of mpox, including 181 deaths, have been reported to the WHO from 117 countries and regions worldwide since January 1, 2022 [5]. In September 2022, the first imported case of mpox was reported in Chongqing, China. In June 2023, a local mpox epidemic occurred in China. As of November 30, 2023, 1611 confirmed cases of mpox have been reported from 30 (96.8%) provinces in mainland China, except Xizang, which included 1610 local cases [11].

The WHO declared mpox an international public health emergency on July 23, 2022 [12], which was reaffirmed on February 15, 2023 [13]. A public health emergency of international concern is described in the International Health Regulations (2005) as an extraordinary event that poses a public health risk to other countries through the international spread of a disease, potentially requiring a coordinated global response [14]. On September 15, 2023, mpox was classified as a category B infectious disease for management by the National Health Commission of China, effective September 20, 2023 [15]. In accordance with the Law of Infectious Disease Prevention and Control of the People’s Republic of China, statutory infectious diseases are identified by their transmission, epidemic intensity, and hazard degree. These diseases are classified into 3 categories, totaling 40 types: 2 types in category A, 27 types in category B, and 11 types in category C. Category A infectious diseases are considered compulsory for management, while category B infectious diseases are strictly regulated [16]. According to the WHO, the confirmation of a single case of mpox in a country is considered an outbreak [5]. Therefore, it is crucial to prioritize the mpox pandemic owing to its potentially disastrous consequences for public health, socioeconomic factors, and overall health care systems. Experiences from previous pandemics, such as SARS, Middle East respiratory syndrome coronavirus, and COVID-19, highlight the roles of health administrators and policymakers to promptly develop comprehensive prevention and control strategies in all countries.

Spatial epidemiological studies have accurately visualized the spatial distribution and clustering of diseases via mapping [17-20]. Using geographically precise and visual evidence, priorities for disease prevention and control could be identified, and their effectiveness could be evaluated [21-24]. Various factors can influence the occurrence and prevalence of infectious diseases: natural factors (climate and geography) and social factors (economy and population density). The existing evidence emphasizes the spatial heterogeneity in mpox distribution, indicating complexity and unevenness in the spatial patterns of mpox cases [25-27]. This highlights the significance of conducting spatial analyses in mpox studies. However, there is a scarcity of reports on spatial analyses of mpox.

The attack rate is a crucial indicator for measuring the frequency and intensity of disease occurrence and for evaluating the effectiveness of disease prevention and control within a short period. Therefore, we conducted a spatial epidemiological analysis to comprehensively and accurately describe and analyze the mpox attack rate. Our analysis aimed to identify spatial distribution characteristics and spatial clustering patterns and to determine the presence of cold and hot spots of spatial clustering. Furthermore, we examined the spatial regression relationship between the mpox attack rate and sociodemographic and socioeconomic factors.

Methods

Study Design

A spatial ecological comparison study was designed to conduct a spatiotemporal epidemiological analysis of mpox attack rates in mainland China at a provincial level, focusing on spatiotemporal distribution characteristics and clustering patterns of the mpox attack rate and its spatial relationship with sociodemographic and socioeconomic factors.

Study Area

This study was conducted in 31 provinces of mainland China, excluding Taiwan, Macao, and Hong Kong (Table S1 in Multimedia Appendix 1). According to the Chinese administrative division, the 31 provinces are divided into 6 regions: north China, northeast China, east China, central south China, southwest China, and northwest China. As of the end of 2022, the estimated population of China was 1411.75 million people. The per capita gross domestic product (PCGDP) of China was US $12,720, and the per capita disposable income (PCDI) was US $5472. These data are sourced from the China Statistical Yearbook (2023) of the National Bureau of Statistics [28]. For a visual representation of the study areas and
administrative divisions, please refer to Figure S1 in Multimedi a Appendix 2.

Study Population
This study analyzed all the locally confirmed cases of mpox in mainland China that occurred between June and November 2023. As of November 30, 2023, a total of 1610 locally confirmed cases of mpox have been reported by the National Health Commission of China [11].

Mpx Attack Rate
The provincial-level crude mpox attack rates per 10 million people were calculated by dividing the total number of confirmed mpox cases in each province by its total population. This rate was multiplied by 10 million to obtain the rate per 10 million people.

Sociodemographic and Socioeconomic Variables
Sociodemographic variables included the proportion of the urban population (PUP), natural population growth rate (NPGR), percentage of the illiterate population older than 15 years (PIP), aging rate older than 65 years (AR), and per capita road area (PCRA) [28]. Socioeconomic variables included the PCGDP, PCDI, per capita consumption expenditure (PCCE), and PCCE for health care (PCCEH) [28].

Spatial Analysis
Spatial description and analysis were conducted using ArcGIS (version 9.0; Environmental Systems Research Institute, Inc), with the province as the spatial analysis unit. Thematic maps were created to visually and intuitively display the spatial distribution of the confirmed cases and attack rates of mpox. Global spatial autocorrelation analysis was conducted to probe the presence of spatial clustering in mpox attack rates at a broader level [22]. This analysis used the global Moran I statistic (−1 ≤ values ≤ +1). Spatial autocorrelation was considered to be present if the P value was below .10, with a test level α of .10. The values of Moran I indicate whether the spatial distribution of the mpox attack rates was random (Moran I=0), dispersed (Moran I<0), or clustered (Moran I>0). However, it is important to note that global spatial autocorrelation analysis does not provide information on specific local locations and patterns of spatial clustering.

To accurately identify the geographic locations and patterns of spatial clustering at the provincial level of mpox attack rates within the study area, a local spatial autocorrelation analysis was conducted. The analysis used Getis-Ord-Gi* and local Moran I statistics, which are commonly used spatial techniques for assessing local spatial autocorrelation [22]. The Getis-Ord Gi* statistics identified 2 types of clustering: hot spots and cold spots. Hot spots are identified by positive \( z \) values, indicating that high values of mpox attack rates are clustered among neighboring provinces, while cold spots are identified by negative \( z \) values, indicating that low values of mpox attack rates are clustered among neighboring provinces. The corresponding \( z \) values for the Getis-Ord Gi* statistic at 90%, 95%, and 99% CIs are ±1.65, ±1.96, and ±2.58, respectively. The local Moran I statistic was used to verify and complement the Getis-Ord Gi* analysis, as it allows the detection of areas where spatial outliers exist. The results of the local Moran I highlighted aspects that may have been overlooked in the Getis-Ord Gi* analysis. Local Moran I examines 4 patterns of spatial clustering: high-high (HH) clustering (positive correlation; high values of mpox attack rates clustered among neighboring provinces), high-low (HL) outlier (negative correlation; provinces with high values of mpox attack rates surrounded by those with low values), low-high (LH) outlier (negative correlation; provinces with low values surrounded by those with high values), and low-low clustering (positive correlation; low values clustered among neighboring provinces).

This study used ordinary least squares (OLS) and geographically weighted regression (GWR) models to examine the spatial regression relationship between the mpox attack rates and sociodemographic and socioeconomic factors. The dependent variable was the cumulative mpox attack rate between June 1 and November 30, 2023, and sociodemographic and socioeconomic factors were considered independent variables. The OLS model was used to estimate the global parameters, whereas the GWR model was used to estimate the local parameters, considering the spatial autocorrelation and spatial heterogeneity of the research factors [29,30]. The GWR model was constructed using the weighted least squares method, with the selection criterion and kernel function playing essential roles. The selection criterion determined the adaptive bandwidth size, while the kernel function calculated the weight matrix. This study used the fixed Gaussian and Akaike information criterion as the kernel function and selection criterion, respectively [30].

Ethical Considerations
This study was approved by the ethical committee of Qiqihar Medical University (approval number [2021] 31). This study adhered to the Helsinki Declaration.

Results
Spatiotemporal Distribution of Confirmed Cases and Attack Rates of Mpox
Between June 1 and November 30, 2023, a total of 1610 locally confirmed cases of mpox were reported in mainland China, resulting in an attack rate of 11.40 per 10 million people. The number of confirmed cases and the attack rate of mpox initially increased and then decreased during the study period, as depicted in Figure 1A-C. The peak period was observed between July and August, which accounted for 61.6% (n=992) of the total cases. In August 2023, mpox reached its highest level with 501 confirmed cases and an attack rate of 3.55 per 10 million people. However, by November, the number of confirmed cases and attack rate reached their lowest points and were even lower than the attack rate in June. There were 80 confirmed cases with an attack rate of 0.57 per 10 million people (Figure 1A).
Between June 1 and November 30, 2023, a total of 30 provinces reported confirmed cases of mpox, with Xizang being the only province without any cases (Table S2 in Multimedia Appendix 1 and Figure S2 in Multimedia Appendix 3). Among these provinces, Guangdong, Beijing, Zhejiang, Sichuan, and Jiangsu had the highest cumulative numbers of confirmed mpox cases, with 327, 245, 176, 132, and 114 cases, respectively (Table S2 in Multimedia Appendix 1 and Figure S2A in Multimedia Appendix 3). Furthermore, the highest cumulative attack rates of mpox were observed in Beijing, Shanghai, Tianjin, Zhejiang, and Guangdong provinces, with rates of 112.18, 35.15, 28.61, 26.76, and 25.84 per 10 million people, respectively (Table S2 in Multimedia Appendix 1 and Figure S2B in Multimedia Appendix 3). The number of regions reporting confirmed mpox cases gradually increased between June and September 2023 (Figure 2A,B). Confirmed mpox cases were reported in 6 provinces in June, whereas in September, the number of regions reporting confirmed mpox cases rose to 28. However, from October to November 2023, there was a gradual decline in the number of regions reporting mpox cases (Figure 2A,B). Beijing
consistently exhibited the highest mpox attack rates (Figure 2B). During July and August 2023, the number of confirmed cases and attack rates of mpox in various regions were particularly elevated, especially in Beijing, Guangdong, Shanghai, Sichuan, Tianjin, and Zhejiang provinces, when compared with other regions and months (Figure 2A,B).

**Figure 2.** Spatiotemporal distribution of confirmed cases (A) and attack rates (B) of mpox from June to November 2023. A higher resolution version of this figure is available in Multimedia Appendix 4.

### Global Spatial Autocorrelation Analysis of Mpox Attack Rates

A global spatial autocorrelation analysis was conducted to examine the spatial patterns of mpox attack rates at the provincial level in mainland China (Table S3 in Multimedia Appendix 1). The results showed that in June (Moran $I=-0.0057$, $P=0.50$), October (Moran $I=0.0969$, $P=0.17$), and November (Moran $I=0.0510$, $P=0.39$), the attack rates of mpox were randomly distributed and not statistically significant. However, there was a significant global spatial autocorrelation in July (Moran $I=0.0938$, $P=0.08$), August (Moran $I=0.1276$, $P=0.08$),
and September (Moran I = 0.0934, P = 0.07). This indicates that, during this period, the attack rates of mpox exhibited a clustered pattern and positive spatial autocorrelation (Figure 3).

**Figure 3.** Results of Moran I statistic for global spatial autocorrelation analysis of mpox attack rates.

Local Spatial Autocorrelation Analysis of Mpox Attack Rates

The Getis-Ord Gi* statistics revealed that Beijing and Tianjin consistently exhibited high mpox attack rates from June to October, categorizing them as hot spots. Shanghai was identified as a hot spot in August and October. Jiangsu and Hainan were identified as hot spots in October and November, respectively. No cold spots with low mpox attack rates were detected using the Getis-Ord Gi* statistics (Figure 4A).
According to the results of the local Moran I statistics, Guangdong province consistently exhibited HH clustering in terms of spatial patterns of mpox attack rates from June to November. Beijing and Tianjin were identified as HH clusters from July to September. Beijing was identified as an HL outlier for June, October, and November. Sichuan was determined to be an HL outlier in July, August, and October. Tianjin was identified as an LH outlier in June and November. Jiangsu was identified as an LH outlier in October. Low-low clusters were mainly located in Inner Mongolia, Xinjiang, Xizang, Qinghai, and Gansu (Figure 4B). This implies that areas with low mpox attack rates are clustered in neighboring regions.

**Spatial Regression Analysis Between Mpox Attack Rates and Sociodemographic and Socioeconomic Factors**

The detailed values of the cumulative mpox attack rates and explanatory variables can be found in Tables S4 and S5 in Multimedia Appendix 1 and Figure S3 in Multimedia Appendix 6. The results of the Pearson correlation analysis indicated that
the cumulative mpox attack rates were not significantly correlated with the NPGR, PIP, AR, and PCRA but were significantly and positively associated with the PUP, PCGDP, PCDI, PCCE, and PCCEH. However, there were significant correlations among the PUP, PCGDP, PCDI, PCCE, and PCCEH (Figure S4 in Multimedia Appendix 7). To address multicollinearity, separate OLS and GWR regression models were conducted for the dependent variable and each independent variable. However, multivariate OLS and GWR regression models have not yet been developed.

First, we used OLS regression models to explore the spatial regression relationship between the cumulative mpox attack rates from June 1 to November 30, 2023, as the dependent variable and sociodemographic and socioeconomic factors as independent variables. The cumulative mpox attack rates were not significantly correlated with the NPGR, PIP, AR, or PCRA; while they were significantly and positively associated with the PUP ($t_{0.05/2,1} = 2.4041$, $P = .02$), PCGDP ($t_{0.05/2,1} = 2.6955$, $P = .01$), PCDI ($t_{0.05/2,1} = 2.8303$, $P = .008$), PCCE ($t_{0.05/2,1} = 2.7452$, $P = .01$), and PCCEH ($t_{0.05/2,1} = 2.5924$, $P = .01$). The $R^2$ values of the OLS models for the PUP, PCGDP, PCDI, PCCE, and PCCEH were 0.4190, 0.5743, 0.6469, 0.5583, and 0.5400, respectively; indicating that these 5 variables can explain the 41.90%, 57.43%, 64.69%, 55.83%, and 54% variance of cumulative mpox attack rates at the provincial level (Table S6 in Multimedia Appendix 1).

The GWR regression models were created using only the statistically significant independent variables identified in the OLS regression models, along with the cumulative mpox attack rates. The coefficients and local $R^2$ values of the explanatory variables in the GWR models are presented in Figure 5 and Tables S7-S10 in Multimedia Appendix 1. Generally, the Akaike Information Criterion values of the GWR models were significantly lower than those of the OLS model, indicating that the GWR had a high explanatory power and better fitting ability (Tables S6 and S7 in Multimedia Appendix 1). The GWR model indicated a positive association and spatial heterogeneity between cumulative mpox attack rates and the PUP, PCGDP, PCDI, and PCCE (Figure 5), with high $R^2$ values in north and northeast China.
Figure 5. Spatial distribution of the coefficients (A) and local $R^2$ (B) values of the explanatory variables in the geographically weighted regression model. PCCE: per capita consumption expenditure; PCCEH: PCCE for health care; PCDI: per capita disposable income; PCGDP: per capita gross domestic product; PUP: proportion of the urban population. A higher resolution version of this figure is available in Multimedia Appendix 8.

Discussion

Principal Findings

This study is the first in mainland China to conduct a spatiotemporal epidemiological analysis to comprehensively identify the spatiotemporal characteristics and clustering patterns of confirmed cases and attack rates of mpox at the provincial level. The findings of this study are of great significance for improving the accuracy and effectiveness of mpox prevention and control.

Since June 2023, local cases of mpox have been reported in 30 of 31 provinces in mainland China, excluding Xizang [11]. In June, 6 provinces reported cases, which rose to 28 regions by September. However, from October to November 2023, there
was a gradual decline in the number of regions reporting mpox cases. As of November 30, 2023, a total of 1610 locally confirmed cases of mpox have been reported [11]. The attack rate of mpox is 11.40 per 10 million. The number of confirmed cases and attack rate of mpox exhibited a pattern of increase from June to August, followed by a decrease from September to November. In November, both the number of confirmed cases and the attack rate reached their lowest points, even lower than the level observed in June. These findings highlight the importance of the implementation of preventive and control measures. On June 10, 2022, the National Health Commission of China formulated the “Diagnosis and Treatment Guidelines for Monkeypox (2022 edition)” [3]. Subsequently, on June 27, 2022, the National Health Commission of China released the Technical Guidelines for Monkeypox Prevention and Control (2022 edition) [4]. The guidelines highlight 4 key areas for action: first, enhancing the evaluation of the mpox pandemic situation; second, intensifying quarantine and monitoring measures for mpox; third, enhancing the professional training of health care workers; and finally, ensuring the availability of diagnostic reagents, therapeutic drugs, and vaccines for mpox [31]. On July 26, 2023, the National Disease Control and Prevention Administration developed the Monkeypox Prevention and Control Plan [32]. Subsequently, on August 2, 2023, the AIDS Prevention and Control Center of China created the Core Information of Monkeypox Prevention and Treatment for Key Populations, aiming to enhance awareness and prevention of mpox [33]. On September 15, 2023, mpox was classified as a category B infectious disease for management by the National Health Commission of China, effective September 20, 2023 [14]. These guidelines provide guidance on the diagnosis, treatment, prevention, and control of mpox in various regions, which also emphasizes the need to strengthen prevention and control measures for mpox.

Although the number of regions reporting mpox cases, confirmed cases, and attack rates declined from October to November 2023, the first confirmed case of mpox was reported in Gansu Province in November 2023. Since January 1, 2022, China, including Taiwan, Macau, and Hong Kong, has ranked 10th globally in the number of confirmed cases reported by the WHO [5]. The number of confirmed mpox cases in China (n=2031) was lower than that in the United States (n=31,800), Brazil (n=10,967), Spain (n=7898), France (n=4195), Colombia (n=4090), Mexico (n=4081), the United Kingdom (n=3892), Germany (n=3816), and Peru (n=3812). Therefore, mpox has emerged as a major public health emergency in mainland China, Germany (n=3816), and Peru (n=3812). Therefore, mpox has emerged as a major public health emergency in mainland China, especially Beijing and Tianjin provinces, indicating that areas with high mpox attack rates were clustered among neighboring provinces. Among these provinces, Guangdong consistently exhibited HH clustering in terms of spatial patterns of mpox attack rates from June to November, while Beijing and Tianjin were identified as HH clusters from July to September. In June, October, and November, Beijing was identified as an HL outlier, suggesting that the areas with high mpox attack rates in Beijing were surrounded by other areas with low values. Similarly, Tianjin was recognized as an LH outlier in June and November, indicating that the areas with low mpox attack rates in Tianjin were surrounded by other areas with high values. This finding aligns with the identification of Beijing as an HL outlier in June and November, emphasizing the consistently high mpox attack rates in Beijing. Further research is necessary to comprehend the reasons behind these high attack rates in Beijing, with possible reasons including Beijing’s status as the capital and most densely populated area in China, along with the presence of a large floating population that may contribute to virus transmission. Therefore, the results of the local Moran I highlighted an HH clustering in Guangdong Province that had been overlooked in the Getis-Ord Gi* analysis and verified the result that Beijing and Tianjin were consistent hot spots identified by the Getis-Ord Gi* analysis. Additionally, Sichuan was identified as an HL outlier in July, August, and October, whereas Jiangsu was identified as an LH outlier in October, suggesting that the surrounding Shanghai and Zhejiang provinces had high mpox attack rates. Therefore, all the hot spots and HH clustering of mpox attack rates identified through local spatial autocorrelation analysis should be considered key areas for precision prevention and control of mpox. Specifically, the Guangdong, Beijing, and Tianjin provinces, which consistently exhibited high-value clustering, should be given high priority for mpox prevention and control. These findings provide geographically precise and visual evidence to assist in identifying key areas for the targeted prevention and control of mpox.

The GWR model was used to analyze the spatially varying relationships between independent and dependent variables while considering spatial autocorrelation [34]. It used the weighted least squares method to estimate varying parameters locally, allowing for varied spatial relationships across different geographical regions. This localized approach facilitated the identification of spatial heterogeneity, revealing trends and distribution patterns of the spatial relationship between independent and dependent variables within specific spatial extents [30,34]. In contrast, the OLS model estimated parameters using the OLS method, assuming that spatial relationships in the model were constant across space [30]. However, this global parameter estimation with constant parameters overlooks spatial heterogeneity in the data and fails to capture spatial variation trends [30]. In this study, the GWR model revealed a positive association and spatial heterogeneity between the cumulative mpox attack rates and the PUP, PCGDP, PCDI, and PCCE. This finding aligns with the results obtained from the application of
the Getis-Ord Gi* statistics, which identified hot spots in the Beijing, Tianjin, Shanghai, Jiangsu, and Hainan provinces. Furthermore, the results of the Local Moran I statistics also supported this finding, indicating HH clustering in Guangdong, Beijing, and Tianjin and HL outliers in Sichuan. These areas, which were identified as key regions for the precise prevention and control of mpox through local autocorrelation analysis, are characterized by robust economic development [28].

Limitations
This study had 2 limitations. First, the spatial analysis was conducted at the provincial level, and further studies should consider smaller geographic units for more accurate evidence. Second, this study focused solely on the relationship between cumulative mpox attack rates and sociodemographic and socioeconomic factors; future studies should also consider natural and environmental risk factors.

Conclusions
A total of 1610 locally confirmed mpox cases were reported in 30 provinces in mainland China from June to November 2023, resulting in an attack rate of 11.40 per 10 million people. The hot spots and HH clustering of mpox attack rates should be considered as key areas for precision prevention and control of mpox. Specifically, Guangdong, Beijing, and Tianjin should be given high priority for mpox prevention and control. These findings provide geographically precise and visual evidence for identifying key areas for targeted prevention and control.

Acknowledgments
The authors sincerely thank the workers at the Centers for Disease Prevention and Control for their efforts in collecting and managing data on confirmed cases of Mpox. The authors express their deepest gratitude to Editage (Cactus Communications Private Limited) for their assistance with language editing. This study was supported by the Doctorate Research Foundation of the Qiqihar Medical University (grant QMSI2023B-16). Generative artificial intelligence had no role in the study design and any portion of the manuscript writing.

Data Availability
The data sets used or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions
YJ and JG did the conceptualization, formal analysis, funding acquisition, and methodology, wrote the original draft, and reviewed and edited the draft. LQ, LD, YQ, LB, and JG did the conceptualization, supervision, and project administration. XC, SM, YH, ZX, and JC did the data curation, software, and formal analysis. All authors read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary Tables S1-S10.
[DOCX File, 56 KB - publichealth_v10i1e57807_app1.docx ]

Multimedia Appendix 2
Spatial distribution of the study areas (A) and administrative divisions (B).
[PNG File, 318 KB - publichealth_v10i1e57807_app2.png ]

Multimedia Appendix 3
Spatial distribution of the cumulative confirmed cases (A) and attack rates (B) of mpox.
[PNG File, 2007 KB - publichealth_v10i1e57807_app3.png ]

Multimedia Appendix 4
Higher resolution version of Figure 2.
[PDF File (Adobe PDF File), 6887 KB - publichealth_v10i1e57807_app4.pdf ]

Multimedia Appendix 5
Higher resolution version of Figure 4.
[PDF File (Adobe PDF File), 5672 KB - publichealth_v10i1e57807_app5.pdf ]

Multimedia Appendix 6
Spatial distribution of the provincial-level values of explanatory variables.

Multimedia Appendix 7
Correlation analysis between the cumulative mpox attack rates and explanatory variables.

Multimedia Appendix 8
Higher resolution version of Figure 5.

References
   [FREE Full text] [doi: 10.1016/S0140-6736(22)02075-X] [Medline: 36403582]


Moderation Effects of Streetscape Perceptions on the Associations Between Accessibility, Land Use Mix, and Bike-Sharing Use: Cross-Sectional Study

Huagui Guo¹,², PhD; Shuyu Zhang¹, BE; Xinwei Xie¹, BE; Jiang Liu¹*, PhD; Hung Chak Ho³*, PhD

¹²³

*these authors contributed equally

Corresponding Author:
Jiang Liu, PhD

Abstract

Background: Cycling is known to be beneficial for human health. Studies have suggested significant associations of physical activity with macroscale built environments and streetscapes. However, whether good streetscapes can amplify the benefits of a favorable built environment on physical activity remains unknown.

Objective: This study examines whether streetscape perceptions can modify the associations between accessibility, land use mix, and bike-sharing use.

Methods: This cross-sectional study used data from 18,019,266 bike-sharing orders during weekends in Shanghai, China. A 500 × 500 m grid was selected as the analysis unit to allocate data. Bike-sharing use was defined as the number of bike-sharing origins. Street view images and a human-machine adversarial scoring framework were combined to evaluate lively, safety, and wealthy perceptions. Negative binomial regression was developed to examine the independent effects of the three perceptual factors in both the univariate model and fully adjusted model, controlling for population density, average building height, distance to nearest transit, number of bus stations, number of points of interest, distance to the nearest park, and distance to the central business district. The moderation effect was then investigated through the interaction term between streetscape perception and accessibility and land use mix, based on the fully adjusted model. We also tested whether the findings of streetscape moderation effects are robust when examinations are performed at different geographic scales, using a small-sample statistics approach and different operationalizations of land use mix and accessibility.

Results: High levels of lively, safety, and wealthy perceptions were correlated with more bike-sharing activities. There were negative effects for the interactions between the land use Herfindahl–Hirschman index with the lively perception ($\beta=-0.63; P=.01$) and safety perception ($\beta=-0.52; P=.001$). The interaction between the lively perception and road intersection density was positively associated with the number of bike-sharing uses ($\beta=0.43; P=.08$). Among these, the lively perception showed the greatest independent effect ($\beta=1.29; P<.001$), followed by the safety perception ($\beta=1.22; P=.001$) and wealthy perception ($\beta=0.72; P=.001$). The findings were robust in the three sensitivity analyses.

Conclusions: A safer and livelier streetscape can enhance the benefits of land use mix in promoting bike-sharing use, with a safer streetscape also intensifying the effect of accessibility. Interventions focused on streetscape perceptions can encourage cycling behavior and enhance the benefits of accessibility and land use mix. This study also contributes to the literature on potential moderators of built environment healthy behavior associations from the perspective of microscale environmental perceptions.

(JMIR Public Health Surveill 2024;10:e58761) doi:10.2196/58761

KEYWORDS

built environment; streetscape perceptions; bike-sharing use; cycling; moderation effect; China

Introduction

Several studies have demonstrated that a favorable built environment can promote walking and bike-sharing use, particularly in Chinese cities [1-3]. However, findings on the effects of built environments are usually mixed [4-6]. One potential explanation is that there may be interactions between built environments at different levels, including the moderation effects of streetscape perceptions [7,8]. An in-depth understanding of streetscape perceptions as moderators is not only an important methodological issue for environmental health research [9] but also informative for creating effective interventions in built environments. Despite some efforts
[10,11], however, it remains unclear whether and to what extent good perceptions of streetscapes can enhance the benefits of favorable built environments.

The ecological model of physical activity provides a theoretical foundation for streetscape moderation effects. According to this model, one’s cycling behavior is determined by factors across multiple levels [9]. Such behavior is shaped by interactions among indicators at different levels [9]. Moreover, understanding the benefits of streetscapes in promoting health is important. Microscale streetscapes provide numerous benefits to people’s physical activity [11-13]. Additionally, the built environment at the neighborhood level has essentially been established and fixed over time, and streets have become a focus of high-quality urban development in the era of inventory planning, particularly in Chinese cities.

However, little attention has been given to the moderation effects of streetscapes [8,14,15]. Furthermore, most studies examining environmental perception as a moderator focus on perceived safety [16-18], while little attention has been given to other perceived aspects. Moreover, streetscapes as potential moderators are usually examined from the objective side [8,10] and are less understood from a perceptual perspective. It has been reported that residents’ safety perceptions are more likely to be correlated with physical activity [8,19], especially for those who do not perceive threats from neighborhoods with high crime rates. However, insufficient attention has been given to streetscapes in studies exploring environmental moderation from a perception perspective [8].

To fill the aforementioned gaps, this study aims to address whether streetscape perceptions can modify the effects of accessibility and land use mix on bike-sharing use. To achieve this, we used data from 18,019,266 GPS-based bike-sharing orders and conducted a cross-sectional study in Shanghai, China. This study contributes to the literature on the potential moderation effects of urban environments from the perspective of microscale environmental perception [8,18,20]. The findings are derived from Chinese cities, where urban forms and the leading perception factors as moderators differ significantly from those of Western countries [10,21], which all advance the development of ecological models of health behavior [9]. This cross-sectional study, with its large data volume and spatial coverage, suggests that urban planning initiatives aiming to encourage healthy behaviors among residents should consider microscale streetscapes, which can promote physical activities and enhance the benefits of accessibility and land use mix.

**Methods**

**Research Area**

We examined the moderation effects of streetscape perceptions in central Shanghai, China. The research area was delineated by the Outer Ring Highway of Shanghai (Figure 1), mainly because of the availability of street view images and bike-sharing use data. As one of the four first-tier cities in China, Shanghai is among the largest cities in the world. As of 2017, there were 1.5 million dockless bikes in Shanghai. The terrain is flat, and the climate is relatively favorable, which provides conducive conditions for people’s cycling activities.
Data Source and Research Design

This is a cross-sectional study in nature, with grids as analysis units to allocate data. Bike-sharing use data were provided by the Mobike Technology Co., Ltd, spanning from August 26 to September 8, 2018. Originally, the number of bike-sharing use records was about 19 billion. Each record included information on trip ID, start time and location, and end time and location. As in many studies [2,10], records on rainy days (eg, September 6, 2018) were not included. Additionally, we excluded records with abnormal durations or lengths, retaining records with a trip duration from 2 minutes to 1 hour. After excluding trips with locations out of our research area, the final number of bike-sharing use records in this study was 18,019,266.

Data on street view images were obtained from the Baidu application programming interface [22], with the sampling points at 50-m intervals along the road network. Each sampling point contains street view images with four headings. Data used to measure macroscale built environments included population data derived from EasyGO data and road networks, buildings, and points of interest (POIs) acquired from Baidu Maps [22].

Deriving Bike-Sharing Use

Bike-sharing use is defined as the number of bike-sharing origins on weekends in each grid. Similar to studies examining environmental effects on bike-sharing use [10,23], a 500 × 500 m grid was selected as the analysis unit to allocate each trip, primarily because the size of the grids selected was usually...
considered to be the main activity space in Chinese neighborhoods [21,24]. Table S1 in Multimedia Appendix 1 presents the summary statistics of bike-sharing use, streetscape perceptions, and macroscale built environments. Figure 2A exhibits the spatial distributions of bike-sharing origins on weekends.

Figure 2. Spatial distributions of the number of bike-sharing uses and assessed streetscape perceptions during 2018 in Shanghai, China.

Assessing Streetscape Perceptions
Streetscape perceptions are three positive perceptual indicators (ie, lively, safety, and wealthy) [25]. They were assessed according to a human-machine adversarial scoring framework developed by Yao et al [25]. Briefly, volunteers scored the perceptions based on the displayed street view image, with the scored values ranging from 0 to 100. Subsequently, a random forest model was used to fit the association of volunteer scorings with visual scenic features extracted by the segmentation approach of the fully convolutional neural network–8. Once the first 50 photos were scored by a volunteer, a random forest set was created to fit the perceived scores in the scoring software. When the volunteer rated the subsequent photos, the software provided the recommended scores based on the rules learned from the volunteer’s previous rating. In this process, an iterative feedback module was used to automatically adjust the recommended scores according to the user’s scoring behaviors. When the difference between the recommended and scored values was lower than 5 points, the scoring procedure stopped and a human-machine adversarial scoring module was created.
As reported, errors in the estimated perceptions using the scoring framework were less than 10% [25]. In this situation, the created scoring module was used to assess the three perceptual indicators and then obtain the perceptual scores for each street view image. The perceptual scores for each sampling point were then calculated by averaging the perceptual scores of each of the four images. Finally, perception scores for each grid were calculated by averaging the perceptual scores of sampling points located in the grid. Figures 2 (B-D) and 3 show the spatial distributions of the three perception indicators assessed and their examples.

**Figure 3.** Examples of the assessment for lively, safety, and wealthy perceptions during 2018 in Shanghai, China.

---

**Measuring Macroscale Built Environments**

Macroscale built environments were measured in terms of the 5Ds framework [1]. Population and average building height were used to measure density. The land use Herfindahl-Hirschman index (HHI) was used to assess diversity, with a high value indicating a less mixed degree of different types of land use [26]. Destination accessibility was measured using the number of POIs and distance to the nearest park. Distance to transit was assessed in terms of distance to the nearest metro station and the number of bus stops. The design was measured through the number of road intersections. Distance to the central business district (CBD) was operationalized as the distance between the centroid of a grid and the CBD (km) to control for the effect of location. Figure
4 shows the spatial distributions of macroscale built environment factors.

Figure 4. Spatial distributions of some built environment elements in 2018 at macroscale in Shanghai, China. POI: point of interest.

**Statistical Analysis**

The negative binomial regression model (NBRM) was used to examine the independent effects of streetscape perceptions. This was performed in the univariate model and then the fully adjusted model, controlling for population density, average building height, distance to nearest transit, number of bus stations, number of POIs, distance to the nearest park, and distance to CBD. The NBRM was selected primarily because the outcome variable (ie, the number of bike-sharing origins) was in number format and discrete. Moreover, the variance of the outcome variable is greater than its mean value. Hence, as in many studies [27,28], the NBRM was chosen for the examination.

Subsequently, streetscape moderation effects were examined through the interaction term between macroscale built environment factors and streetscape perceptions using the fully adjusted NBRM. Factors at the macroscale level included land use mix and accessibility, while the three perception indicators were lively, safety, and wealthy. Consequently, 6 total models were produced, with each model examining different combinations of these factors. To mitigate multicollinearity and facilitate comparison between independent variables, independent factors were all standardized into z scores.
Finally, we conducted three sensitivity analyses. Recognizing the potential influence of an uncertain geographic context problem [29], we examined whether the results of streetscape moderation roles are sensitive when investigations are performed at different geographic scales. Consistent with previous studies [3,30], a 1000 × 1000 m grid was selected as the analysis unit. Furthermore, apart from the big data analysis, we further tested whether there were moderation effects of streetscapes using a classic (small sample) statistics approach. Similar to the sampling rate of prior studies [10,31], 20% of the total grids were randomly chosen for further examination. Moreover, we tested streetscape moderation roles when land use mix and accessibility were operationalized differently. Road length and land use HHI calculated based on area of interest data were used as further proxies.

Ethical Considerations

This study was approved by the Human and Artefacts Ethics Sub-Committee at the City University of Hong Kong (reference 27609824). The bike-sharing record data were anonymous and deidentified, including information on trip ID, start time and location, and end time and location, in this study.

Results

Effects of Streetscape Perceptions

In the univariate model, each of the three positive perceptual factors was positively associated with the number of bike-sharing uses on weekends (Table S2 in Multimedia Appendix 1). Among this, the lively perception showed the greatest effect ($\beta$=5.50; $P$<.001), followed by the safety perception ($\beta$=5.38; $P$<.001) and wealthy perception ($\beta$=4.45; $P$<.001). A similar pattern of results was observed in the fully controlled model (Table 1). In particular, the most influential effect was found for the lively perception ($\beta$=1.45; $P$<.001), followed by the safety perception ($\beta$=1.19; $P$<.001) and wealthy perception ($\beta$=0.79; $P$<.001).

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>For lively</th>
<th>For safety</th>
<th>For wealthy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception</td>
<td>0.16 (0.02)</td>
<td>0.12 (0.02)</td>
<td>0.08 (0.02)</td>
</tr>
<tr>
<td>Population</td>
<td>0.58 (0.03)</td>
<td>0.57 (0.03)</td>
<td>0.59 (0.03)</td>
</tr>
<tr>
<td>Average building height</td>
<td>0.08 (0.01)</td>
<td>0.09 (0.01)</td>
<td>0.09 (0.01)</td>
</tr>
<tr>
<td>Land use mix</td>
<td>–0.21 (0.02)</td>
<td>–0.22 (0.02)</td>
<td>–0.22 (0.02)</td>
</tr>
<tr>
<td>Number of road intersections</td>
<td>0.05 (0.02)</td>
<td>0.05 (0.02)</td>
<td>0.05 (0.02)</td>
</tr>
<tr>
<td>Distance to nearest transit</td>
<td>–0.30 (0.03)</td>
<td>–0.30 (0.02)</td>
<td>–0.30 (0.03)</td>
</tr>
<tr>
<td>Number of bus stations</td>
<td>0.12 (0.01)</td>
<td>0.13 (0.01)</td>
<td>0.12 (0.01)</td>
</tr>
<tr>
<td>Number of POIs a</td>
<td>0.11 (0.02)</td>
<td>0.11 (0.02)</td>
<td>0.12 (0.02)</td>
</tr>
<tr>
<td>Distance to nearest parks</td>
<td>–0.21 (0.02)</td>
<td>–0.20 (0.02)</td>
<td>–0.22 (0.02)</td>
</tr>
<tr>
<td>Distance to CBD b</td>
<td>–0.17 (0.03)</td>
<td>–0.17 (0.03)</td>
<td>–0.17 (0.03)</td>
</tr>
</tbody>
</table>

aPOI: point of interest.
bCBD: central business district.

Moderation Effects of Streetscape Perceptions

All streetscape perceptions positively moderated the association between the land use HHI and the number of bike-sharing uses on weekends (Table 2). Specifically, streetscape perceptions and the land use HHI were significantly associated with the number of bike-sharing uses on weekends; among the three perceptual indicators, the lively perception showed the greatest effect. Regarding the moderation effect, the effect of the interaction between the land use HHI and each of the streetscape perceptions was smaller than that of each of the two factors; the lively perception presented the strongest moderation role on the effect of the land use HHI, followed by the wealthy perception and safety perception. Specifically, the effect of the interaction between the lively perception and the land use HHI was –0.63 ($P$=.01), greater than those of interactions between the land use HHI with the wealthy and safety perceptions at –0.60 ($P$=.008) and –0.52 ($P$=.001), respectively. This means that livelier, safer, and wealthier street environments can enhance the benefits of land use mix to promote cycling behavior.

There were positive interactions in the number of road intersections with the lively and wealthy perceptions (Table 3). The effects of the number of road intersections, both of the two perceptions, and their interactions were all positive. Different from that of land use mix, the effect of the number of road intersections was close to the impact of the interaction with the
lively and wealthy perceptions. With respect to the moderation role, the lively perception showed the largest moderation effect, although no significant moderation effect was observed for the safety perception. This means that good lively and wealthy perceptions of streetscapes are likely to magnify the role of accessibility in encouraging cycling activities.

Table. Moderation effects of lively, safety, and wealthy perceptions on the association between land use mix and bike-sharing use on weekends in 2018 in Shanghai.

| Independent variables | For lively | | For safety | | For wealthy | |
|-----------------------|------------|----------------|------------|----------------|-------|----------------|-------|
|                       | β (SD) | P values | β (SD) | P values | β (SD) | P values |
| Perception            | 0.13 (0.02) | <.001 | 0.12 (0.02) | <.001 | 0.07 (0.02) | <.001 |
| Population            | 0.57 (0.03) | <.001 | 0.57 (0.03) | <.001 | 0.58 (0.03) | <.001 |
| Average building height | 0.08 (0.01) | <.001 | 0.09 (0.01) | <.001 | 0.08 (0.01) | <.001 |
| Land use mix          | –0.25 (0.03) | <.001 | –0.25 (0.02) | <.001 | –0.25 (0.02) | <.001 |
| Number of road intersections | 0.04 (0.02) | .01 | 0.05 (0.02) | <.001 | 0.05 (0.02) | <.001 |
| Distance to nearest transit | –0.29 (0.03) | <.001 | –0.30 (0.02) | <.001 | –0.30 (0.03) | <.001 |
| Number of bus stations | 0.12 (0.01) | <.001 | 0.12 (0.01) | <.001 | 0.12 (0.01) | <.001 |
| Number of POIs a      | 0.12 (0.02) | <.001 | 0.11 (0.02) | <.001 | 0.13 (0.02) | <.001 |
| Distance to nearest parks | –0.16 (0.03) | <.001 | –0.17 (0.03) | <.001 | –0.17 (0.03) | <.001 |
| Distance to CBD b     | –0.24 (0.03) | <.001 | –0.23 (0.02) | <.001 | –0.23 (0.02) | <.001 |
| Land use mix × perception c | –0.63 (0.03) | .01 | –0.52 (0.02) | <.001 | –0.60 (0.02) | .01 |

aPOI: point of interest.
bCBD: central business district.
cFor value = original value × 10.
Table. Moderation effects of lively, safety, and wealthy perceptions on the associations between accessibility and bike-sharing use on weekends in 2018 in Shanghai, China.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>For lively</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>For safety</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>For wealthy</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$ (SD)</td>
<td>$P$ values</td>
<td>$\beta$ (SD)</td>
<td>$P$ values</td>
<td>$\beta$ (SD)</td>
<td>$P$ values</td>
<td>$\beta$ (SD)</td>
<td>$P$ values</td>
<td>$\beta$ (SD)</td>
<td>$P$ values</td>
<td>$\beta$ (SD)</td>
<td>$P$ values</td>
<td>$\beta$ (SD)</td>
<td>$P$ values</td>
<td>$\beta$ (SD)</td>
<td>$P$ values</td>
</tr>
<tr>
<td>Perception</td>
<td>0.14 (0.02)</td>
<td>&lt;.001</td>
<td>0.12 (0.02)</td>
<td>&lt;.001</td>
<td>0.08 (0.02)</td>
<td>&lt;.001</td>
<td>0.08 (0.02)</td>
<td>&lt;.001</td>
<td>0.08 (0.02)</td>
<td>&lt;.001</td>
<td>0.08 (0.02)</td>
<td>&lt;.001</td>
<td>0.08 (0.02)</td>
<td>&lt;.001</td>
<td>0.08 (0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Population</td>
<td>0.57 (0.03)</td>
<td>&lt;.001</td>
<td>0.57 (0.03)</td>
<td>&lt;.001</td>
<td>0.57 (0.03)</td>
<td>&lt;.001</td>
<td>0.57 (0.03)</td>
<td>&lt;.001</td>
<td>0.57 (0.03)</td>
<td>&lt;.001</td>
<td>0.57 (0.03)</td>
<td>&lt;.001</td>
<td>0.57 (0.03)</td>
<td>&lt;.001</td>
<td>0.57 (0.03)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Average building height</td>
<td>0.08 (0.01)</td>
<td>&lt;.001</td>
<td>0.09 (0.01)</td>
<td>&lt;.001</td>
<td>0.08 (0.01)</td>
<td>&lt;.001</td>
<td>0.08 (0.01)</td>
<td>&lt;.001</td>
<td>0.08 (0.01)</td>
<td>&lt;.001</td>
<td>0.08 (0.01)</td>
<td>&lt;.001</td>
<td>0.08 (0.01)</td>
<td>&lt;.001</td>
<td>0.08 (0.01)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Land use mix</td>
<td>−0.21 (0.02)</td>
<td>&lt;.001</td>
<td>−0.22 (0.02)</td>
<td>&lt;.001</td>
<td>−0.22 (0.02)</td>
<td>&lt;.001</td>
<td>−0.22 (0.02)</td>
<td>&lt;.001</td>
<td>−0.22 (0.02)</td>
<td>&lt;.001</td>
<td>−0.22 (0.02)</td>
<td>&lt;.001</td>
<td>−0.22 (0.02)</td>
<td>&lt;.001</td>
<td>−0.22 (0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of road intersections</td>
<td>0.05 (0.01)</td>
<td>.001</td>
<td>0.05 (0.01)</td>
<td>&lt;.001</td>
<td>0.06 (0.02)</td>
<td>&lt;.001</td>
<td>0.06 (0.02)</td>
<td>&lt;.001</td>
<td>0.06 (0.02)</td>
<td>&lt;.001</td>
<td>0.06 (0.02)</td>
<td>&lt;.001</td>
<td>0.06 (0.02)</td>
<td>&lt;.001</td>
<td>0.06 (0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Distance to nearest transit</td>
<td>−0.29 (0.03)</td>
<td>&lt;.001</td>
<td>−0.30 (0.02)</td>
<td>&lt;.001</td>
<td>−0.30 (0.03)</td>
<td>&lt;.001</td>
<td>−0.30 (0.03)</td>
<td>&lt;.001</td>
<td>−0.30 (0.03)</td>
<td>&lt;.001</td>
<td>−0.30 (0.03)</td>
<td>&lt;.001</td>
<td>−0.30 (0.03)</td>
<td>&lt;.001</td>
<td>−0.30 (0.03)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of bus stations</td>
<td>0.18 (0.01)</td>
<td>&lt;.001</td>
<td>0.12 (0.01)</td>
<td>&lt;.001</td>
<td>0.12 (0.01)</td>
<td>&lt;.001</td>
<td>0.12 (0.01)</td>
<td>&lt;.001</td>
<td>0.12 (0.01)</td>
<td>&lt;.001</td>
<td>0.12 (0.01)</td>
<td>&lt;.001</td>
<td>0.12 (0.01)</td>
<td>&lt;.001</td>
<td>0.12 (0.01)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of POIs</td>
<td>0.11 (0.02)</td>
<td>&lt;.001</td>
<td>0.11 (0.02)</td>
<td>&lt;.001</td>
<td>0.13 (0.02)</td>
<td>&lt;.001</td>
<td>0.13 (0.02)</td>
<td>&lt;.001</td>
<td>0.13 (0.02)</td>
<td>&lt;.001</td>
<td>0.13 (0.02)</td>
<td>&lt;.001</td>
<td>0.13 (0.02)</td>
<td>&lt;.001</td>
<td>0.13 (0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Distance to nearest park</td>
<td>−0.17 (0.03)</td>
<td>&lt;.001</td>
<td>−0.17 (0.03)</td>
<td>&lt;.001</td>
<td>−0.17 (0.03)</td>
<td>&lt;.001</td>
<td>−0.17 (0.03)</td>
<td>&lt;.001</td>
<td>−0.17 (0.03)</td>
<td>&lt;.001</td>
<td>−0.17 (0.03)</td>
<td>&lt;.001</td>
<td>−0.17 (0.03)</td>
<td>&lt;.001</td>
<td>−0.17 (0.03)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Distance to CBD</td>
<td>−0.26 (0.03)</td>
<td>&lt;.001</td>
<td>−0.21 (0.02)</td>
<td>&lt;.001</td>
<td>−0.21 (0.02)</td>
<td>&lt;.001</td>
<td>−0.21 (0.02)</td>
<td>&lt;.001</td>
<td>−0.21 (0.02)</td>
<td>&lt;.001</td>
<td>−0.21 (0.02)</td>
<td>&lt;.001</td>
<td>−0.21 (0.02)</td>
<td>&lt;.001</td>
<td>−0.21 (0.02)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of road intersections × perception $^c$</td>
<td>0.43 (0.02)</td>
<td>.08</td>
<td>0.17 (0.02)</td>
<td>.29</td>
<td>0.39 (0.02)</td>
<td>&lt;.001</td>
<td>0.39 (0.02)</td>
<td>&lt;.001</td>
<td>0.39 (0.02)</td>
<td>&lt;.001</td>
<td>0.39 (0.02)</td>
<td>&lt;.001</td>
<td>0.39 (0.02)</td>
<td>&lt;.001</td>
<td>0.39 (0.02)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

$^a$POI: point of interest.

$^b$CBD: central business district.

$^c$For value = original value × 10.

Sensitivity Analyses

As indicated in the sensitivity analysis using a 1000 × 1000 m grid as the analysis unit (Figure 5 A-C), there were significant interactions between land use mix (ie, land use HHI) and both lively and safety perceptions of the streetscape, but not the wealthy perception. Similarly, a moderation role on the impact of the number of road intersections was found for the lively perception but not for the wealthy perspective. Figure S1 G-I in Multimedia Appendix 1 illustrates the sensitivity analysis in terms of a randomly selected small sample. Overall, positive moderation roles persisted for all streetscape perceptions on the effect of land use mix and for the lively perception on the impact of the number of road intersections. Furthermore, among the three perception indicators, the lively perception still played the most substantial moderation role in relation to the impact of the land use HHI.

Figure 5 D-F exhibits the sensitivity analysis using the land use HHI calculated using area of interest data and road length as proxies for land use mix and accessibility, respectively. Generally, lively, wealthy, and safety perceptions of streetscapes continued to exhibit positive moderation on the effect of the land use HHI. In particular, the absolute effect of the interaction between land use mix and lively perception was 0.10 ($P$=.001), accounting for 66.14% and 94.22% of the main effects of the two factors, respectively. With regards to accessibility, operationalized by road length in grids, its interactions with lively and wealthy perceptions were still correlated with the number of bike-sharing uses on weekends.
Figure 5. Sensitivity analyses of streetscape moderation effects examined at different geographic scales (A-C) and using different operationalizations (D-F) in 2018 in Shanghai, China.

Discussion

Principal Findings and Comparisons With Prior Work

We found that streetscape perceptions were positively associated with bike-sharing activities. This is consistent with many prior studies [17,32,33]. In particular, a cross-sectional study conducted in Boston, Massachusetts using running data from the Strava Heatmap suggested the positive effects of wealthy and safety perceptions on the amount of running [13]. The mechanisms linking environmental perceptions and physical activity are highly complex and beyond the scope of this work, highlighting the need for further investigation, especially with evidence from a neuroscience perspective.

However, we observed that among three perceptual factors, the lively perception not only plays the strongest effect on bike-sharing use but also emerges as the most influential moderator, which is inconsistent with findings from Western studies [7,13,20]. Typically, among environmental perceptions such as safety and lively, the former usually plays the strongest moderation role on the impact of macroscale built environments in studies outside of mainland China [8,16,20]. In particular, a US study indicated that compared to perceived pleasure, the safety perception plays a greater moderation role in the impact of accessibility on healthy behaviors [8].

The underlying mechanisms behind the lively perception not only playing the largest independent effect but also showing...
the strongest moderation effect are highly complex. There are two potential explanations. On the one hand, among the environmental perceptions (qualities), safety plays a crucial role in physical activity in many Western cities [7,13,20]. However, the lively perception usually takes precedence in residents’ concerns about healthy behavior in the Chinese context [17,34]. This is also consistent with the findings of this study. Regarding accessibility, the main effect of the lively perception on the number of bike-sharing uses on weekends was 1.39 (P<.001), higher than that of the safety and wealthy perceptions at 1.12 (P<.001) and 0.75 (P<.001), respectively. A similar pattern of results was observed for land use mix, with the greatest main effect for the lively perception (β=1.39; P<.001), followed by the safety perception (β=1.22; P<.001) and wealthy perception (β=0.75; P<.001).

On the other hand, livelier places tend to provide various benefits beyond urban liveliness itself. In many Chinese cities, more vibrant and lively urban places are associated with large population inflow and increased greenery [35,36]. A large number of people on the street and more green space can strengthen the sense of safety [37-39], thus promoting healthy behaviors more effectively [8,10]. The findings extend the results of prior studies by highlighting that, in the Chinese context, a lively perception, rather than a safety perception, usually plays the strongest moderation effect on the built environment–healthy behavior relationship.

This study provides significant implications for public health. Usually, urban planning initiatives aiming to promote residents’ physical activities target interventions in built environments at the macroscale level, which are time-consuming and less actionable. Our findings suggest that favorable microscale environments can promote cycling behavior and enhance the benefits of accessibility and land use mix as well. This underscores the need for interventions on microscale streetscapes, which are less time-consuming and more readily modifiable than those on macroscale built environments.

Limitations
Several limitations should be clarified. First, similar to many studies [6,8,19], the use of 500 × 500 m grids as the analysis unit to examine moderation effects of streetscape perceptions may be susceptible to the uncertain geographic context problem [29]. In our study, the selected analysis unit was usually considered as the main activity space in most Chinese neighborhoods [21,24]. Moreover, a 1000 × 1000 m grid was used as the analysis unit to test the robustness of streetscape moderation effects. However, careful research design is essential to address the uncertain geographic context problem in future work.

Second, this is an aggregated study in nature with grids to allocate bike-sharing data. Hence, as in many bike-sharing studies [2,19], the impacts of individual characteristics on cycling were not controlled. Furthermore, streetscape perceptions were assessed by the public in terms of street view images, which may collectively influence the findings. Despite the negligible differences between respondents’ socioeconomic characteristics reported in some prior studies [40,41], future research should associate environmental perceptions with riders to draw more scientifically grounded conclusions. Finally, further study can delve into the mechanisms of moderation roles by microscale urban environments, either on the objective or perceived side, which remain unclear and could be better understood with individual-level data.

Conclusions
Livelier and safer perceptions of streetscapes can magnify the benefits of land use mix for cycling activity; the lively perception has a similar effect on accessibility. The findings emphasize that sufficient microscale streetscape interventions can encourage cycling behavior and amplify the positive effects of accessibility and land use mix.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (52308055), General Project of Ministry of Education Foundation on Humanities and Social Sciences (23YJCZH061), Social Science Foundation of Fujian Province (FJ2023C084), and the Startup Foundation of Fuzhou University (511034).

Data Availability
The data sets used or analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
HG was responsible for the conceptualization, methodology, formal analysis, and writing of the original draft. SZ was responsible for some of the formal analysis and investigation. XX provided part of the data sources and wrote some parts of the paper. JL was responsible for the methodology and investigation. HCH was responsible for the conceptualization and writing of the original draft. All authors reviewed the manuscript. JL (443171001@qq.com) and HCH (hungcho2@cityu.edu.hk) are co-corresponding authors for this manuscript.

Conflicts of Interest
None declared.
References


Abbreviations
- CBD: central business district
- HHI: Herfindahl-Hirschman index
- NBRM: negative binomial regression model
- POI: point of interest
Guo et al

Edited by A Mavragani; submitted 24.03.24; peer-reviewed by L Yang, Y Zhang; revised version received 04.06.24; accepted 06.06.24; published 03.07.24.

Please cite as:
Guo H, Zhang S, Xie X, Liu J, Ho HC
Moderation Effects of Streetscape Perceptions on the Associations Between Accessibility, Land Use Mix, and Bike-Sharing Use: Cross-Sectional Study
JMIR Public Health Surveill 2024;10:e58761
URL: https://publichealth.jmir.org/2024/1/e58761
doi:10.2196/58761

© Huagui Guo, Shuyu Zhang, Xinwei Xie, Jiang Liu, Hung Chak Ho. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 3.7.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Association Between Nitrogen Dioxide Pollution and Cause-Specific Mortality in China: Cross-Sectional Time Series Study

Jie Zeng1*, MS; Guozhen Lin2,3*, PhD; Hang Dong2,3*, PhD; Mengmeng Li4, PhD; Honglian Ruan5, PhD; Jun Yang5, PhD

1Department of Internet Medical Center, Guangdong Second Provincial General Hospital, Guangzhou, China
2Guangzhou Center for Disease Control and Prevention, Guangzhou, China
3Institute of Public Health, Guangzhou Medical University and Guangzhou Center for Disease Control and Prevention, Guangzhou, China
4State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, China
5School of Public Health, Guangzhou Medical University, Guangzhou, China

*these authors contributed equally

Corresponding Author:
Jun Yang, PhD
School of Public Health
Guangzhou Medical University
No. 1 Xinzao Road
Xinzao Town, Panyu District
Guangzhou, 511436
China
Phone: 86 020 37103532
Email: yangjun_eci@jnu.edu.cn

Abstract

Background: Nitrogen dioxide (NO₂) has been frequently linked to a range of diseases and associated with high rates of mortality and morbidity worldwide. However, there is limited evidence regarding the risk of NO₂ on a spectrum of causes of mortality. Moreover, adjustment for potential confounders in NO₂ analysis has been insufficient, and the spatial resolution of exposure assessment has been limited.

Objective: This study aimed to quantitatively assess the relationship between short-term NO₂ exposure and death from a range of causes by adjusting for potential confounders in Guangzhou, China, and determine the modifying effect of gender and age.

Methods: A time series study was conducted on 413,703 deaths that occurred in Guangzhou during the period of 2010 to 2018. The causes of death were classified into 10 categories and 26 subcategories. We utilized a generalized additive model with quasi-Poisson regression analysis using a natural cubic splines function with lag structure of 0 to 4 days to estimate the potential lag effect of NO₂ on cause-specific mortality. We estimated the percentage change in cause-specific mortality rates per 10 μg/m³ increase in NO₂ levels. We stratified meteorological factors such as temperature, humidity, wind speed, and air pressure into high and low levels with the median as the critical value and analyzed the effects of NO₂ on various death-causing diseases at those high and low levels. To further identify potentially vulnerable subpopulations, we analyzed groups stratified by gender and age.

Results: A significant association existed between NO₂ exposure and deaths from multiple causes. Each 10 μg/m³ increment in NO₂ density at a lag of 0 to 4 days increased the risks of all-cause mortality by 1.73% (95% CI 1.36%-2.09%) and mortality due to nonaccidental causes, cardiovascular disease, respiratory disease, endocrine disease, and neoplasms by 1.75% (95% CI 1.38%-2.12%), 2.06% (95% CI 1.54%-2.59%), 2.32% (95% CI 1.51%-3.13%), 2.40% (95% CI 0.84%-3.98%), and 1.18% (95% CI 0.59%-1.78%), respectively. Among the 26 subcategories, mortality risk was associated with 16, including intentional self-harm, hypertensive disease, and ischemic stroke disease. Relatively higher effect estimates of NO₂ on mortality existed for low levels of temperature, relative humidity, wind speed, and air pressure than with high levels, except a relatively higher effect estimate was present for endocrine disease at a high air pressure level. Most of the differences between subgroups were not statistically
significant. The effect estimates for NO$_2$ were similar by gender. There were significant differences between the age groups for mortality due to all causes, nonaccidental causes, and cardiovascular disease.

**Conclusions:** Short-term NO$_2$ exposure may increase the risk of mortality due to a spectrum of causes, especially in potentially vulnerable populations. These findings may be important for predicting and modifying guidelines for NO$_2$ exposure in China.

**KEYWORDS**
nitrogen dioxide; cause-specific mortality; stratification effect; vulnerable subpopulations; China

**Introduction**

Ambient air pollution is one of the greatest environmental risks to human health, with 8 million deaths reported worldwide every year [1]. Rapid industrialization in China over the last decades has significantly increased the emission of pollutants. Both short and long-term exposure to ambient air pollution increases the risk of death, years of life lost, and years of disability, thereby increasing the burden of disease [2-6]. Exposure to air pollution, including particulate matter (PM), nitrogen oxides (NO$_x$), ozone, and sulfur dioxide, may increase the risk of respiratory disease through oxidative damage, which occurs via inflammatory injury and the production of reactive oxygen species [7].

Nitrogen dioxide (NO$_2$) is a toxic gas that is present in traffic emissions, and chronic exposure to NO$_2$ is associated with respiratory inflammation, allergies, infections, and other symptoms. The current acceptable limits of NO$_2$ exposure in China are 105 ppb (200 μg/m$^3$) over 1 hour and an average of 21 ppb (40 μg/m$^3$) annually [8]. In recent decades, however, the concentration of NO$_2$ in China has exceeded the global average level [9]. Guangzhou is the capital and largest city of Guangdong province in China, with a population of 13,964,637 [10]. According to data released by the World Health Organization [11] in 2018, the annual average concentration of NO$_2$ reached 50 μg/m$^3$, which exceeds air quality standards (25 μg/m$^3$).

Evidence clearly demonstrates that short-term exposure to NO$_2$ is related to an increased risk of mortality from different diseases, showing independent and linear trends [12-16]. For instance, a multicity analysis showed NO$_2$ exposure increased the all-cause mortality rates, as well as mortality due to cardiovascular and respiratory diseases [12]. However, other studies have reported a lower mortality risk due to chronic diseases in the event of NO$_2$ exposure [17]. Research on the mechanisms by which NO$_2$ affects mortality has come to limited conclusions. For example, it is not yet known whether NO$_2$ exposure can cause endocrine disease and neoplasms.

Regarding the effect of NO$_2$ on diseases, there is less evidence for the different effects caused by stratification of meteorological factors. The significance of other meteorological risk factors, such as temperature, humidity, air pressure, and wind speed, has been less frequently explored. In addition, most of the studied disease outcomes affected by NO$_2$ have included only cardiovascular and respiratory diseases. Moreover, the influence of age and gender on the association between NO$_2$ and mortality is still unclear. Therefore, it is essential to fully explore predisposing diseases associated with NO$_2$ exposure.

To this end, we aimed to examine the relationship between short-term NO$_2$ exposure and a spectrum of causes of mortality in Guangzhou to identify the influencing factors. Our research may offer unique insights into the health effects of NO$_2$ and the underlying factors and help formulate public policies to protect potentially vulnerable subgroups [18].

**Methods**

**Sample and Data**

The daily death count in Guangzhou during the period between 2010 and 2018 was obtained from the city registry, which included all permanent residents and is published by the Guangzhou Center for Disease Control and Prevention [19]. Using the 10th Revision of the International Classification of Diseases (ICD-10), the causes of the deaths were classified as all causes (A00-Z99), nonaccidental causes (A00-R99), cardiovascular diseases (I00-I99), neoplasms (C00-D48), endocrine diseases (D50-D89, E00-E90), nervous system diseases (G00-G99), respiratory diseases (J00-J99), digestive diseases (K00-K93), genitourinary diseases (N00-N99), and external causes (V01-Y89). These 10 groups were divided further into 26 subgroups, with each group containing at least an average number of deaths per day, ensuring that the model was convergent.

**Measures of Variables**

Data on the average daily level of NO$_2$ from 2010 to 2018 were obtained from the Guangzhou Bureau of Environmental Protection. The data were collected from 11 fixed-point air pollution monitoring stations across the city (Figure 1). At the same time, the daily concentration of atmospheric dynamic diameter particulate matter (PM$_{10}$ or PM$_{2.5}$), ozone (O$_3$), sulfur dioxide (SO$_2$), carbon monoxide (CO), and other gaseous pollutants were obtained at the same sites and were used to adjust for potential hybrid co-pollutants during model building. Exposure levels were determined by calculating the average daily concentration level at each monitoring station. General data on the mean temperature (°C), mean relative humidity (%), mean air pressure (hPa), sunshine duration (h), wind speed (km/h), and daily air quality index were obtained from the China Meteorological Data Sharing Service System [20].
Models and Data Analysis Procedure

A generalized additive model with Poisson regression analysis was used to determine the relationship between NO$_2$ exposure and each cause of mortality after adjusting for potential covariates. A natural cubic spline function (NS) of 7 degrees of freedom ($df$) per year was used to control for long-term and seasonal trends in mortality. In the $df$ adjustment of other weather factors, the NS functions of 6, 3, and 3 $dfs$ were used to adjust the $dfs$ of temperature, relative humidity, wind speed, and atmospheric pressure, respectively. The day of the week and public holidays were added to the model as indicator variables. The method and selection of model parameters have been described in a previous research analysis [21,22]. After establishing the basic model, we used an NS of 2 to 6 $dfs$ to investigate the relationship between NO$_2$ concentration and mortality as a result of each cause of death. Based on the most likely days of lag, a moving average method was used to correlate the cumulative effects of pollutants over the defined lag period. This is consistent with a previous study [15,23] that selected an empirical maximum of 4 $dfs$ for NO$_2$ and used a lag model structure of 0 to 4 days to obtain an estimate of the cumulative risk of NO$_2$ exposure during the first 4 days.

To identify potentially vulnerable subpopulations at a particular risk of mortality due to NO$_2$ exposure on a spectrum of mortality causes, stratified analyses were conducted based on age groups (0-64 and ≥65 years) and gender (men and women) based on the causes of death that showed a significant association with NO$_2$. In addition, in order to analyze the interactive health effects between meteorological factors and NO$_2$, the stratification analyses were conducted based on the levels of temperature, humidity, wind speed, and air pressure, with their median values as the cutoff. The difference between the 2 groups was analyzed using a $z$ test, as follows [24]:

$$E_1 - E_2 \pm \sigma_{E_1 - E_2}$$

where $E_1$ and $E_2$ are the logarithms of risk ratios for 2 categories, such as men and women or 0-64 years and ≥65 years, and $SE_{E_1}$ and $SE_{E_2}$ were used to indicate the relevant standard error [25,26].

Multiple sensitivity analyses were conducted to evaluate the robustness of the model and parameter validation estimates. The $dfs$ for the time variable were set as 3 years to 7 years, and the $dfs$ for the weather variable were set as 3 to 6. To adjust for the mixing effect of temperature during longer lag days, 2-dimensional, cross-basis NSs and 4 $dfs$ were used for temperature and lag dimensions, respectively, and a maximum lag of 21 days was used for the distributed lag nonlinear model (DLNM) [27]. Dual pollutant models were constructed to verify the robustness of the main model after adjusting for confounders by adding another air pollutant (O$_3$, PM$_{2.5}$, PM$_{10}$, CO, SO$_2$) to the main model.

All statistical analyses were performed using the mgcv and dlnm packages of R software (version 3.4.1) by major models [28]. The change in mortality was calculated for every 10 $\mu$g/m$^3$ increase in NO$_2$ grade. All statistical results are presented as 2-sided values, while a $P$ value <.05 was considered to indicate statistical significance.

Ethics Approval

Ethics approval was not required for secondary analysis of the anonymous data in this study.
Results

The average annual temperature, humidity, daily NO$_2$ concentrations, and number of cause-specific deaths are summarized in Table 1. The Spearman correlation between air pollution and weather conditions is shown in Figure S1 in Multimedia Appendix 1. The correlation coefficients between NO$_2$ and other constituents ranged from 0.1 to 0.94. Temperature and wind speed were negatively correlated with NO$_2$, while NO$_2$ was negatively correlated with other meteorological factors and air pollutants.

Table 1. Summary statistics of environmental monitoring and daily death counts during 2010 to 2018 in Guangzhou, China.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Air pollutant concentrations</strong></td>
<td></td>
</tr>
<tr>
<td>NO$_2$ ($\mu$g/m$^3$)</td>
<td>47 (19)</td>
</tr>
<tr>
<td>Particulate matter (PM)$_{2.5}$ ($\mu$g/m$^3$)</td>
<td>38 (22)</td>
</tr>
<tr>
<td>SO$_2$ ($\mu$g/m$^3$)</td>
<td>18 (12)</td>
</tr>
<tr>
<td>O$_3$ ($\mu$g/m$^3$)</td>
<td>82 (47)</td>
</tr>
<tr>
<td>CO (mg/m$^3$)</td>
<td>0.9 (0.2)</td>
</tr>
<tr>
<td>Air quality index</td>
<td>72 (31)</td>
</tr>
<tr>
<td><strong>Weather conditions</strong></td>
<td></td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>22 (6)</td>
</tr>
<tr>
<td>Relative humidity (%)</td>
<td>79 (11)</td>
</tr>
<tr>
<td>Air pressure (hPa)</td>
<td>1007 (56)</td>
</tr>
<tr>
<td>Wind speed (km/h)</td>
<td>22 (11)</td>
</tr>
<tr>
<td><strong>Daily numbers of deaths</strong></td>
<td></td>
</tr>
<tr>
<td>All causes</td>
<td>126 (27)</td>
</tr>
<tr>
<td>Nonaccidental causes</td>
<td>119 (26)</td>
</tr>
<tr>
<td>Circulatory disease</td>
<td>49 (14)</td>
</tr>
<tr>
<td>Respiratory disease</td>
<td>19 (7)</td>
</tr>
<tr>
<td>Digestive disease</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Nervous disease</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Genitourinary disease</td>
<td>2 (1)</td>
</tr>
<tr>
<td>External causes</td>
<td>7 (3)</td>
</tr>
<tr>
<td>Endocrine disease</td>
<td>5 (3)</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>35 (8)</td>
</tr>
</tbody>
</table>

The average annual NO$_2$ level was 47 $\mu$g/m$^3$. During the study period of 2010 to 2018, 413,703 all-cause deaths were recorded in Guangzhou, with 126 deaths occurring daily, on average. The majority (391,543/413,703, 94.6%) of the deaths were due to nonaccidental causes. Other causes included cardiovascular diseases (160,226/413,703, 38.7%), respiratory diseases (63,290/413,703, 15.3%), digestive diseases (12,787/413,703, 3.1%), nervous system diseases (3659/413,703, 0.9%), genitourinary diseases (5314/413,703, 1.3%), external causes (22,147/413,703, 5.4%), endocrine diseases (14,972/413,703, 3.6%), and neoplasms (115,703/413,703, 28%; Table S1 in Multimedia Appendix 1).

The association between NO$_2$ exposure and mortality due to specific causes over 6 days (lags 0 to 6) and stratified by gender and age are shown in Table 2 and Figure S2 in Multimedia Appendix 1. For all-cause mortality, the highest risk of death due to NO$_2$ exposure was observed at the lag of 2 or 3 days and was usually limited to 4 days. Effect estimates of NO$_2$ were stronger for women and older subgroups for different diseases before the lag of 4 days, and the effects stabilized thereafter. The mortality risk increased linearly with NO$_2$ concentrations at a lag of 0 to 4 days (Figure 2). Finally, we examined the stability of the model by using lag 0 days for NO$_2$ and observed a similar linear relationship between NO$_2$ and mortality (Figure S4 in Multimedia Appendix 1). The results of the generalized Poisson regression are shown in Table S2 in Multimedia Appendix 1. The goodness of fit index $R^2$ of the model was 0.69, which meant that the model was well constructed.
Table 2. The relative risk of death resulting from the different causes associated with an increase of 10 μg/m³ in NO₂ levels with a 1-day delay (lag days of 0, 1, 2, 3, 4, 5, and 6) for each gender and age group.

<table>
<thead>
<tr>
<th>Causes of death</th>
<th>All, % change (95% CI)</th>
<th>Gender, % change (95% CI)</th>
<th>Age (years), % change (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>All causes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.009 (1.006-1.011)</td>
<td>1.009 (1.005-1.012)</td>
<td>1.009 (1.005-1.012)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.011 (1.008-1.014)</td>
<td>1.010 (1.007-1.013)</td>
<td>1.013 (1.010-1.017)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>1.010 (1.007-1.013)</td>
<td>1.009 (1.006-1.012)</td>
<td>1.012 (1.009-1.016)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>1.007 (1.005-1.010)</td>
<td>1.007 (1.003-1.010)</td>
<td>1.008 (1.004-1.011)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>1.003 (1.001-1.007)</td>
<td>1.004 (1.001-1.007)</td>
<td>1.004 (1.000-1.007)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>1.001 (0.998-1.003)</td>
<td>1.001 (0.998-1.005)</td>
<td>0.999 (0.996-1.003)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.997 (0.995-1.999)</td>
<td>0.997 (0.994-1.999)</td>
<td>0.997 (0.993-1.000)</td>
</tr>
<tr>
<td>Nonaccidental causes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.008 (1.006-1.011)</td>
<td>1.009 (1.005-1.012)</td>
<td>1.008 (1.004-1.012)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.011 (1.009-1.014)</td>
<td>1.010 (1.007-1.013)</td>
<td>1.013 (1.010-1.017)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>1.010 (1.008-1.013)</td>
<td>1.008 (1.005-1.012)</td>
<td>1.013 (1.010-1.017)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>1.007 (1.005-1.010)</td>
<td>1.007 (1.003-1.010)</td>
<td>1.008 (1.005-1.012)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>1.004 (1.002-1.007)</td>
<td>1.004 (1.001-1.007)</td>
<td>1.004 (1.000-1.008)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>1.001 (0.998-1.003)</td>
<td>1.001 (0.998-1.005)</td>
<td>1.000 (0.996-1.003)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.997 (0.995-1.999)</td>
<td>0.997 (0.994-1.999)</td>
<td>0.997 (0.993-1.000)</td>
</tr>
<tr>
<td>Circulatory diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.009 (1.005-1.012)</td>
<td>1.008 (1.004-1.013)</td>
<td>1.009 (1.004-1.014)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.013 (1.009-1.016)</td>
<td>1.011 (1.004-1.013)</td>
<td>1.014 (1.004-1.014)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>1.012 (1.008-1.016)</td>
<td>1.009 (1.004-1.014)</td>
<td>1.015 (1.004-1.014)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>1.009 (1.006-1.013)</td>
<td>1.010 (1.005-1.014)</td>
<td>1.009 (1.004-1.014)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>1.007 (1.003-1.011)</td>
<td>1.009 (1.004-1.014)</td>
<td>1.005 (1.000-1.010)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>1.003 (0.999-1.007)</td>
<td>1.004 (1.004-1.014)</td>
<td>1.001 (0.996-1.006)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.997 (0.993-1.000)</td>
<td>0.997 (0.992-1.001)</td>
<td>0.997 (0.996-1.006)</td>
</tr>
<tr>
<td>Respiratory diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.010 (1.004-1.015)</td>
<td>1.012 (1.005-1.020)</td>
<td>1.006 (0.997-1.014)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.015 (1.010-1.021)</td>
<td>1.015 (1.008-1.022)</td>
<td>1.017 (1.008-1.025)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>1.017 (1.011-1.022)</td>
<td>1.015 (1.008-1.022)</td>
<td>1.019 (1.011-1.028)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>1.008 (1.003-1.014)</td>
<td>1.010 (1.003-1.017)</td>
<td>1.005 (0.997-1.014)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>1.005 (0.999-1.011)</td>
<td>1.008 (1.001-1.015)</td>
<td>1.001 (0.992-1.009)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>1.001 (0.996-1.007)</td>
<td>1.001 (0.994-1.008)</td>
<td>1.002 (0.993-1.010)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.997 (0.992-1.003)</td>
<td>0.996 (0.989-1.003)</td>
<td>1.000 (0.991-1.008)</td>
</tr>
<tr>
<td>Digestive diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.004 (0.992-1.015)</td>
<td>0.997 (0.982-1.012)</td>
<td>1.014 (0.996-1.033)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.003 (0.992-1.015)</td>
<td>0.997 (0.983-1.012)</td>
<td>1.010 (0.992-1.029)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>0.999 (0.988-1.010)</td>
<td>0.996 (0.981-1.010)</td>
<td>1.003 (0.985-1.022)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>0.998 (0.987-1.009)</td>
<td>0.997 (0.982-1.011)</td>
<td>1.000 (0.981-1.019)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>0.997 (0.985-1.008)</td>
<td>0.990 (0.975-1.005)</td>
<td>1.008 (0.989-1.027)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>0.991 (0.980-1.003)</td>
<td>0.988 (0.973-1.003)</td>
<td>0.997 (0.978-1.016)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.995 (0.980-1.003)</td>
<td>0.995 (0.981-1.010)</td>
<td>0.995 (0.976-1.013)</td>
</tr>
<tr>
<td>Causes of death</td>
<td>All, % change (95% CI)</td>
<td>Gender, % change (95% CI)</td>
<td>Age (years), % change (95% CI)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------</td>
<td>---------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Nervous diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>0.995 (0.973-1.017)</td>
<td>0.986 (0.957-1.016)</td>
<td>1.005 (0.972-1.038)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>0.992 (0.971-1.014)</td>
<td>0.982 (0.953-1.012)</td>
<td>1.004 (0.973-1.037)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>0.999 (0.978-1.021)</td>
<td>0.995 (0.966-1.025)</td>
<td>1.004 (0.973-1.037)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>1.010 (0.981-1.032)</td>
<td>1.015 (0.985-1.045)</td>
<td>1.005 (0.973-1.038)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>1.009 (0.987-1.032)</td>
<td>1.025 (0.995-1.057)</td>
<td>0.991 (0.958-1.025)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>0.985 (0.963-1.007)</td>
<td>0.995 (0.965-1.026)</td>
<td>0.973 (0.940-1.006)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.980 (0.959-1.002)</td>
<td>0.977 (0.949-1.007)</td>
<td>0.983 (0.951-1.016)</td>
</tr>
<tr>
<td>Genitourinary diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.009 (0.991-1.028)</td>
<td>1.008 (0.984-1.032)</td>
<td>1.012 (0.985-1.040)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.007 (0.989-1.025)</td>
<td>1.010 (0.986-1.034)</td>
<td>1.004 (0.977-1.032)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>1.004 (0.986-1.022)</td>
<td>1.018 (0.995-1.043)</td>
<td>0.986 (0.959-1.013)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>0.998 (0.979-1.016)</td>
<td>1.005 (0.981-1.030)</td>
<td>0.989 (0.961-1.017)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>0.994 (0.976-1.013)</td>
<td>0.986 (0.961-1.011)</td>
<td>1.005 (0.977-1.034)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>0.991 (0.973-1.010)</td>
<td>0.986 (0.962-1.011)</td>
<td>0.998 (0.970-1.027)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.988 (0.970-1.007)</td>
<td>0.995 (0.971-1.019)</td>
<td>0.980 (0.953-1.008)</td>
</tr>
<tr>
<td>External causes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.013 (1.004-1.022)</td>
<td>1.008 (0.996-1.020)</td>
<td>1.020 (1.006-1.034)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.010 (1.001-1.020)</td>
<td>1.008 (0.996-1.019)</td>
<td>1.014 (1.000-1.029)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>1.006 (0.996-1.015)</td>
<td>1.010 (0.998-1.022)</td>
<td>0.999 (0.985-1.013)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>1.002 (0.993-1.012)</td>
<td>1.005 (0.993-1.017)</td>
<td>0.998 (0.984-1.012)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>1.001 (0.991-1.010)</td>
<td>1.001 (0.989-1.014)</td>
<td>1.000 (0.985-1.014)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>0.998 (0.988-1.007)</td>
<td>1.002 (0.990-1.014)</td>
<td>0.991 (0.977-1.006)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.998 (0.989-1.007)</td>
<td>0.999 (0.987-1.011)</td>
<td>0.996 (0.982-1.010)</td>
</tr>
<tr>
<td>Endocrine diseases</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.011 (1.000-1.022)</td>
<td>1.006 (0.991-1.022)</td>
<td>1.015 (1.000-1.030)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.020 (1.009-1.030)</td>
<td>1.017 (1.002-1.032)</td>
<td>1.022 (1.008-1.037)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>1.020 (1.009-1.031)</td>
<td>1.019 (1.003-1.034)</td>
<td>1.021 (1.007-1.036)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>1.007 (0.997-1.018)</td>
<td>1.003 (0.987-1.019)</td>
<td>1.011 (0.996-1.026)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>0.998 (0.987-1.010)</td>
<td>0.996 (0.980-1.012)</td>
<td>1.001 (0.986-1.016)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>0.997 (0.986-1.008)</td>
<td>0.996 (0.980-1.013)</td>
<td>0.997 (0.982-1.012)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>0.988 (0.977-0.999)</td>
<td>0.981 (0.966-0.997)</td>
<td>0.994 (0.979-1.008)</td>
</tr>
<tr>
<td>Neoplasms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag 0</td>
<td>1.006 (1.002-1.010)</td>
<td>1.008 (1.002-1.013)</td>
<td>1.003 (0.996-1.010)</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.007 (1.003-1.011)</td>
<td>1.007 (1.002-1.012)</td>
<td>1.007 (1.000-1.013)</td>
</tr>
<tr>
<td>Lag 2</td>
<td>1.007 (1.003-1.011)</td>
<td>1.007 (1.002-1.012)</td>
<td>1.007 (1.000-1.013)</td>
</tr>
<tr>
<td>Lag 3</td>
<td>1.006 (1.002-1.010)</td>
<td>1.003 (0.998-1.008)</td>
<td>1.011 (1.004-1.018)</td>
</tr>
<tr>
<td>Lag 4</td>
<td>1.002 (0.998-1.006)</td>
<td>0.999 (0.994-1.005)</td>
<td>1.006 (1.000-1.013)</td>
</tr>
<tr>
<td>Lag 5</td>
<td>1.000 (0.996-1.005)</td>
<td>1.001 (0.996-1.006)</td>
<td>0.999 (0.992-1.006)</td>
</tr>
<tr>
<td>Lag 6</td>
<td>1.000 (0.996-1.004)</td>
<td>1.001 (0.996-1.007)</td>
<td>0.999 (0.992-1.005)</td>
</tr>
</tbody>
</table>
Figure 2. The dose-effect of the curves of NO$_2$ and mortality as a result of (A) all causes, (B) nonaccidental causes, (C) cardiovascular diseases, (D) respiratory diseases, (E) digestive diseases, (F) nervous system diseases, (G) genitourinary diseases, (H) external causes, (I) endocrine diseases, and (J) neoplasms at 7 degrees of freedom with a lag of 0 to 4 days, stratified by gender and age (years).

Among 10 categories, for every 10 $\mu$g/m$^3$ increment in NO$_2$ at a lag of 0 to 4 days, the risks of mortality increased due to all causes by 1.73% (95% CI 1.36%-2.09%) and due to nonaccidental causes, cardiovascular diseases, respiratory diseases, endocrine diseases, and neoplasms by 1.75% (95% CI 1.38%-2.12%), 2.06% (95% CI 1.54%-2.59%), 2.32% (95% CI 1.51%-3.13%), 2.40% (95% CI 0.84%-3.98%), and 1.18% (95% CI 0.59%-1.78%), respectively. On the other hand, NO$_2$ did not significantly impact deaths due to digestive diseases, genitourinary diseases, nervous diseases, and external causes (Table 3 and Figure S3 in Multimedia Appendix 1).
JMIR PUBLIC HEALTH AND SURVEILLANCE

Zeng et al

Table 3. Percentage increase in mortality as a result of different diseases per 10 μg/m3 increase in NO2 with a lag of 0 to 4 days in Guangzhou, China,
stratified by gender and age groups.
Causes of death

All, % change (95% CI) Gender, % change (95% CI)

Age (years), % change (95% CI)

Male

Female

0-64

≥65

All causes

1.73 (1.36 to 2.09)

1.56 (1.11 to 2.00)

1.95 (1.45 to 2.46)

1.15 (0.66 to 1.63)

2.15 (1.68 to 2.62)

Nonaccidental causes

1.75 (1.38 to 2.12)

1.57 (1.12 to 2.03)

1.98 (1.47 to 2.50)

1.13 (0.62 to 1.63)

2.18 (1.71 to 2.66)

Cardiovascular diseases

2.06 (1.54 to 2.59)

1.97 (1.29 to 2.65)

2.17 (1.46 to 2.90)

1.13 (0.29 to 1.98)

2.50 (1.87 to 3.12)

Chronic rheumatic heart diseases

2.32 (1.56 to 3.08)

2.07 (1.05 to 3.10)

2.62 (1.55 to 3.71)

1.62 (0.37 to 2.88)

2.68 (1.76 to 3.60)

Hypertensive diseases

3.39 (1.80 to 4.99)

2.79 (0.50 to 5.13)

3.90 (1.75 to 6.10)

1.88 (–1.15 to 5.01)

3.88 (2.08 to 5.71)

Ischemic heart diseases

1.46 (0.70 to 2.23)

1.61 (0.56 to 2.66)

1.32 (0.26 to 2.40)

0.59 (–0.74 to 1.93)

1.82 (0.91 to 2.73)

Acute ischemic heart disease

1.27 (0.20 to 2.34)

1.71 (0.30 to 3.13)

0.71 (–0.85 to 2.30)

0.84 (–0.84 to 2.56)

1.51 (0.18 to 2.87)

Acute myocardial infarction

1.23 (0.13 to 2.34)

1.69 (0.24 to 3.16)

0.65 (–0.98 to 2.30)

1.05 (–0.70 to 2.82)

1.33 (–0.05 to 2.73)

Myocardial infarction

1.23 (0.12 to 2.34)

1.67 (0.22 to 3.14)

0.66 (–0.97 to 2.31)

1.04 (–0.70 to 2.82)

1.33 (–0.05 to 2.73)

Chronic ischemic heart disease

1.65 (0.60 to 2.72)

1.54 (0.02 to 3.07)

1.77 (0.33 to 3.23)

0.21 (–1.99 to 2.45)

2.05 (0.85 to 3.26)

Other forms of heart disease

–0.61 (–3.09 to 1.93)

–0.66 (–3.97 to 2.76) –0.59 (–4.34 to 3.30)

–1.28 (–4.81 to 2.39) 0.04 (–3.41 to 3.62)

Cerebrovascular diseases

2.48 (1.70 to 3.27)

2.21 (1.16 to 3.27)

2.80 (1.69 to 3.91)

1.64 (0.35 to 2.94)

2.89 (1.95 to 3.84)

Stroke

2.48 (1.55 to 3.42)

2.30 (1.03 to 3.58)

2.71 (1.35 to 4.08)

1.47 (–0.01 to 2.97)

3.07 (1.90 to 4.25)

Intracerebral hemorrhagic
stroke

1.75 (0.21 to 3.30)

2.18 (0.21 to 4.18)

1.19 (–1.18 to 3.61)

1.23 (–0.90 to 3.40)

2.18 (0.06 to 4.35)

Ischemic stroke

2.84 (1.35 to 4.35)

2.35 (0.30 to 4.43)

3.35 (1.24 to 5.50)

0.95 (–1.72 to 3.69)

3.60 (1.85 to 5.39)

Arteries, arterioles, and capil- 0.27 (–3.75 to 4.45)
laries

0.44 (–4.40 to 5.53)

–0.29 (–7.17 to 7.09)

2.04 (–3.73 to 8.16)

–1.24 (–6.66 to 4.50)

Respiratory diseases

2.32 (1.51 to 3.13)

2.53 (1.50 to 3.57)

2.03 (0.82 to 3.27)

2.15 (0.41 to 3.92)

2.36 (1.46 to 3.27)

Influenza and pneumonia

2.47 (1.35 to 3.60)

2.72 (1.20 to 4.27)

2.19 (0.62 to 3.80)

2.50 (0.06 to 5.00)

2.45 (1.21 to 3.71)

Chronic lower respiratory
disease

2.33 (1.16 to 3.52)

2.57 (1.17 to 3.99)

1.90 (–0.04 to 3.87)

2.38 (–0.14 to 4.96)

2.32 (1.02 to 3.63)

Chronic obstructive pulmonary disease

2.33 (1.12 to 3.56)

2.47 (1.03 to 3.93)

2.10 (0.08 to 4.16)

2.71 (0.01 to 5.48)

2.25 (0.93 to 3.58)

Other respiratory disease

1.52 (–0.69 to 3.77)

1.31 (–1.81 to 4.53)

1.68 (–1.41 to 4.87)

3.19 (–1.64 to 8.27)

1.09 (–1.34 to 3.59)

Digestive diseases

–0.09 (–1.71 to 1.55)

–1.18 (–3.26 to 0.94) 1.61 (–1.07 to 4.35)

–0.71 (–3.19 to 1.83) 0.42 (–1.73 to 2.63)

Esophagus, stomach, and
duodenum disease

0.34 (–3.10 to 3.89)

–0.95 (–5.55 to 3.89) 1.90 (–3.14 to 7.21)

1.25 (–5.01 to 7.93)

Liver disease

0.28 (–2.63 to 3.27)

–0.42 (–3.71 to 2.98) 2.44 (–3.58 to 8.84)

–1.34 (–4.64 to 2.08) 5.14 (–0.77 to 11.39)

Other digestive disease

–0.07 (–3.72 to 3.71)

–1.16 (–5.95 to 3.88) 1.39 (–4.29 to 7.41)

0.49 (–6.45 to 7.94)

–0.34 (–4.55 to 4.06)

Nervous diseases

0.32 (–2.78 to 3.51)

0.03 (–4.20 to 4.44)

0.64 (–3.93 to 5.42)

1.18 (–2.92 to 5.46)

–0.95 (–5.54 to 3.86)

Genitourinary diseases

0.57 (–2.04 to 3.24)

1.12 (–2.31 to 4.67)

–0.07 (–3.97 to 3.98)

–0.29 (–4.12 to 3.71) 1.17 (–2.35 to 4.81)

Urinary diseases

0.57 (–2.04 to 3.24)

1.12 (–2.31 to 4.67)

–0.07 (–3.97 to 3.98)

–0.29 (–4.12 to 3.71) 1.17 (–2.35 to 4.81)

Renal failure

0.82 (–2.75 to 4.52)

2.22 (–2.51 to 7.16)

–0.87 (–6.17 to 4.74)

–0.20 (–5.41 to 5.30) 1.68 (–3.08 to 6.67)

External causes

1.32 (–0.01 to 2.67)

1.36 (–0.37 to 3.11)

1.26 (–0.76 to 3.33)

1.38 (–0.33 to 3.12)

Road traffic injury

0.28 (–2.89 to 3.55)

0.68 (–2.85 to 4.34)

–1.10 (–7.86 to 6.17)

–0.40 (–3.66 to 2.97) 13.02 (–0.64 to 28.55)

Intentional self-harm

3.99 (0.22 to 7.90)

2.08 (–2.59 to 6.97)

7.09 (1.06 to 13.48)

3.05 (–0.93 to 7.20)

9.85 (–0.31 to 21.03)

Endocrine diseases

2.40 (0.84 to 3.98)

1.76 (–0.48 to 4.06)

2.95 (0.81 to 5.14)

1.01 (–1.36 to 3.43)

3.41 (1.35 to 5.51)

Diabetes

2.04 (0.12 to 3.99)

1.15 (–1.54 to 3.91)

2.97 (0.25 to 5.76)

1.02 (–1.72 to 3.83)

3.00 (0.31 to 5.77)

Neoplasm

1.18 (0.59 to 1.78)

1.03 (0.28 to 1.77)

1.45 (0.48 to 2.42)

1.06 (0.34 to 1.79)

1.41 (0.41 to 2.42)

https://publichealth.jmir.org/2024/1/e44648

XSL• FO
RenderX

–0.03 (–4.08 to 4.20)

1.27 (–0.80 to 3.38)

JMIR Public Health Surveill 2024 | vol. 10 | e44648 | p.1797
(page number not for citation purposes)


Among the 26 disease subcategories, after NO\textsubscript{2} exposure, there were significant increases in the top 10 causes with the highest risk of mortality: intentional self-harm (3.99%, 95% CI 0.22%-7.90%), hypertensive diseases (3.39%, 95% CI 1.80%-4.99%), ischemic stroke (2.84%, 95% CI 1.35%-4.35%), cerebrovascular diseases (2.48%, 95% CI 1.70%-3.27%), stroke (2.48%, 95% CI 1.55%-3.42%), influenza and pneumonia (2.47%, 95% CI 1.35%-3.60%), chronic obstructive pulmonary disease (2.33%, 95% CI 1.12%-3.56%), chronic lower respiratory disease (2.33%, 95% CI 1.16%-3.52%), chronic rheumatic heart disease (2.32%, 95% CI 1.56%-3.08%), and diabetes (2.04%, 95% CI 0.12%-3.99%; Table 3 and Figure S3 in Multimedia Appendix 1).

The results of the stratified analysis are summarized in Table 3. For all-cause mortality, the impact of NO\textsubscript{2} was relatively higher among women than men, but the difference was not statistically significant (P=.32). Older adults were at a higher risk than the younger age groups (P=.02), and their death rates due to NO\textsubscript{2} exposure increased by 2.15% (95% CI 1.68%-2.62%) and 1.15% (95% CI 0.66%-1.63%), respectively. Age was a particularly significant factor for deaths due to nonaccidental causes and cardiovascular diseases.

Stratification analyses by meteorological factor levels were further conducted (Table 4). We observed relatively higher effect estimates of NO\textsubscript{2} on mortality for low levels of temperature, relative humidity, and wind speed than high levels, although most of the differences between the subgroups were not statistically significant. For air pressure, relatively higher effect estimates were found for all causes, nonaccidental causes, cardiovascular diseases, respiratory diseases, and neoplasms at low levels of air pressure than at high levels of air pressure. However, the opposite was true for endocrine diseases.

The sensitivity analysis showed that each of the different modeling strategies provided roughly similar results. First, the calendar date and weather variables were varied using 3 to 7 \textit{dfs} and 3 to 6 \textit{dfs}, respectively, year-on-year. Second, we used DLNM to adjust for temperature and found that the estimated effect of NO\textsubscript{2} was slightly attenuated but still statistically significant. Third, the main findings were stable with additional adjustment for SO\textsubscript{2}, PM\textsubscript{2.5}, O\textsubscript{3}, PM\textsubscript{10}, and CO in the 2-pollutant model. As shown in Figure S5 in Multimedia Appendix 1, the model demonstrated stability and reliability after using alternative \textit{dfs} to control for the meteorological confounders.

### Table 4. Stratified analysis of the percent increase in mortality as a result of different diseases per 10 \(\mu\text{g/m}^3\) increase in NO\textsubscript{2} at a lag of 0 to 4 days during 2010 to 2018 in Guangzhou, China.

<table>
<thead>
<tr>
<th>Causes of death</th>
<th>Total, % change (95% CI)</th>
<th>Gender, % change (95% CI)</th>
<th>Age (years), % change (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All, % change (95% CI)</td>
<td>Female, % change (95% CI)</td>
<td>Male, % change (95% CI)</td>
</tr>
<tr>
<td></td>
<td>Temperature, % change (95% CI)</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Pancreas disease</td>
<td>0.22 (-3.12 to 3.69)</td>
<td>2.35 (-2.30 to 7.24)</td>
<td>-2.11 (-6.76 to 2.78)</td>
</tr>
</tbody>
</table>

The principal findings of this study indicate a significant correlation between NO\textsubscript{2} exposure and multiple causes of mortality, with notable increases in cardiovascular, respiratory, and endocrine diseases. The results highlight the importance of considering both acute and long-term effects of NO\textsubscript{2} on public health, necessitating further research into the mechanisms underlying this association and potential interventions to mitigate health impacts.
increased the risk of all-cause mortality by 1.73% and had a similar impact on deaths due to nonaccidental causes (1.75% increase), cardiovascular diseases (2.06% increase), respiratory diseases (2.32% increase), endocrine diseases (2.40% increase), and neoplasms (1.18% increase). Following NO\textsubscript{2} exposure, older adults were at a higher risk of death due to all causes, nonaccidental causes, and cardiovascular diseases. Furthermore, the impact of NO\textsubscript{2} was higher in warm temperatures than in cold temperatures.

Multiple studies have found that low wind speeds lead to high atmospheric stability and high levels of air pollution, which in turn have led to higher numbers of COVID-19–related infections and deaths [29,30]. Atmospheric stability based on low wind speeds reduces the diffusion of gases and air pollution particles [29]. The results of our analyses were similar. However, several studies found that wind speed and COVID-19 are positively correlated. An inverted U-shaped dose-response curve was found for wind speed and COVID-19 [30]. A Japanese study involving 74 participants reported that low humidity and high air pressure contribute to brain hemorrhage [31]. In contrast, a hospital-based study in Mexico showed no significant relationship between barometric pressure and stroke [32]. However, the influence of humidity and air pressure may be not as strong as other weather conditions such as temperature. Therefore, we included humidity, wind speed, and air pressure as confounding factors in the model to avoid potential bias.

Previous studies on the association between NO\textsubscript{2} and mortality mainly focused on overall causes, cardiovascular diseases, and respiratory diseases [13,15,16]. Our study also showed a statistically significant association between NO\textsubscript{2} and endocrine system diseases. We found that the impact of NO\textsubscript{2} exposure on mortality rates associated with endocrine, respiratory, and cardiovascular diseases was higher than that reported in a previous meta-analysis [33]. Studies previously conducted in China also reported that, for every 10 μg/m\textsuperscript{3} increase in NO\textsubscript{2} concentration, mortality rates related to all causes, endocrine diseases, cardiovascular diseases, and respiratory diseases increased by 0.96%, 1.13%, 1.01%, and 1.22%, respectively [34,35]. In western countries, available data indicate that NO\textsubscript{2} exposure leads to an increase in total mortality as well as mortality associated with endocrine diseases, cardiovascular diseases, and respiratory diseases, by 0.33%, 0.38%, 0.40%, and 0.38%, respectively [13,36]. These differences could be attributed to racial factors and the varying analytical techniques among the different studies.

NO\textsubscript{2} can affect disease-related mortality through various mechanisms. For instance, altered immune responses and inflammatory reactions following NO\textsubscript{2} exposure may lead to endocrinological disorders. Inhaled NO\textsubscript{2} triggers production of proinflammatory cytokines by alveolar macrophages, resulting in local oxidative stress. Furthermore, the cytokines enter systemic circulation and affect distant tissues, contributing to autoimmune responses and metabolic dysfunction [37,38]. NO\textsubscript{2} is also a respiratory irritant that generates highly reactive, free radicals [39], which can cause severe lung injury, and death can occur depending on the dose and duration of exposure [40]. In addition, for individuals exposed to NO\textsubscript{2}, levels of inflammatory markers such as IL-12 and C-reactive protein in the blood are elevated, which may destabilize atherosclerotic plaques and result in their rupture, leading to increased blood pressure and enhanced thrombosis formation [41].

Gender and age affect the level of sensitivity to air pollution [42,43]. The impact of the aforementioned factors on the effects generated by NO\textsubscript{2} were largely inconclusive in previous studies. We found that gender was not a determinant of mortality risk due to NO\textsubscript{2} exposure, which is consistent with previous studies [15,44-46]. However, one study showed that men are more susceptible to the detrimental effects of NO\textsubscript{2} [14], which can be attributed to occupational and physiological differences. Furthermore, several studies have shown that older adults are predisposed to NO\textsubscript{2}-related mortality risk [13,15,23,47]. Consistent with this, the older adults in our cohort showed a higher risk of death due to all causes, nonaccidental causes, and cardiovascular diseases following NO\textsubscript{2} exposure. A possible explanation is that older people have higher rates of chronic diseases and weaker immune systems, which may exacerbate the pathological effects of environmental NO\textsubscript{2} pollution [48]. However, one study showed a link between NO\textsubscript{2} exposure and the first episode of atrial fibrillation in young adults but not older adults [49]. NO\textsubscript{2} is a significant risk factor for sudden death and melancholia among younger individuals, which may explain these results. Considering the rapidly aging population and changes in disease patterns, public policymaking, risk assessment, and air pollution standards should be modified to reduce the impact of NO\textsubscript{2} on public health in China.

The highest risk of death due to NO\textsubscript{2} exposure was observed at a lag of 1 or 2 days and persisted for 4 days. However, some studies have observed the strongest link between NO\textsubscript{2} levels and the number of hospital admissions on day 0 [46]. Similar trends in the effects of a lag of NO\textsubscript{2} exposure have been reported for the mortality due to all causes, cardiovascular diseases, and respiratory diseases [13]. Moreover, the shape of the lag effects can be described by 2 different patterns. The first is that the NO\textsubscript{2} levels recorded on the previous 2 days have a greater effect on mortality due to all causes, nonaccidental causes, circulatory diseases, respiratory diseases, endocrine diseases, and tumors. The second is that mortality resulting from digestive, nervous, and urogenital diseases as well as external systems tends to be more evenly distributed during the first 6 days. These inconsistencies may be a result of differences in the biological mechanisms that underlie the biological effects of NO\textsubscript{2}.

We also found that the impact of NO\textsubscript{2} exposure was higher on cold days, which is consistent with the findings of previous studies. For instance, studies conducted in Shanghai and Wuhan have shown that the effects of NO\textsubscript{2} are highest in the winter [50,51]. Moreover, other research groups in China reported no substantial variation in the effects of NO\textsubscript{2} on mortality due to nonaccidental causes and cardiovascular diseases across different seasons [23,52,53]. However, a large time series analysis conducted in the northeast United States reported a peak mortality risk from NO\textsubscript{2} exposure in the summer [54].
Likewise, an Italian study conducted between April and September also found a stronger association between NO$_2$ and mortality due to natural causes, cardiac diseases, and respiratory diseases [14]. A meta-analysis further demonstrated that NO$_2$ had a greater impact on hospital admissions and mortality in hot weather (May 1 to September 30) [55]. One possible explanation is that people are exposed to higher levels of NO$_2$ during the warmer periods of the year due to greater involvement in outdoor activities [14]. Furthermore, the production of NO$_2$ also increases at higher temperatures [56]. Another reason is that heat may promote thrombosis by increasing blood viscosity and secondary cholesterol levels [57]. Finally, individual susceptibility to air pollutants may increase in the summer [58]. These inconsistent results may be due to the influence of concentrations of NO$_2$ components, regional gaseous pollutants, climatic conditions, resident exposure patterns, socioeconomic characteristics, and different data analysis methods.

This study has several salient points. First, the study showed that NO$_2$ was related not only to all-cause mortality but also to cause-specific mortality. Second, the relationship between NO$_2$ and mortality due to different causes and the potential modifying effects of gender, age, and season were analyzed. These findings may have important implications for developing targeted strategies to protect vulnerable populations from the harmful effect of NO$_2$. The detrimental effects of NO$_2$ can be minimized by reducing time spent outdoors and physical exertion, wearing a mask when going out, avoiding heavily trafficked roads, and increasing immunity [59].

Nevertheless, there are several limitations that need to be considered. First, our study was limited to the Guangzhou area, and the results cannot be generalized to other regions due to differences in the population structure and environmental characteristics. Second, environmental monitoring data may only indicate population averages rather than individual levels, which may have introduced inevitable measurement bias [60]. Third, the classification of the cause of death was based only on ICD-10 codes and does not detail the actual cause of death of the patient, which may also lead to bias. Finally, causal relationships and the underlying pathophysiology of our findings could not be ascertained and warrant further research.

**Conclusions**

In summary, our study comprehensively examined the association between NO$_2$ exposure and a spectrum of causes for the risk of mortality. Through subgroup analysis, we found that gender did not significantly modify the NO$_2$-related mortality risks, while older adults were more susceptible to death due to all causes, nonaccidental causes, and cardiovascular diseases when exposed to NO$_2$. Furthermore, the association between NO$_2$ and cause-specific mortality is stronger in the warm season. We found a relatively higher effect of NO$_2$ on health on days with higher temperatures, relative humidity, and wind speed but lower pressure levels, though the between-group differences were not statistically significant. Considering heterogeneity in socioeconomic characteristics and population structure among different regions, as well as the limitations of the inaccuracy of monitoring data and measurement error of ICD classifications, our findings still need to be confirmed in other regions. Our study provides new evidence to develop prevention-oriented health policies, highlighting that it is necessary to strengthen air quality standards to protect public health from NO$_2$ pollution.

**Acknowledgments**

The study was supported by Guangzhou Science and Technology Program Project (No. 2023A03J0284) and Guangdong Province Medical Science and Technology Research Project (No. A2023042).

**Authors' Contributions**

JY conceptualized the study, acquired the funding, and supervised the study. GL and HD provided the resources for the study. GL, HD, and HR performed the investigation. JZ performed the formal analysis. JZ created the visualizations. JZ wrote the original manuscript draft. GL, HD, ML, HR, and JY reviewed and edited the manuscript draft.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Supplementary tables and figures.

[DOCX File, 1482 KB - publichealth_v10i1e44648_app1.docx]

**References**


Abbreviations

DLNM: distributed lag nonlinear model
ICD-10: 10th Revision of the International Classification of Diseases
NS: natural cubic spline function
PM: particulate matter

©Jie Zeng, Guozhen Lin, Hang Dong, Mengmeng Li, Honglian Ruan, Jun Yang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 05.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Eating Habits and Lifestyle Factors Related to Childhood Obesity Among Children Aged 5-6 Years: Cluster Analysis of Panel Survey Data in Korea

Heemoon Lim1*, MSN; Hyejung Lee2*, PhD

1College of Nursing, Yonsei University, Seoul, Republic of Korea
2Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea
*all authors contributed equally

Abstract

Background: Childhood obesity has emerged as a major health issue due to the rapid growth in the prevalence of obesity among young children worldwide. Establishing healthy eating habits and lifestyles in early childhood may help children gain appropriate weight and further improve their health outcomes later in life.

Objective: This study aims to classify clusters of young children according to their eating habits and identify the features of each cluster as they relate to childhood obesity.

Methods: A total of 1280 children were selected from the Panel Study on Korean Children. Data on their eating habits (eating speed, mealtime regularity, consistency of food amount, and balanced eating), sleep hours per day, outdoor activity hours per day, and BMI were obtained. We performed a cluster analysis on the children’s eating habits using k-means methods. We conducted ANOVA and chi-square analyses to identify differences in the children’s BMI, sleep hours, physical activity, and the characteristics of their parents and family by cluster.

Results: At both ages (ages 5 and 6 years), we identified 4 clusters based on the children’s eating habits. Cluster 1 was characterized by a fast eating speed (fast eaters); cluster 2 by a slow eating speed (slow eaters); cluster 3 by irregular eating habits (poor eaters); and cluster 4 by a balanced diet, regular mealtimes, and consistent food amounts (healthy eaters). Slow eaters tended to have the lowest BMI (P<.001), and a low proportion had overweight and obesity at the age of 5 years (P=.03) and 1 year later (P=.005). There was a significant difference in sleep time (P=.01) and mother’s education level (P=.03) at the age of 5 years. Moreover, there was a significant difference in sleep time (P=.03) and the father’s education level (P=.02) at the age of 6 years.

Conclusions: Efforts to establish healthy eating habits in early childhood may contribute to the prevention of obesity in children. Specifically, providing dietary guidance on a child’s eating speed can help prevent childhood obesity. This research suggests that lifestyle modification could be a viable target to decrease the risk of childhood obesity and promote the development of healthy children. Additionally, we propose that future studies examine long-term changes in obesity resulting from lifestyle modifications in children from families with low educational levels.

(JMIR Public Health Surveill 2024;10:e51581) doi:10.2196/51581

https://publichealth.jmir.org/2024/1/e51581
**KEYWORDS**
BMI; body mass index; childhood obesity; cluster analysis; healthy eating; healthy lifestyle; pediatric obesity; preschool child; prevention; unsupervised machine learning

**Introduction**

Childhood obesity has emerged as a major health issue due to the rapid growth in the prevalence of obesity among young children and the higher risk of developing cardiovascular and metabolic diseases in adulthood [1,2]. To address these health problems, childhood obesity has been studied for decades, and great efforts have been made to identify and characterize potential predictors of childhood obesity [3]. However, more studies are needed to understand the factors involved and their complex relationship with the development of childhood obesity [4].

Obesity can be caused by a combination of biological factors such as an individual’s genes, insulin resistance, disease, and metabolic processes, as well as socioeconomic factors such as the surrounding family and environment leading to obesity-related behaviors [5-7]. Although, fundamentally, excessive energy due to an imbalance between energy intake and consumed energy is known to cause fat formation and obesity, Davison and Birch [8] explain the various causes of childhood obesity as a micro- and macrosystem surrounding the child and provide evidence of the need for great efforts to change behavior to improve children’s health.

Eating habits affect dietary intake and obesity through various behaviors such as meal frequency, amount, speed, and snacking habits [9]. A prospective cohort study in which eating habits were measured repeatedly confirmed that there were individual differences in the development of food enjoyment and satiety responsiveness, which affect eating habits after the age of 4 years. These results suggest that eating habits are dynamic behaviors in the first years of life and may change beyond preschool age [10]. The GUSTO study measured the eating habits of children aged 5 and 6 years and found that obesity and overweight in children were related to rapid eating speed [11]. Therefore, understanding the early-life factors that influence these behaviors may help identify areas for intervention to curb the progression of being overweight or obese in children [12].

In terms of obesity prevention, the period of childhood before the age of 5 years is very important as an opportunity to establish new behaviors rather than change existing ones that have become entrenched in adulthood, which presents a difficult challenge [13]. Children’s eating habits begin with solid foods at the age of 3-6 months. From that time until the age of 5 years, preschool children learn autonomous eating habits from their parents and form eating habits based on their own preferences and previous experiences [14]. Additionally, preschool children aged 5 years or younger who are obese are more likely than children with a normal weight to become overweight during adolescence and are 5 times more likely to become obese as adults. Thus, prevention through healthy lifestyle habits early in childhood is important [15]. The importance of these early childhood lifestyle habits is highlighted by the World Health Organization’s guidelines for children’s health, which also discuss the importance of forming lifestyle habits in children before the age of 5 years [16].

Establishing a healthy lifestyle early in life is important to improve health outcomes later [17]. A recent literature review on childhood obesity revealed that, to prevent childhood obesity, changes need to be made in children’s overall lifestyle, including their daily living habits, rather than limiting management to food intake [18,19]. However, a number of studies generally recommend limiting the intake of high-calorie foods, sugary drinks, and fast foods and eating more fruit and vegetables to prevent childhood obesity [20-22]. As an eating practice guideline, dietary habits, such as eating breakfast, balanced eating, and eating slowly, are recommended during mealtimes, but studies on the relationship between these eating habits and early childhood obesity are limited [23,24]. Additionally, previous obesity research using machine learning explored the relationship between demographic factors, some behaviors, and childhood obesity but had limitations due to single cross-sectional methods and small sample sizes [25,26]. Therefore, this study is designed to identify characteristic patterns of preschool children’s eating habits using unsupervised machine learning techniques and to determine the impact of these eating habits on children’s BMI. Our results provide evidence that can guide healthy eating habits to prevent childhood obesity.

**Methods**

**Study Design and Data**

This study used data from the Panel Study on Korean Children (PSKC), which was designed to follow a sample of children from 2008 to 2027 to confirm the impact of families and communities on children’s growth and development. The PSKC is a nationally representative sample using stratified sampling that considers all the regions in South Korea. For this panel survey, parents with children born between April and July 2008 were recruited from 30 hospitals. In the first survey, a total of 2150 parents participated in face-to-face interviews and completed a self-administered questionnaire [27]. However, only 1280 children were included in this study because they had both the sixth and seventh surveys of the PSKC. For this study, the sixth data set (at the age of 5 years) and seventh data set (at the age of 6 years) were obtained for the data analysis after excluding missing and incomplete data (Multimedia Appendix 1). The data used for this study is considered a representative sample of national data in terms of the national demographics (male 649/1280, 50.7% and female 631/1280, 49.3%) and prevalence of childhood obesity (overweight 122/1280, 9.5% and obese 54/1280, 4.2%) [28].

**Measurements**

**Eating Habits**

Eating habits were assessed based on four questions that mothers (or fathers) were asked to answer: “Is your child’s eating speed
fast?” “Does your child have meals at regular times?” “Is the amount of food your child eats consistent?” and “Does your child eat all kinds of food?” Responses to each question were assessed using a 5-point Likert scale ranging from “not at all” to “agree very strongly.” A higher score indicated a greater tendency in the diet habit.

**BMI**

Obesity status, the primary outcome of this study, was defined according to BMI, which was calculated using the children’s weight and height [29]. The categories, such as normal, overweight, and obese, were defined based on the Korean child growth chart: children in the 85 to 95 percentile were categorized as overweight, and those over the 95 percentile were categorized as obese [30,31].

**Physical Activity**

For the children’s activity levels, we used hours spent in outdoor activity as perceived by their mothers. They calculated the average number of hours their child spent daily on outdoor activities.

**Sleep Duration**

For the children’s sleep hours, we used the average amount of sleep time as perceived by their mothers. The child’s average sleep time at night was calculated as the difference between the mother’s reported bedtime and wake-up time.

**Characteristics of Parents and Family**

Parental age, education level, and employment status were obtained as parent characteristics; the number of family members and family income (Korean won per month) were obtained as family characteristics.

**Ethical Considerations**

This study was approved by the Hospital Ethics Committee, Seoul, South Korea (No. 4-2023-0418). The PSKC database was created with the voluntary consent of participants to investigate the growth and development of Korean children. If a participant decides to withdraw, they are excluded from the database. Digitally anonymized data sets were obtained after obtaining consent from PSKC in relation to the data. This study rigorously followed the guidelines recommended by the PSKC [32].

**Statistical Analyses**

All the continuous variables were tested for normality using the Shapiro-Wilk and Kolmogorov tests. The Shapiro-Wilk statistic was significant ($P<.001$), and the plots (regression of standardized residuals) showed no clear signs of violating the normality assumption [33].

Cluster analyses were performed in R (version 4.1.3; The R Project for Statistical Computing) using the packages “tidyverse,” “cluster,” “factoextra,” and “NbClust.” Clustering is an unsupervised machine learning technique to find natural groupings of participants based on a data set’s inherent structure. To identify the clusters, we used 4 eating habits (eating speed, mealtime regularity, food amount consistency, and balanced eating) as the input variables. For the ordinal variables measured on a 5-point Likert scale, we scaled by considering the means and SDs of the variables [26]. Principal component analysis was used to check the data distribution and independence of the 4 eating habits. Before the data set was considered significantly clusterable, the Hopkins statistic was applied iteratively using a threshold of 0.5, and the data set was confirmed to be above the threshold. The clustering analysis was performed by applying 2 hierarchical clustering methods (agglomeration and division) and Ward’s approach based on Euclidean distance and $k$-means [34]. We used the R NbClust package to explore the optimal number of clusters in our data set by varying all combinations of cluster number, distance measure, and clustering method and considered the Elbow method ($k=4$) and Scott index ($k=4$) for optimal cluster selection (Multimedia Appendix 2) [35]. Finally, the number of clusters ($k=4$) was selected by visually inspecting the data (clusters 1, 2, 3, and 4). After selecting the number of clusters, clusters were formed by repeating the $k$-means algorithm, which is most commonly used in unsupervised machine learning techniques, until the center value of the cluster did not change. Once the clusters were identified [35], a radar chart was created to explore the functionality of the final clusters (Figure 1).
A 1-way ANOVA with a Bonferroni posthoc comparison and a chi-square test were used to assess the group differences [36]. The power of this study was calculated using G^*power (version 3.1) with a 95% degree of confidence, and the number of participants and study design were considered through a comparison of the differences between clusters. Statistical significance was defined as a 2-sided $P<.05$.

**Results**

**Children’s Characteristics**

Of the children analyzed, 50.7% (649/1280) were boys, and the mean birth weight was 3.26 (SD 0.41) kg. The average baseline BMI was 15.99 (1.60) kg/m$^2$ at the age of 5 years and 16.20 (2.01) kg/m$^2$ at the age of 6 years. At the age of 5 years, the children’s average time spent on outdoor activities was 1.14 (0.81) hours per day, and the average sleep duration per day was 9.87 (0.73) hours. The children’s average time spent on outdoor activities was 1.08 (0.71) hours per day, and the average sleep duration per day was 9.76 (0.68) hours at the age of 6 years.

A bachelor’s degree was the most prevalent education level for both mothers (485/1280, 37.89%) and fathers (552/1280, 43.13%). Most of the fathers (1226/1280, 96.09%) were employed, and the average family income was ₩4,275,100 (US $3186.64) per month (Multimedia Appendix 3).

**Cluster Developed by Eating Habits**

The analysis identified 4 clusters, and their characteristics are similar on a radar chart (Figure 1). Cluster 1 (fast eaters) is characterized by a high eating speed and represents 512 children (aged 5 years) and 440 children (aged 6 years); cluster 2 (slow eaters) represents 293 five- and 415 six-year-old children with a slow eating speed. Cluster 3 (poor eaters) represents 283 children (aged 5 years) and 243 children (aged 6 years) with irregular mealtimes, inconsistent food amounts, and imbalanced eating habits, and cluster 4 (healthy eaters) represents 192 children (aged 5 years) and 182 children (aged 6 years) with regular mealtimes, consistent food amounts, and balanced eating (Table 1).

**Table 1.** Comparison of the children’s eating habits by cluster (N=1280).

<table>
<thead>
<tr>
<th>Eating habit</th>
<th>At 5 years old, mean (SD)</th>
<th>At 6 years old, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fast eater (n=512)</td>
<td>Slow eater (n=293)</td>
</tr>
<tr>
<td>Eating speed</td>
<td>3.24 (0.48)</td>
<td>1.94 (0.50)</td>
</tr>
<tr>
<td>Regularity of mealtime</td>
<td>3.91 (0.39)</td>
<td>3.97 (0.50)</td>
</tr>
<tr>
<td>Consistency of food amount</td>
<td>3.85 (0.38)</td>
<td>3.72 (0.50)</td>
</tr>
<tr>
<td>Balanced eating</td>
<td>3.65 (0.66)</td>
<td>2.76 (0.82)</td>
</tr>
</tbody>
</table>

**Cluster Changes From 5 to 6 Years of Age**

Changes in clusters according to children’s eating habits are shown in Table 2. Of the total 1028 children, 553 (53.8%) remained in the same eating habits cluster as classified as being aged 5 years. A total of 44.7% (229/512) of children were in the fast eater group, 52.6% (154/293) in the slow eater group, 38.2% (108/283) in the poor eater group, and 32.3% (62/192) in the healthy eater group remained in the same cluster a year later.
Table 2. The cluster changes from 5 to 6 years of age. The bar graph represents the number of participants in the clusters at 5 years old who moved to another cluster at 6 years old.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency, n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fast eaters</strong></td>
<td></td>
</tr>
<tr>
<td>Maintain</td>
<td>229/512 (44.7)</td>
</tr>
<tr>
<td>Slow eater</td>
<td>131/512 (25.6)</td>
</tr>
<tr>
<td>Poor eater</td>
<td>68/512 (13.3)</td>
</tr>
<tr>
<td>Healthy eater</td>
<td>84/512 (16.4)</td>
</tr>
<tr>
<td><strong>Slow eaters</strong></td>
<td></td>
</tr>
<tr>
<td>Maintain</td>
<td>154/293 (52.6)</td>
</tr>
<tr>
<td>Fast eater</td>
<td>66/293 (22.5)</td>
</tr>
<tr>
<td>Poor eater</td>
<td>52/293 (17.8)</td>
</tr>
<tr>
<td>Healthy eater</td>
<td>21/293 (7.2)</td>
</tr>
<tr>
<td><strong>Poor eaters</strong></td>
<td></td>
</tr>
<tr>
<td>Maintain</td>
<td>108/283 (38.2)</td>
</tr>
<tr>
<td>Fast eater</td>
<td>78/283 (27.6)</td>
</tr>
<tr>
<td>Slow eater</td>
<td>82/283 (29)</td>
</tr>
<tr>
<td>Healthy eater</td>
<td>15/283 (5.3)</td>
</tr>
<tr>
<td><strong>Healthy eaters</strong></td>
<td></td>
</tr>
<tr>
<td>Maintain</td>
<td>62/192 (32.3)</td>
</tr>
<tr>
<td>Fast eater</td>
<td>69/192 (35.9)</td>
</tr>
<tr>
<td>Slow eater</td>
<td>48/192 (25)</td>
</tr>
<tr>
<td>Poor eater</td>
<td>13/192 (6.8)</td>
</tr>
</tbody>
</table>

The cluster changes from 5 to 6 years of age in the relevant cluster.

**Characteristic Differences by Clusters at the Age of 5 Years**

Among children’s characteristics, there was a significant difference in BMI at the age of 5 years between groups \((P<.001)\). Fast eaters also had the highest BMI at the age of 5 years (mean 16.17 kg/m\(^2\)); slow eaters had the lowest BMI (mean 15.59 kg/m\(^2\)). The proportion of children with obesity differed significantly between groups \((P=.03)\). A higher proportion of children with obesity, based on BMI at the age of 5 years, was reported among poor eaters (16/283, 5.7%) and fast eaters (28/512, 5.5%). There was a significant difference in sleep duration; fast eaters were associated with longer sleep duration (9.93 hours per day) than slow eaters \((P=.005)\). Among parental and family characteristics, healthy eaters had a higher proportion of mothers with master’s degrees or higher, and poor eaters had a higher proportion of mothers who were high school graduates or lower \((P=.03)\). There were no significant differences in the time children spent outdoors, family income, or parents’ employment status (Table 3).
Table 3. Differences in characteristics by cluster at the age of 5 years (N=1280).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fast eater (n=512)</th>
<th>Slow eater (n=293)</th>
<th>Poor eater (n=283)</th>
<th>Healthy eater (n=192)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Posthoc test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristics of child</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>261 (51)</td>
<td>146 (49.8)</td>
<td>145 (51.2)</td>
<td>97 (50.5)</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>251 (49)</td>
<td>147 (50.2)</td>
<td>138 (48.8)</td>
<td>95 (49.5)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Birth weight (kg), mean (SD)</td>
<td>3.28 (0.40)</td>
<td>3.22 (0.39)</td>
<td>3.26 (0.42)</td>
<td>3.28 (0.42)</td>
<td>.18</td>
<td>N/A</td>
</tr>
<tr>
<td>BMI at the age of 5 years (kg/m&lt;sup&gt;2&lt;/sup&gt;), mean (SD)</td>
<td>16.17 (1.63)</td>
<td>15.59 (1.31)</td>
<td>16.03 (1.76)</td>
<td>16.06 (1.61)</td>
<td>&lt;.001</td>
<td>a,c,d&gt;b</td>
</tr>
<tr>
<td>Overweight, n (%)</td>
<td>53 (10.4)</td>
<td>21 (7.2)</td>
<td>25 (8.8)</td>
<td>23 (12)</td>
<td>.03</td>
<td>N/A</td>
</tr>
<tr>
<td>Obese, n (%)</td>
<td>28 (5.5)</td>
<td>4 (1.4)</td>
<td>16 (5.7)</td>
<td>6 (3.1)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Physical activity (hour/day), mean (SD)</td>
<td>1.16 (0.83)</td>
<td>1.17 (0.81)</td>
<td>1.06 (0.81)</td>
<td>1.13 (0.75)</td>
<td>.37</td>
<td>N/A</td>
</tr>
<tr>
<td>Sleep duration (hour/day), mean (SD)</td>
<td>9.93 (0.71)</td>
<td>9.79 (0.73)</td>
<td>9.79 (0.74)</td>
<td>9.93 (0.76)</td>
<td>.01</td>
<td>a&gt;b</td>
</tr>
<tr>
<td><strong>Characteristics of mother</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>35.9 (3.46)</td>
<td>36.0 (3.43)</td>
<td>35.9 (3.74)</td>
<td>36.3 (3.42)</td>
<td>.77</td>
<td>N/A</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
<td>N/A</td>
</tr>
<tr>
<td>High school or less</td>
<td>152 (29.7)</td>
<td>78 (26.6)</td>
<td>94 (33.2)</td>
<td>50 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College degree</td>
<td>132 (25.8)</td>
<td>96 (32.8)</td>
<td>78 (27.6)</td>
<td>50 (26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>210 (41)</td>
<td>99 (33.8)</td>
<td>100 (35.3)</td>
<td>76 (39.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree or higher</td>
<td>18 (3.5)</td>
<td>20 (6.8)</td>
<td>11 (3.9)</td>
<td>16 (8.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.06</td>
<td>N/A</td>
</tr>
<tr>
<td>Employed</td>
<td>206 (40.2)</td>
<td>124 (42.3)</td>
<td>111 (39.2)</td>
<td>97 (50.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>306 (59.8)</td>
<td>169 (57.7)</td>
<td>172 (60.8)</td>
<td>95 (49.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Characteristics of father</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>38.5 (3.97)</td>
<td>38.9 (3.70)</td>
<td>38.1 (4.0)</td>
<td>38.6 (3.79)</td>
<td>.10</td>
<td>N/A</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
<td>N/A</td>
</tr>
<tr>
<td>High school or less</td>
<td>146 (28.5)</td>
<td>74 (25.3)</td>
<td>83 (29.3)</td>
<td>44 (22.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College degree</td>
<td>100 (19.5)</td>
<td>62 (21.2)</td>
<td>65 (23)</td>
<td>28 (14.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>213 (41.6)</td>
<td>136 (46.4)</td>
<td>105 (37.1)</td>
<td>98 (51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree or higher</td>
<td>53 (10.4)</td>
<td>21 (7.2)</td>
<td>30 (10.6)</td>
<td>22 (11.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.57</td>
<td>N/A</td>
</tr>
<tr>
<td>Employed</td>
<td>489 (95.5)</td>
<td>284 (96.9)</td>
<td>268 (94.7)</td>
<td>185 (96.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>23 (4.5)</td>
<td>9 (3.1)</td>
<td>15 (5.3)</td>
<td>7 (3.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Characteristics of family, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of family members</td>
<td>4.29 (0.86)</td>
<td>4.27 (0.77)</td>
<td>4.20 (0.87)</td>
<td>4.25 (0.87)</td>
<td>.38</td>
<td>N/A</td>
</tr>
<tr>
<td>Income (₩10,000; US $7.45)</td>
<td>427.48 (227.85)</td>
<td>430.15 (173.44)</td>
<td>418.34 (211.31)</td>
<td>437.08 (175.79)</td>
<td>.79</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup>P value was calculated from ANOVA and chi-square test.

<sup>b</sup>N/A: not applicable.

**Characteristic Differences by Clusters at the Age of 6 Years**

Among children’s characteristics, there was a significant difference between groups in BMI at the age of 6 (<i>P</i>&lt;.001). Fast eaters also had the highest BMI at the age of 6 years (mean 16.55 kg/m<sup>2</sup>); slow eaters had the lowest BMI (mean 15.85 kg/m<sup>2</sup>). The proportion of children with obesity differed significantly between groups (<i>P</i>&lt;0.01). A higher proportion of children with obesity, based on BMI at the age of 6 years, was reported among fast eaters (40/440, 9.1%) and healthy eaters (16/182, 8.8%). There was a significant difference in sleep duration; healthy eaters were associated with longer sleep
duration (9.89 hours per day) than fast eaters ($P=.03$). Among parental and family characteristics, healthy eaters had a higher proportion of fathers with master’s degrees or higher, and poor eaters had a higher proportion of fathers who were high school graduates or lower ($P=.03$). There were no significant differences in the time children spent outdoors, family income, or parents’ employment status (Table 4).

Table 4. Differences in characteristics by cluster at the age of 6 years (N=1280).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fast eater (n=440)</th>
<th>Slow eater (n=415)</th>
<th>Poor eater (n=243)</th>
<th>Healthy eater (n=182)</th>
<th>$P$ value$^a$</th>
<th>Posthoc test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics of child</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>223 (50.7)</td>
<td>218 (52.5)</td>
<td>128 (52.7)</td>
<td>80 (44)</td>
<td></td>
<td>N/A$^b$</td>
</tr>
<tr>
<td>Female</td>
<td>217 (49.3)</td>
<td>197 (47.5)</td>
<td>115 (47.3)</td>
<td>102 (56)</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Birth weight (kg), mean (SD)</td>
<td>3.28 (0.39)</td>
<td>3.25 (0.40)</td>
<td>3.20 (0.41)</td>
<td>3.32 (0.45)</td>
<td>.02</td>
<td>d&gt;c</td>
</tr>
<tr>
<td>BMI at the age of 6 years ($kg/m^2$), mean (SD)</td>
<td>16.55 (2.11)</td>
<td>15.85 (1.74)</td>
<td>16.04 (2.01)</td>
<td>16.42 (2.17)</td>
<td>&lt;.001</td>
<td>a&gt;b;c; d&gt;b</td>
</tr>
<tr>
<td>Overweight, n (%)</td>
<td>45 (10.2)</td>
<td>28 (6.7)</td>
<td>28 (11.5)</td>
<td>22 (12.1)</td>
<td>.005</td>
<td>N/A</td>
</tr>
<tr>
<td>Obese, n (%)</td>
<td>40 (9.1)</td>
<td>17 (4.1)</td>
<td>12 (4.9)</td>
<td>16 (8.8)</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Physical activity (hour/day), mean (SD)</td>
<td>1.08 (0.69)</td>
<td>1.05 (0.69)</td>
<td>1.05 (0.74)</td>
<td>1.21 (0.75)</td>
<td>.07</td>
<td>N/A</td>
</tr>
<tr>
<td>Sleep duration (hour/day), mean (SD)</td>
<td>9.72 (0.67)</td>
<td>9.77 (0.69)</td>
<td>9.73 (0.69)</td>
<td>9.89 (0.68)</td>
<td>.03</td>
<td>d&gt;a</td>
</tr>
<tr>
<td>Characteristics of mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>35.8 (3.55)</td>
<td>36.0 (3.43)</td>
<td>35.9 (3.74)</td>
<td>36.3 (3.42)</td>
<td>.77</td>
<td>N/A</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
<td>N/A</td>
</tr>
<tr>
<td>High school or less</td>
<td>123 (28)</td>
<td>111 (26.7)</td>
<td>91 (37.4)</td>
<td>49 (26.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College degree</td>
<td>127 (28.9)</td>
<td>112 (27)</td>
<td>70 (28.8)</td>
<td>47 (25.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>169 (38.4)</td>
<td>167 (40.2)</td>
<td>73 (30)</td>
<td>76 (41.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree or higher</td>
<td>21 (4.8)</td>
<td>25 (6)</td>
<td>9 (3.7)</td>
<td>10 (5.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.21</td>
<td>N/A</td>
</tr>
<tr>
<td>Employed</td>
<td>206 (40.2)</td>
<td>124 (42.3)</td>
<td>111 (39.2)</td>
<td>97 (50.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>306 (59.8)</td>
<td>169 (57.7)</td>
<td>172 (60.8)</td>
<td>95 (49.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characteristics of father</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>N/A</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>38.5 (3.97)</td>
<td>38.9 (3.70)</td>
<td>38.1 (4.0)</td>
<td>38.6 (3.79)</td>
<td>.10</td>
<td>N/A</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>N/A</td>
</tr>
<tr>
<td>High school or less</td>
<td>116 (26.4)</td>
<td>112 (27)</td>
<td>81 (33.3)</td>
<td>38 (20.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College degree</td>
<td>93 (21.1)</td>
<td>72 (17.3)</td>
<td>59 (24.3)</td>
<td>31 (17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>185 (42)</td>
<td>190 (45.8)</td>
<td>86 (35.4)</td>
<td>99 (50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree or higher</td>
<td>46 (10.5)</td>
<td>41 (9.9)</td>
<td>17 (7)</td>
<td>22 (12.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.98</td>
<td>N/A</td>
</tr>
<tr>
<td>Employed</td>
<td>489 (95.5)</td>
<td>284 (96.9)</td>
<td>268 (94.7)</td>
<td>185 (96.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>23 (4.5)</td>
<td>9 (3.1)</td>
<td>15 (5.3)</td>
<td>7 (3.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characteristics of family, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of family members</td>
<td>4.26 (0.84)</td>
<td>4.25 (0.80)</td>
<td>4.28 (0.96)</td>
<td>4.30 (0.80)</td>
<td>.93</td>
<td>N/A</td>
</tr>
<tr>
<td>Income ($₩10,000; US $7.45)</td>
<td>440.38 (229.69)</td>
<td>416.80 (165.07)</td>
<td>407.74 (239.91)</td>
<td>440.00 (169.47)</td>
<td>.08</td>
<td>N/A</td>
</tr>
</tbody>
</table>

$^a$P value was calculated from ANOVA and chi-square test.

$^b$N/A: not applicable.
**Discussion**

**Overview**

Using a nationally representative sample, we identified 4 distinct clusters based on the eating habits of children aged 5 and 6 years. These children’s eating habits showed a pattern of 4 clusters a year later. However, in approximately half of the children, individual children changed their eating habit cluster after 1 year in this study, providing valuable insight into the timing of early obesity management [37]. Additionally, the higher proportion of children who had overweight and obesity and the higher BMI of children who ate quickly indicate that fast eating is associated with obesity. Children aged 5 years who are fast eaters need age-appropriate training to reduce their eating speed. Importantly, strategies to prevent progression from having overweight to having obesity in young children should be developed to improve children’s overall health status.

While a few studies have attempted to identify the relationship between eating habits and obesity in young children [4,23,38,39], none of them investigated eating habits concurrently, such as eating speed, balanced eating, mealtime regularity, and food amount consistency. In this study, we found that a large number of young children fell into the category of fast eaters and had a higher BMI. This relationship between fast eating habits and obesity can be explained by the mechanism that fast eating lowers satiety and consequently increases food intake by delaying the effects of brain signals and hormones [40]. The results of this study also suggest childhood obesity could be prevented by increasing eating time [39]. Regarding another eating habit related to childhood obesity, a recent systematic review identified mealtimes as a mechanism that explains obesity by affecting changes in metabolic efficiency, hormones, and gut microbiota throughout the day [14]. Paoli et al [41] emphasized the importance of regular eating times for obesity prevention through the regularity of fasting periods between meals.

Regarding the relationship between sleep duration and childhood obesity, insufficient sleep could contribute to the development of obesity through appetite, diet, and daytime activity levels [42,43]. These studies did not support previous reports of late bedtimes and short sleep duration in children with obesity [44]. However, in this study, sleep duration per cluster was more than 9 hours, so all the children had enough sleep, which limits exploration of the relationship between sleep duration and obesity in this study. Additional research on sleep and childhood obesity is recommended.

Among the risk factors for childhood obesity, a more important explanatory factor is sedentary time, such as TV watching time, rather than outdoor activity [45]. Similarly, this study found no difference in outdoor activity levels between the groups classified based on eating habits. The relationship between childhood obesity and physical activity levels should be explored by considering various activities, such as sedentary time and indoor activities, in addition to outdoor activities.

Besides children’s sleep duration and activity levels, the parents’ education level differed among the clusters. The proportion of parents with a high school diploma or lower was higher in the group of poor eaters than in the other eating groups. These results are similar to those of previous studies showing that mothers of children with overweight or obesity had a lower education level [4]. This suggests that parents’ level of education is a factor related to the development of childhood obesity. However, the single influence of the mother or father by age, rather than both parents, is thought to account for differences in parental influence as the child ages and requires further research [46]. The prevalence of childhood obesity is known to be high in low-income households and communities with low socioeconomic status, and the relationship between childhood obesity and families’ economic status has been well documented in large samples in several countries [42,47,48]. However, family income in this study did not differ between the clusters, which suggests that primary caregiver education is the key factor in forming eating habits to consider for the prevention of obesity in children [43].

**Limitations**

This study has several strengths and limitations that should be noted. Although the individual clusters had different BMI tendencies, the inclusion of only eating habit variables in the cluster analysis could limit our understanding of the development of childhood obesity since it omits interrelated variables such as physical activity and sleep. In addition, the results should be interpreted with caution because each eating habit was measured using a single item based on a Likert scale. However, we found evidence for the impact of dietary guidelines to prevent obesity in young children, including specific eating habits. Another limitation of this study was the reliance on parental reports. Nevertheless, the study is meaningful in that it used panel data that are representative longitudinal data of a country’s child population, and the derived results used machine learning techniques to solve complex phenomena targeting a relatively large sample of children.

**Conclusions**

Our results show that eating habits, such as eating speed, regularity of meals, meal amount consistency, and balanced eating habits, can be considered risk factors for developing childhood obesity. In addition, changed clusters by eating habits within 2 years in children highlights the need for early childhood obesity management. Besides eating habits, children’s sleep duration and maternal education levels differed significantly across the clusters. These findings suggest that a modification in lifestyle could be a good target to decrease the risk of childhood obesity and develop healthy children. In addition, we also propose that future studies examine long-term changes in obesity with lifestyle modification in children from families with low educational levels.
Acknowledgments
This work was supported by a 2022 Faculty Research Grant granted by Yonsei University College of Nursing (6-2022-0036).

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
H Lim and H Lee conceptualized the study and contributed to the study’s design. H Lim worked on the data curation and analysis. H Lim and H Lee drafted the manuscript with critical input. All the authors approved the final version of the manuscript and agreed to its publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Flowchart of the process of selecting study participants. PSKC: Panel Study on Korean Children.

Multimedia Appendix 2
Evidence of clustering decision. (A) Result of the dendrogram at the age of 5 years; (B) Result of the dendrogram at the age of 6 years; (C) Results of choosing the number of clusters at the age of 5 years; (D) Results of selecting the number of clusters at the age of 6 years. (A) and (B) visually show the hierarchical method as one of the methods for selecting the number of clusters. (C) and (D) are the results of using the NbClust package of the R program and are the process for selecting the number of clusters through the D index and elbow method.

Multimedia Appendix 3
Characteristics of samples at the age of 5 and 6 years (N=1280).

References


31. Panel Study on Korean Children (PSKC) data use manual for 1st-7th study. Korea Institute of Child Care and Education. URL: https://panel.kicce.re.kr/pskc/board/view.do?menu_id=4&board_id=33366&manage_id=27&old_menu_id=0&old_board_id=0&group_depth=0&parent_id=0&group_idx=0&group_ord=0&viewMode=NORMAL&search_type=title&search_text=&rowCount=10&viewPage=1 [accessed 2022-08-19]


Abbreviations

PSKC: Panel Study on Korean Children
Association Between Early-Life Exposure to Antibiotics and Development of Child Obesity: Population-Based Study in Italy

Anna Cantarutti1, PhD; Paola Rescigno2, MD; Claudia Da Borso1, MSc; Joaquin Gutierrez de Rubalcava Doblas3, MD; Silvia Bressan3, PhD, Prof Dr; Elisa Barbieri3, PhD; Carlo Giaquinto3, PhD, Prof Dr; Cristina Canova2, PhD, Prof Dr

1Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy
2Department of Cardio-Thoraco-Vascular Sciences and Public Health, University of Padova, Padova, Italy
3Department of Women’s and Children’s Health, University of Padova, Padova, Italy

Corresponding Author:
Anna Cantarutti, PhD
Department of Statistics and Quantitative Methods
University of Milano-Bicocca
Via Bicocca degli Arcimboldi, 8
Milano, 20126
Italy
Phone: 39 0264485828
Email: anna.cantarutti@unimib.it

Abstract

Background: Childhood obesity is a significant public health problem representing the most severe challenge in the world. Antibiotic exposure in early life has been identified as a potential factor that can disrupt the development of the gut microbiome, which may have implications for obesity.

Objective: This study aims to evaluate the risk of developing obesity among children exposed to antibiotics early in life.

Methods: An Italian retrospective pediatric population-based cohort study of children born between 2004 and 2018 was adopted using the Pedianet database. Children were required to be born at term, with normal weight, and without genetic diseases or congenital anomalies. We assessed the timing of the first antibiotic prescription from birth to 6, 12, and 24 months of life and the dose-response relationship via the number of antibiotic prescriptions recorded in the first year of life (none, 1, 2, and ≥3 prescriptions). Obesity was defined as a BMI z score >3 for children aged ≤5 years and >2 for children aged >5 years, using the World Health Organization growth references. The obese incidence rate (IR) × 100 person-years and the relative 95% CI were computed using infant sex, area of residence, preschool and school age, and area deprivation index, which are the covariates of interest. A mixed-effect Cox proportional hazards model was used to estimate the hazard ratio and 95% CI for the association between antibiotic exposure in early life and child obesity between 24 months and 14 years of age, considering the family pediatricians as a random factor. Several subgroup and sensitivity analyses were performed to assess the robustness of our results.

Results: Among 121,540 children identified, 54,698 were prescribed at least an antibiotic within the first year of life and 26,990 were classified as obese during follow-up with an incidence rate of 4.05 cases (95% CI 4.01-4.10) × 100 person-year. The risk of obesity remained consistent across different timings of antibiotic prescriptions at 6 months, 1 year, and 2 years (fully adjusted hazard ratio [aHR] 1.07, 95% CI 1.04-1.10; aHR 1.06, 95% CI 1.03-1.09; and aHR 1.07, 95% CI 1.04-1.10, respectively). Increasing the number of antibiotic exposures increases the risk of obesity significantly (P trend<.001). The individual-specific age analysis showed that starting antibiotic therapy very early (between 0 and 5 months) had the greatest impact (aHR 1.12, 95% CI 1.08-1.17) on childhood obesity with respect to what was observed among those who were first prescribed antibiotics after the fifth month of life. These results were consistent across subgroup and sensitivity analyses.

Conclusions: The results from this large population-based study support the association between early exposure to antibiotics and an increased risk of childhood obesity. This association becomes progressively stronger with both increasing numbers of antibiotic prescriptions and younger age at the time of the first prescription.

(JMIR Public Health Surveill 2024;10:e51734) doi:10.2196/51734
KEYWORDS
childhood obesity; BMI z score; pediatric population-based; antibiotics; real-world data; association; exposure; child obesity; obesity; population-based; gut microbiome; early life; pediatric; prescription

Introduction

Childhood obesity is a significant public health problem representing the most severe challenge in the world that can have long-term consequences [1]. In the United States, 1 in 3 children and adolescents are overweight or obese [2]. In Europe, the prevalence has been increasing over the past few decades. According to the World Health Organization (WHO) and the European Childhood Obesity Surveillance Initiative, 38% of Italian children were found to be overweight (including obesity), with boys more affected than girls (41% and 35%, respectively). Moreover, the overall prevalence of obesity in Italy was 16% and more common in boys (20%) than girls (13%) [3].

Obesity is a chronic and complex disease characterized by excessive adiposity that can impair health, leading to various health complications, including an increased risk of developing chronic conditions such as type 2 diabetes, cardiovascular diseases, and mental health issues. It can also have social and emotional consequences, such as low self-esteem and stigmatization [4]. Considering these dramatic consequences, it is crucial to implement interventions to prevent the occurrence of this condition in children, which is also the goal of Health4EUkids, the European Joint Action for the implementation of Best Practices for the promotion of health and the prevention of obesity [5].

Early childhood obesity has been associated with several factors, such as maternal prepregnancy BMI, nutritional intake, physical activity, sleep duration, and screen time [6]. In particular, sleep deprivation can disrupt hormones that regulate appetite and metabolism, and excessive screen time can contribute to early childhood obesity by reducing physical activity and promoting unhealthy eating habits. In addition, emerging evidence suggests that the composition and function of the gut microbiome can influence various aspects of human health, including energy metabolism and body weight regulation. Previous studies demonstrated that the intestinal microbiome plays an important role in host energy metabolism, including gene expression that impacts energy availability from short-chain fatty acids and the processing of otherwise indigestible polysaccharides [2]. The microbial ecosystem begins taxonomic diversification at birth and completes its development during the early years of life [7]. The establishment and maturation of the gut microbiome are influenced by a complex interplay of internal and external factors, including environmental factors, the type of delivery (natural birth or cesarean delivery), and diet. While the gut microbiome is relatively stable, it can experience periods of acute or chronic perturbation in certain disease states or due to new exposures such as antibiotic therapies [8]. Antibiotic exposure in early life has been identified as a potential factor that can disrupt the development and composition of the gut microbiome, which may have implications for obesity later in life [7]. Previous epidemiological studies have provided valuable insights into the relationship between the gut microbiome, antibiotic exposure in early life, and obesity [2]. Early-life antibiotic exposures can modify the bacterial diversity of the intestinal microbiome in infants and delay microbiota maturation. These effects were most pronounced with antibiotic exposure during the first year of life, while no significant effect was observed with later exposures [9]. However, it is important to note that other confounding factors, such as genetic predisposition, diet, lifestyle, and socioeconomic factors, could contribute to both antibiotic use and obesity risk. Furthermore, the specific antibiotics used, the timing and duration of exposure, and the individual's age at the time of exposure may influence the observed associations.

The objective of this study is to examine the risk of developing obesity among children exposed to antibiotics early in life in a large population-based Italian birth cohort with a detailed assessment of antibiotic use and long-term follow-up to assess the development of obesity identified through standardized anthropometric measurements.

Methods

Study Population

We used data from Pedianet [10], an independent network of more than 400 family pediatricians (FPs) established in 1998, to collect information from outpatient routine clinical care in Italy; detailed information was explained elsewhere [11]. In particular, in this study, we used information regarding demographic data, prescriptions (pharmaceutical prescriptions identified by the Anatomical Therapeutical Chemical code), and growth parameters. In Italy, FPs have been considered by the Ministry of Health as physicians responsible for performing regular “mandatory” well-child visits for preventive medicine purposes at specific time points during which the child’s anthropometric measurements are recorded [12].

We identified all children born between 2004 and 2018, followed from birth to at least 4 years (maximum follow-up of 14 years). Children were required to have at least 2 visits during the first 2 years of life at least 6 months apart [13], born at term (≥27 gestational weeks) and with birth weight greater than 2500 g, and without genetic diseases (ie, achondroplasia, Cornelia de Lange, Down, Prader-Willi, Turner, and Williams syndromes [14-21]) or congenital anomalies. The final cohort consisted of 121,540 children (Figure S1 in Multimedia Appendix 1).

Ethical Considerations

This is an observational, retrospective, noninterventional study. According to a bylaw on the classification and implementation of observational drug-related research, as issued by the Italian National Drug Agency (an entity belonging to the Italian Ministry of Health), this study does not require approval by an ethics committee in Italy (Italian Drug Agency note on August 3, 2007). This study was conducted in accordance with the tenets of the Declaration of Helsinki and was compliant with the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance’s Guide on Methodological Standards in
Pharmacoepidemiology. Data generated during routine patient care were collected and handled anonymously, in compliance with Italian regulations, and stored under a unique numerical identifier. Ethical approval of the study and access to the database was approved by the Internal Scientific Committee of So.Se.Te. Srl, the legal owner of Pedianet.

Infant Antibiotic Exposure Assessment

Antibiotic exposure (Anatomical Therapeutical Chemical code: J01.*) was assessed using the outpatient prescription data recorded during the primary care visits. We assessed (1) the timing of the first antibiotic prescription from birth to 6, 12 (ie, the primary exposure of interest), and 24 months of life (Table S1 in Multimedia Appendix 1) and (2) the dose-response relationship via the number of antibiotic prescriptions recorded in the first year of life (none, 1, 2, and ≥3 prescriptions). Moreover, specific classes of antibiotics were categorized based on the spectrum of action (ie, narrow and broad-spectrum antibiotics) and independent assessment (ie, penicillins, cephalosporins, macrolides, and others; Table S2 in Multimedia Appendix 1) [22].

Child Obesity Assessment

Data for weight and height were retrieved from the growth parameters recorded during the primary care visits and ascertained longitudinally from 2 years until the last parameter recorded at the end of the study or the completion of the 14th year of life. BMI was calculated as weight (kg)/length (m²) and transformed into age- and sex-specific z score using the WHO Growth References [23-25]. Obesity was defined as a z score >3 SDs above the mean for children aged ≤5 years and a z score >2 SDs for children aged >5 years [26]. Children without growth parameters recorded from the second birthday or with implausible BMI z score values (ie, z score <4 and z score >8) were excluded from the analysis [27]. We also performed a sensitivity analysis where obesity was redefined based on the Centers for Disease Control and Prevention (CDC) Growth Charts, identifying children as obese for values of the BMI z score >2 [28].

Covariates

Additional factors such as the birth year, infant sex, area of residence (ie, north, center, south, and islands of Italy), the area deprivation index (ADI — least deprived: ADI 2, 3, 4, 5—most deprived; and ADI missing [29]), and the FP were considered of interest. The ADI is based on 5 items that recurrently describe social and material deprivation and is categorized in quintiles based on the regional ADI level to ensure within-region appropriately represented categories.

Statistical Analysis

Demographic characteristics were summarized through frequency and percentages and compared obese with nonobese children. The chi-square test was used to assess differences. Moreover, the obese incidence rate (IR) per 100 person-year (PY) and the relative 95% CI were computed using infant sex, area of residence, preschool (ie, obese within ≤5 years of age), school-aged children (ie, obese over >5 years of age), and ADI.

A mixed-effect Cox proportional hazards model was used to estimate the hazard ratio and 95% CI for the association between antibiotic exposure within the first year of life (exposure at 1 year) and child obesity between 24 months and 14 years of age, considering the FPs as a random factor. The proportional hazard assumption for the time-fixed covariates was tested using Schoenfeld residuals [30]. The follow-up began after 2 years of age and ended with the last anthropometric measure available at the end of the study (ie, December 31, 2022), the completion of the 14th year of life, or the end of pediatric assistance. We performed 2 levels of adjustment, including (1) infant sex, area of residence, and year of birth, and (2) infant sex, area of residence, year of birth, and ADI as categorical variables (considering missing as the sixth category).

In addition, we evaluated the association between (1) the timing of the first antibiotic prescription from birth to 6 (exposure at 6 months) and 24 (exposure at 2 years) months, (2) the dose-response relationship by assessing the number of antibiotic prescriptions recorded at 1 year, and (3) the individual-specific age of the antibiotic therapy initiation by assessing the age at the first prescription within 1 year (0–<5, 5–<8, and 8–12 months) and the development of childhood obesity. The SAS software (version 9.4; SAS Institute) and R (R Foundation for Statistical Computing) were used for the analyses. For all hypotheses tested, 2-tailed P < .05 were considered to be significant.

Subgroup and Sensitivity Analyses

We also performed several subgroup analyses. We assessed the association between child obesity with (1) the type of antibiotics prescribed in the first year based on the spectrum of action (ie, narrow and broad-spectrum antibiotics vs none) and the antibiotic class (ie, penicillins, cephalosporins, macrolides, and others vs none), (2) the preschool and school age children, and (3) the infant sex.

We performed several sensitivity analyses to examine the robustness of our findings. We repeated the main analysis (1) including only children with a birth weight within the normal range (ie, not exceeding 4000 g; n = 41,443, 34.4% excluded), (2) excluding from the cohort those children with missing values for the ADI (n = 16,582, 13.6% excluded), (3) restricting the cohort to only those children resident in the Veneto region since it is the most representative region of Italy in Pedianet (n = 49,772, 41% included), and (4) redefining the outcome variable calculating the BMI z score using the CDC Growth Charts.

Results

Description of the Cohort

Among 121,540 children identified in the Pedianet network within birth, 45% (54,698/121,540) were prescribed at least 1 antibiotic within the first year of life. The median follow-up period was approximately 6 (IQR 5–10) years.

Of the overall cohort, 22% (26,990/121,540) were classified as obese during follow-up, with an IR of 4.05 cases (95% CI 4.01–4.10) × 100 PY (Table 1). The IR of obesity was quite similar among male and female participants (IR 3.96 vs 4.16 × 100 PY, respectively), more than double in school-aged than school-aged children (ie, obese over >5 years of age), and ADI.
preschool-aged children (IR 3.68 vs 1.48 × 100 PY, respectively), and higher in children from the south and islands (IR 3.56, 4.01, and 5.16 × 100 PY, respectively for children from north, center, and south of Italy). Moreover, the incidence of obesity increased with increasing deprivation index (IR 3.69 vs 4.23×100 PY for least and most deprived children, respectively; Table 1).

Among children without antibiotic prescription during their first year of life, 21% (13,956/66,842) were observed to be obese, while 24% (13,034/54,698) of those with antibiotic exposure were obese, with an IR of 3.82 × 100 PY (95% CI 3.76-3.89) and 4.34 × 100 PY (95% CI 4.26-4.41), respectively. The incidence of obesity was higher among antibiotic-exposed children than unexposed children across all baseline characteristics considered (Table 2; Table S3 in Multimedia Appendix 1).

### Table 1. Baseline characteristics of children within the cohort and incidence rate of obesity among characteristics of interest (Pedianet, Italy; 2004-2018; N=121,540).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value, n (%)</th>
<th>IR(^a) × 100 PY(^b) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Obese—WHO(^c) definition</strong></td>
<td>26,990 (22.21)</td>
<td>4.05 (4.01-4.1)</td>
</tr>
<tr>
<td><strong>Obese—CDC(^d) definition</strong></td>
<td>12,676 (10.43)</td>
<td>1.86 (1.83-1.89)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>63,025 (51.86)</td>
<td>3.96 (3.89-4.02)</td>
</tr>
<tr>
<td>Female</td>
<td>58,515 (48.14)</td>
<td>4.16 (4.09-4.23)</td>
</tr>
<tr>
<td><strong>Year of birth</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004-2007</td>
<td>35,248 (29)</td>
<td>—(^e)</td>
</tr>
<tr>
<td>2008-2011</td>
<td>39,237 (32.28)</td>
<td>—</td>
</tr>
<tr>
<td>2012-2017</td>
<td>47,055 (38.72)</td>
<td>—</td>
</tr>
<tr>
<td><strong>Obesity age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preschool aged</td>
<td>—</td>
<td>1.48 (1.44-1.52)</td>
</tr>
<tr>
<td>School aged</td>
<td>—</td>
<td>3.68 (3.63-3.73)</td>
</tr>
<tr>
<td><strong>Local area of residence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>69,661 (57.32)</td>
<td>3.56 (3.5-3.62)</td>
</tr>
<tr>
<td>Center</td>
<td>18,037 (14.84)</td>
<td>4.01 (3.9-4.13)</td>
</tr>
<tr>
<td>South and Islands</td>
<td>33,842 (27.84)</td>
<td>5.13 (5.03-5.24)</td>
</tr>
<tr>
<td><strong>Deprivation index</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>16,582 (13.64)</td>
<td>4.19 (4.06-4.33)</td>
</tr>
<tr>
<td>Low</td>
<td>20,546 (16.90)</td>
<td>3.69 (3.58-3.8)</td>
</tr>
<tr>
<td>Medium low</td>
<td>21,666 (17.83)</td>
<td>3.93 (3.82-4.04)</td>
</tr>
<tr>
<td>Medium</td>
<td>22,695 (18.67)</td>
<td>4.13 (4.01-4.24)</td>
</tr>
<tr>
<td>Medium high</td>
<td>20,979 (17.26)</td>
<td>4.22 (4.1-4.34)</td>
</tr>
<tr>
<td>High</td>
<td>19,072 (15.69)</td>
<td>4.23 (4.1-4.36)</td>
</tr>
</tbody>
</table>

\(^a\)IR: incidence rate.  
\(^b\)PY: person-year.  
\(^c\)WHO: World Health Organization.  
\(^d\)CDC: Centers for Disease Control and Prevention.  
\(^e\)Not applicable.
Table 2. IR\(^a\) and 95% CI of obesity among characteristics of interest by exposure groups (Pedianet, Italy; 2004-2018; N=121,540).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Unexposed children (n=66,842)</th>
<th>Antibiotic-exposed children (n=54,698)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3.76 (3.67-3.85)</td>
<td>4.18 (4.08-4.28)</td>
</tr>
<tr>
<td>Female</td>
<td>3.88 (3.79-3.97)</td>
<td>4.52 (4.41-4.63)</td>
</tr>
<tr>
<td><strong>Obesity age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preschool aged</td>
<td>1.32 (1.27-1.37)</td>
<td>1.67 (1.61-1.73)</td>
</tr>
<tr>
<td>School aged</td>
<td>3.53 (3.46-3.59)</td>
<td>3.86 (3.79-3.94)</td>
</tr>
<tr>
<td><strong>Local area of residence</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>3.44 (3.37-3.52)</td>
<td>3.76 (3.66-3.86)</td>
</tr>
<tr>
<td>Center</td>
<td>4.05 (3.89-4.22)</td>
<td>3.97 (3.83-4.14)</td>
</tr>
<tr>
<td>South and Islands</td>
<td>4.83 (4.67-4.99)</td>
<td>5.36 (5.22-5.5)</td>
</tr>
<tr>
<td><strong>Deprivation index</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>3.87 (3.69-4.05)</td>
<td>4.58 (4.36-4.79)</td>
</tr>
<tr>
<td>Medium low</td>
<td>3.72 (3.58-3.87)</td>
<td>4.19 (4.02-4.37)</td>
</tr>
<tr>
<td>Medium</td>
<td>3.93 (3.78-4.08)</td>
<td>4.36 (4.19-4.53)</td>
</tr>
<tr>
<td>Medium high</td>
<td>3.99 (3.83-4.15)</td>
<td>4.5 (4.32-4.68)</td>
</tr>
<tr>
<td>High</td>
<td>4.02 (3.85-4.19)</td>
<td>4.47 (4.27-4.66)</td>
</tr>
</tbody>
</table>

\(^a\)IR: incidence rate.  
\(^b\)PY: person-year.

### Association Between Early Antibiotic Exposure and Incidence of Obesity

The findings presented in Table 3 are directly relevant to the central goal of this research investigation. The risk of obesity remained consistent across different timing of antibiotic prescriptions at 6 months, 1 year, and 2 years (fully adjusted HR [aHR] 1.07, 95% CI 1.04-1.10; aHR 1.06, 95% CI 1.03-1.09; aHR 1.07, 95% CI 1.04-1.10, respectively). No significant differences were observed between partially and fully adjusted analyses, even when restricting the cohort to complete cases for the ADI. Increasing the number of antibiotic exposures increases the risk of obesity significantly (\(P\) for trend<.001). Compared to children with no antibiotic prescriptions, those with 1, 2, or \(\geq\)3 prescriptions had an increased risk of 4% (95% CI 1-7), 6% (95% CI 2-10), and 14% (95% CI 9-18), respectively. The individual-specific age analysis showed that starting antibiotic therapy very early (between 0 and 5 months) had the greatest impact (aHR 1.12, 95% CI 1.08-1.17) on childhood obesity with respect to what was observed among those who were first prescribed antibiotics between 5 and 8 months of life (aHR 1.08, 95% CI 1.04-1.12) and the 8-12 months of life (aHR 1.03, 95% CI 0.99-1.06).

Subgroup analyses showed a higher risk of obesity among preschool-aged children (aHR 1.11, 95% CI 1.05-1.18) than school-aged children (aHR 1.05, 95% CI 1.02-1.09). When examining the subgroup of male and female participants separately, antibiotic exposure in the first year of life was associated with an increased risk of obesity of 4% (95% CI 1%-8%) and 10% (95% CI 7%-15%), respectively. However, the difference between female and male participants in the risk of obesity was not significant (\(P=.11\)). The increased risk of obesity among exposed children was consistent when analyzing only children with a birth weight within the normal range (aHR 1.08, 95% CI 1.04-1.12) and from the Veneto Region (aHR 1.07, 95% CI 1.02-1.11). In addition, using the CDC definition of obesity showed results consistent with the WHO definition (aHR 1.08, 95% CI 1.04-1.13).

Table 4 presents analyses based on the spectrum of action and class of antibiotic in the first year of life. Exposure to narrow-spectrum antibiotics appears associated with a higher risk of obesity (aHR 1.09, 95% CI 1.06-1.12) compared to broad-spectrum antibiotics (aHR 1.06, 95% CI 1.02-1.09). However, the difference between narrow- and broad-spectrum antibiotics in the risk of obesity was not significant (\(P=.52\)). In addition, the class of macrolides showed a stronger association with the risk of obesity (aHR 1.13, 95% CI 1.08-1.17) compared to other categories (aHR 1.10, 95% CI 1.00-1.20), while penicillins (aHR 1.07, 95% CI 1.04-1.10) and cephalosporins (aHR 1.07, 95% CI 1.03-1.12) had a similar extent of association.
Table 3. Partially and fully adjusted models assessing the association between antibiotic exposure and the risk of obesity (Pedianet, Italy; 2004-2018; N=121,540).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Children, n</th>
<th>Children exposed to antibiotics, n</th>
<th>Obese children exposed to antibiotics, n</th>
<th>HR(^a) (95% CI) (b)</th>
<th>(P) for trend</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure at 1 year</td>
<td>121,540</td>
<td>54,698</td>
<td>13,034</td>
<td>1.06 (1.03-1.09)</td>
<td></td>
</tr>
<tr>
<td>Partially adjusted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully adjusted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete case analysis(^c)</td>
<td>104,958</td>
<td>47,284</td>
<td>11,226</td>
<td>1.06 (1.03-1.09)</td>
<td></td>
</tr>
<tr>
<td>Partially adjusted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully adjusted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dose-response analysis (fully adjusted)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Exposure at 1 year</td>
<td>121,540</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>66,842</td>
<td>13,956</td>
<td>1.00 (Reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 RX(^d)</td>
<td>26,619</td>
<td>5856</td>
<td>1.04 (1.01-1.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 RXs</td>
<td>13,401</td>
<td>3175</td>
<td>1.06 (1.02-1.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥3 RXs</td>
<td>14,678</td>
<td>4003</td>
<td>1.14 (1.09-1.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual-specific age for exposure at 1 year</td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>121,540</td>
<td>66,842</td>
<td>13,956</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>≥0 months and &lt;5 months</td>
<td>121,540</td>
<td>15,020</td>
<td>3849</td>
<td>1.12 (1.08-1.17)</td>
<td></td>
</tr>
<tr>
<td>≥5 months and &lt;8 months</td>
<td>121,540</td>
<td>16,542</td>
<td>3950</td>
<td>1.08 (1.04-1.12)</td>
<td></td>
</tr>
<tr>
<td>≥8 months and ≤12 months</td>
<td>121,540</td>
<td>23,136</td>
<td>5235</td>
<td>1.03 (0.99-1.06)</td>
<td></td>
</tr>
<tr>
<td>Exposure at 6 months (fully adjusted)</td>
<td>121,540</td>
<td>23,576</td>
<td>5889</td>
<td>1.07 (1.04-1.10)</td>
<td></td>
</tr>
<tr>
<td>Exposure at 2 years (fully adjusted)</td>
<td>121,540</td>
<td>83,846</td>
<td>19,405</td>
<td>1.07 (1.04-1.10)</td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal birth weight (fully adjusted)</td>
<td>117,397</td>
<td>52,543</td>
<td>12,409</td>
<td>1.06 (1.03-1.09)</td>
<td></td>
</tr>
<tr>
<td>Preschool aged (fully adjusted)</td>
<td>121,540</td>
<td>5986</td>
<td>3046</td>
<td>1.11 (1.05-1.18)</td>
<td></td>
</tr>
<tr>
<td>School-aged (fully adjusted)</td>
<td>77,537</td>
<td>21,004</td>
<td>9988</td>
<td>1.05 (1.02-1.09)</td>
<td></td>
</tr>
<tr>
<td>Female (fully adjusted)</td>
<td>58,515</td>
<td>25,210</td>
<td>6254</td>
<td>1.10 (1.07-1.15)</td>
<td></td>
</tr>
<tr>
<td>Male (fully adjusted)</td>
<td>63,025</td>
<td>29,488</td>
<td>6780</td>
<td>1.04 (1.01-1.08)</td>
<td></td>
</tr>
<tr>
<td>Veneto region (fully adjusted)</td>
<td>49,772</td>
<td>9518</td>
<td>3826</td>
<td>1.07 (1.02-1.11)</td>
<td></td>
</tr>
<tr>
<td>CDC(^e) definition (fully adjusted)</td>
<td>121,553</td>
<td>12,676</td>
<td>6413</td>
<td>1.08 (1.04-1.13)</td>
<td></td>
</tr>
</tbody>
</table>

\(a\)HR: hazard ratio.

\(b\)Not applicable.

\(c\)The complete case analysis included 104,858 children.

\(d\)RX: prescription.

\(e\)CDC: Centers for Disease Control and Prevention.
Table 4. Fully adjusted models assessing the association between the spectrum of action and class of antibiotic therapy and the risk of obesity (Pedianet, Italy; 2004-2018; N=121,540).

<table>
<thead>
<tr>
<th>Spectrum of action</th>
<th>Children, n</th>
<th>Children exposed to antibiotics, n</th>
<th>Obese children exposed to antibiotics, n</th>
<th>HR* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrow-spectrum antibiotics</td>
<td>99,170</td>
<td>32,328</td>
<td>7856</td>
<td>1.09 (1.06-1.12)</td>
</tr>
<tr>
<td>Broad-spectrum antibiotics</td>
<td>101,844</td>
<td>35,002</td>
<td>8503</td>
<td>1.06 (1.02-1.09)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class of antibiotic therapy</th>
<th>Children exposed to antibiotics, n</th>
<th>HR* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penicillins</td>
<td>107,516</td>
<td>35,295</td>
</tr>
<tr>
<td>Cephalosporins</td>
<td>81,356</td>
<td>12,355</td>
</tr>
<tr>
<td>Macrolides</td>
<td>81,989</td>
<td>13,042</td>
</tr>
<tr>
<td>Others</td>
<td>69,333</td>
<td>2110</td>
</tr>
</tbody>
</table>

*aHR: hazard ratio.

Discussion

Principal Findings

In this pediatric population–based cohort study of 121,540 children, we found a 6% (aHR 1.06, 95% CI 1.03-1.09) increased risk of developing childhood obesity among children exposed to antibiotics within the first year of life compared to unexposed children. This relationship is stronger as the number of prescriptions increases and as the individual’s age at the first prescription of antibiotics decreases. The results were consistent across all sensitivity and subgroup analyses conducted.

Obesity is a significant public health concern worldwide, especially among children and young people. Italy is one of the countries in Europe with the highest rate of childhood obesity: according to the latest report by “OKkio alla Salute” from the Ministry of Health in 2019, 38% are overweight and 16.5% are obese among school-aged children; they also used the WHO Growth References [31]. There continued to be a large difference between countries. Overweight prevalence in children varied remarkably from 6% in Tajikistan to 43% in Cyprus. Similarly, obesity rates spanned a wide spectrum, from a minimal 1% in Tajikistan to a concerning 19% in Cyprus. In our cohort, 22% of children were classified as obese at least in 1 measurement during the follow-up when the age- and sex-specific z score was calculated using the WHO Growth References, with respect to 10% when the CDC Growth Charts were used. The reason for this discrepancy relies on the fact that while the CDC 2000 Growth Charts represent the reference growth charts for the US pediatric population, the WHO 2006 ones are intended to serve as growth chart standards, describing how children should grow globally and not how they did grow in a specific nation. Indeed, while the CDC Growth Charts were developed from 5 nationally representative survey data sets from the United States (the National Health Examination Surveys), the WHO ones were based on data from a Multicenter Growth Reference Study that collected a highly selective sample of children from 6 sites around the world (Brazil, Ghana, India, Norway, Oman, and the United States), consisting of children who were not subjected to socioeconomic constraints on growth, who were fed according to the study feeding recommendations, who were healthy term singleton births, and whose mothers did not smoke; the growth of these children was considered to represent optimal growth. Several studies have subsequently observed a significant difference in the rate of overweight or underweight children depending on the growth charts used in national prevalence studies [32-34].

Previous studies have investigated the association between early antibiotic exposure and childhood obesity and reported inconsistent results [2,29-33]. Our results confirm previous findings demonstrating a positive association through several exhaustive analyses, even with attenuated estimations [35-39]. Moreover, our results confirmed previous evidence that the risk of childhood obesity was more significant in those children who received more than 3 antibiotics in the early years of life [37] and with earlier start-exposure timing [39]. However, our results did not support previous evidence showing stronger associations between obesity risk and the male sex of the children. In fact, we found an increased risk of obesity among female participants (aHR 1.10, 95% CI 1.07-1.15) compared with male participants (aHR 1.04, 95% CI 1.01-1.08) [39].

Furthermore, we explored the impact of different spectrums of action and types of antibiotics on childhood obesity. Conversely to Bailey et al [2] who found a stronger association for broad-spectrum antibiotics but not for narrow-spectrum therapy, our findings revealed a positive association with both narrow- and broad-spectrum antibiotics.

The causes of childhood obesity are complex and multifactorial, although the hypothesis that gut microbiota plays a crucial role in the pathogenesis of obesity is well established [40-43]. Recent studies have indicated that antibiotic use leads to alterations in the gut microbiota, potentially affecting nutrient absorption and resulting in metabolic imbalances that contribute to obesity [7]. Our results corroborate what has already been shown in various laboratory models, supporting the notion that weight gain induced by antibiotics is mediated through the drug’s impact on the gut microbiome and the consequently altered circulating levels of substances (eg, short-chain fatty acids, secondary biliary acids, and branched and aromatic amino acids) that influence human metabolism [7]. Changes in microbiota caused by antibiotic use are defined as dysbiosis, a microbial imbalance correlated with impaired health, including increased susceptibility to infections, impaired immune function,
gastrointestinal symptoms, and even long-term effects on metabolic health and obesity [44]. Obesity is a chronic and complex disease with significant long-term effects; hence, it is crucial to implement interventions to prevent this condition in children. Preventing and addressing childhood obesity require a comprehensive, multisectoral approach involving individuals, families, communities, educational institutions, health care systems, and policy makers. Strategies may include promoting healthy eating habits, increasing physical activity opportunities, improving food environments, and implementing policies that restrict the marketing of unhealthy foods to children as also supported by the Health4EUkids project; the European Joint Action for the Implementation of Best Practices for the promotion of health and the prevention of obesity [5].

There are several potential limitations to consider in this research. First, exposure to antibiotics was defined by prescription, assuming that a prescription led to actual medication use. However, this assumption may not always hold true, leading to possible misclassification of the exposure. Anyhow, this misclassification is likely to bias the results toward the null hypothesis. Moreover, antibiotics prescribed in a private setting or hospitals are not recorded. Second, important confounding factors related to both pregnancy and early life were not captured in Pedianet. This includes factors like maternal BMI before and during pregnancy, breastfeeding, method of complementary feeding, dietary patterns, sugar-sweetened beverages consumption, eating behavior (eg, skipping breakfast and family dinners), meal frequency and composition (fast foods and snacking), portion size, physical activity, screen media exposure, and sedentary behavior [45,46]. These factors could potentially influence the results and lead to residual confounding. We adjusted the models for various sociodemographic characteristics to address some of these limitations. Notably, we included the area-level socioeconomic deprivation index, which is commonly used in public health research. It serves to quantify the extent of geographically determined social inequalities in health or assess the independent effect that area characteristics have on health beyond individual socioeconomic position. In addition, this index can help substitute missing individual-level data in epidemiological studies and account for confounding socioeconomic factors.

This study also has several strong points. First, using a pediatric population–based registry allowed for a large unbiased cohort with extended follow-up, thereby minimizing selection and recall bias. Second, to ensure accuracy and avoid potential exposure or outcome misclassification, we included only children who were consistently monitored by their FP during the first 2 years of life and had at least 1 reliable BMI measurement after completing the second year of life. Furthermore, setting the end of follow-up as the date of the last BMI measurement for each child helped minimize outcome misclassification. Finally, we used all the anthropometric measures (N=564,066) recorded during the pediatric visits performed by children included in the cohort. To our knowledge, no previous study has been conducted with this level of comprehensiveness and detail in terms of anthropometric data collection.

Conclusions

In conclusion, the results from this large population-based study support the association between early exposure to antibiotics and an increased risk of childhood obesity. This relationship becomes more pronounced as the number of prescriptions increases and as the age specific for the first prescription of antibiotics decreases.

Acknowledgments

This research is funded by the Ministry of University and Research within the Complementary National Plan PNC-I.1 “Research initiatives for innovative technologies and pathways in the health and welfare sector,”—D.D. 931 of 06/06/2022, PNC0000002 DARE—Digital Lifelong Prevention CUP: B53C22006440001. The authors gratefully acknowledge the contributions of Luigi Cantarutti (Pedianet), and all Pedianet family pediatricians (Pedianet): Alongi Angelo, Andreola Barbara, Angelini Roberta, Apostolo Maria Grazia, Arcangeli Bruno, Argo Giovanna, Avarelo Giovanni, Azzone Lucia Barbazza, Maria Carolina, Barbieri Patrizia, Belluzzi Gabriele, Benetti Eleonora, Biasci Paolo, Boe Franca, Bollettini Stefano, Bonaiuto Francesco, Bontempelli Anna Maria, Brusaterra Ivana, Budassi Roberto, Caccini Massimo, Calif Laura, Cantalupi Laura, Cardarelli Chiara Garri Giovanni, Carnazza Sylvia, Castaldo Massimo, Castelli Stefano, Cavedagni Monica, Cera Giuseppe Egidio, Chillemi Chiara, Cichello Francesca, Cicione Giuseppe, Ciscato Carla, Clerici Schoeller Marialunga, Collacciani Giuseppe, Comaita Fabrizio, Conte Valeria, Conte Ugo Alfredo, Corro’ Roberta, Costaglolia Rosaria, Costanzo Nicola, Cozzani Sandra, Cuboni Giancarlo, Curia Giorgia, D’alia Caterina, D’Amanti Vito Francesco, D’Avino Antonio, De Clara Roberto, De Giovani Lorenzo, De Marchi Annamaria, Del Ponte Gigiola, Di Giampietro Tiziana, Di Mauro Giuseppe, Di Santo Giuseppe, Di Saverio Piero, Di Mattei Dolci, Marco, Doria Mattia, El Mazzoum Dania, Fadda Maria Carmen, Falco Pietro Fama Mario, Faraci Marco, Farina Maria Immucolata, Favilli Tania, Federico Mariagrazia, Felici Michele, Ferraiuolo Maurizio, Ferrara Enrico Ferretti Michele, Forcina Paolo, Foti Patrizia, Frison Ezzo, Fusco Fabrizio, Gallo Giovanni, Galvagno Andrea, Genova Grazia, Gentilucci Pierfrancesco, Giampaolo Francesco, Girotto Francesca, Girotto Giuseppe, Grelloni Mauro, Gruelli Mirco, Isca Antonina, Lagrasta Urania, Lisabetta Landi, Letta Maria Rosaria, Lietti Giuseppe, Lista Cinzia, Lucantonio Riccardo, Luise Francesca, Macropodio Nadia, Marine Francesca, Marinelli Lorenzo, Marostica Gabriella, Masotti Sergio, Meneghetti Stefano, Milani Massimo, Milone Stella Vittoria, Moggia Donatella, Monteleone Angela Maria, Mussini Pierangelina, Naccari Anna, Nicoloso Flavia, Olimpi Laura Maria, Ongaro Riccardo, Palma Maria Maddalena Pasinato Angela, Passarella Andrea, Pazzola Pasquale, Perin Monika, Perri Danilo, PescoSoldi Silvana Rosa, Petrazzoli Giovanni, Petrotto Giuseppe, Picco Patrizia, Pirola Ambrogina, Pisanello Lorena, Pitarello Daniele, Porro Elena, Porto Adolfo Francesco, Puocci Maria Paola, Righetti Andrea, Rizzare Rosari, Rosas Paolo, Ruoffato

https://publichealth.jmir.org/2022/1/e51734

JMIR Public Health Surveill 2022 | vol. 10 | e51734 | p.1823

(page number not for citation purposes)
Data Availability
The data supporting this study’s findings are available on request from the corresponding author (AC). The data are not publicly available due to restrictions (containing information that could compromise the privacy of research participants).

Authors’ Contributions
All authors contributed to the study’s conception and design. Material preparation, data collection, and analysis were performed by AC and CC. AC, PR, CDB, JGRD, SB, EB, CG, and CC wrote the first draft of the paper. All authors read and approved the final paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary material.

References


32. Flegal KM, Ogden CL. Childhood obesity: are we all speaking the same language? Adv Nutr 2011;2(2):159S-166S [FREE Full text] [doi: 10.3945/an.111.00307] [Medline: 22332047]


Abbreviations

ADI: area deprivation index
aHR: adjusted hazard ratio
CDC: Centers for Disease Control and Prevention
FP: family pediatrician
IR: incidence rate
PY: person-year
WHO: World Health Organization

Edited by A Mavragani, T Sanchez; submitted 10.08.23; peer-reviewed by A Catalano, H Li; comments to author 09.01.24; revised version received 30.01.24; accepted 11.03.24; published 31.05.24.

Please cite as:
Cantarutti A, Rescigno P, Da Borso C, Gutierrez de Rubalcava Doblas J, Bressan S, Barbieri E, Giaquinto C, Canova C
Association Between Early-Life Exposure to Antibiotics and Development of Child Obesity: Population-Based Study in Italy
JMIR Public Health Surveill 2024;10:e51734
URL: https://publichealth.jmir.org/2024/1/e51734
doi:10.2196/51734
PMID:38820572

©Anna Cantarutti, Paola Rescigno, Claudia Da Borso, Joaquin Gutierrez de Rubalcava Doblas, Silvia Bressan, Elisa Barbieri, Carlo Giaquinto, Cristina Canova. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 31.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic
information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Predicting the Transition to Metabolically Unhealthy Obesity Among Young Adults With Metabolically Healthy Obesity in South Korea: Nationwide Population-Based Study

HyunHae Lee1*, MSN, RN; Ji-Su Kim2*, PhD, RN; Hyerine Shin2*, MSN, RN

1School of Nursing, University of Washington, Seattle, WA, United States
2Department of Nursing, Chung-Ang University, Seoul, Republic of Korea
*all authors contributed equally

Corresponding Author:
Ji-Su Kim, PhD, RN
Department of Nursing
Chung-Ang University
84 Heukseok-ro Dongjak-Gu
Seoul, 06974
Republic of Korea
Phone: 82 2 820 5991
Fax: 82 2 824 7961
Email: jisu80@cau.ac.kr

Abstract

Background: Globally, over 39% of individuals are obese. Metabolic syndrome, usually accompanied by obesity, is regarded as a major contributor to noncommunicable diseases. Given this relationship, the concepts of metabolically healthy and unhealthy obesity, considering metabolic status, have been evolving. Attention is being directed to metabolically healthy people with obesity who have relatively low transition rates to noncommunicable diseases. As obesity rates continue to rise and unhealthy behaviors prevail among young adults, there is a growing need for obesity management that considers these metabolic statuses. A nomogram can be used as an effective tool to predict the risk of transitioning to metabolically unhealthy obesity from a metabolically healthy status.

Objective: The study aimed to identify demographic factors, health behaviors, and 5 metabolic statuses related to the transition from metabolically healthy obesity to unhealthy obesity among people aged between 20 and 44 years and to develop a screening tool to predict this transition.

Methods: This secondary analysis study used national health data from the National Health Insurance System in South Korea. We analyzed the customized data using SAS (SAS Institute Inc) and conducted logistic regression to identify factors related to the transition from metabolically healthy to unhealthy obesity. A nomogram was developed to predict the transition using the identified factors.

Results: Among 3,351,989 people, there was a significant association between the transition from metabolically healthy to unhealthy obesity and general characteristics, health behaviors, and metabolic components. Male participants showed a 1.30 higher odds ratio for transitioning to metabolically unhealthy obesity than female participants, and people in the lowest economic status were also at risk for the transition (odds ratio 1.08, 95% CI 1.05-1.1). Smoking status, consuming >30 g of alcohol, and insufficient regular exercise were negatively associated with the transition. Each relevant variable was assigned a point value. When the nomogram total points reached 295, the shift from metabolically healthy to unhealthy obesity had a prediction rate of >50%.

Conclusions: This study identified key factors for young adults transitioning from healthy to unhealthy obesity, creating a predictive nomogram. This nomogram, including triglycerides, waist circumference, high-density lipoprotein-cholesterol, blood pressure, and fasting glucose, allows easy assessment of obesity risk even for the general population. This tool simplifies predictions amid rising obesity rates and interventions.

(JMIR Public Health Surveill 2024;10:e52103) doi:10.2196/52103

https://publichealth.jmir.org/2024/1/e52103
KEYWORDS
metabolically healthy obesity; metabolic syndrome; metabolically unhealthy obesity; nomogram; obesity; young adult; male; noncommunicable disease; South Korea; population-based study; intervention

Introduction
In the last 40 years, the number of people with obesity has more than tripled worldwide; as of 2016, 39% of the global population was estimated to have a BMI ≥25 kg/m² [1]. Obesity, usually accompanied by metabolic syndrome (MetS), is a major risk factor for noncommunicable diseases (NCDs), such as cancer, musculoskeletal problems, and cardiovascular diseases [2]. Epidemiological studies showed that 20%-45% of people have MetS. Furthermore, MetS prevalence is anticipated to increase to >50% by 2035 [3].

Recent studies have increasingly emphasized metabolically healthy obesity (MHO) and metabolically unhealthy obesity (MUO), with rising interest in the connection between obesity and MetS [4-6]. Studies revealed that individuals with MHO have a significantly reduced incidence of common obesity-related disorders, such as stroke, cancer, cardiovascular disease, and type 2 diabetes [7]. However, recent research indicates that individuals with MHO might eventually transition to MUO [6], with up to 50% of the population with obesity possibly falling into the MHO category [8]. This underlines the need for research and treatment to prevent or interrupt the transition from MHO to MUO.

On the other hand, young adults face higher mortality from obesity than older people. As social relationships that foster independence evolve, the prevalence of unhealthy lifestyle behaviors, such as smoking, alcohol consumption, and physical inactivity, has been rapidly increasing, contributing to chronic diseases in the population with obesity [9,10]. Additionally, the COVID-19 pandemic caused increased psychological stress, decreased physical activity, and altered eating patterns in young adults, contributing to obesity [11]. Young adults are more aware of obesity than older populations but engage less in weight loss programs and achieve less intentional weight loss [12], underscoring the need for proactive interventions and tailored programs for this age group.

This study identified risk factors and developed a screening tool with a nomogram to prevent the transition from MHO to MUO. A nomogram, a graphical tool that can perform a complicated approximation calculation, provided a clear interpretation of which predictors could be more critical factors [13]. The use of this scale also allows the general population to intuitively and easily assess the likelihood of transitioning to MUO. Therefore, this study aimed to predict the transition from MHO to MUO among young individuals with obesity using the nomogram.

Methods

Study Setting
The Korean government provides health checkups for 97% of the South Korean population through public health insurance, while the remaining 3% receive these checkups through medical aid programs. Residents are encouraged to undergo biennial health checkups, covering measurements such as height, weight, BMI, blood tests, and urinary tests. The National Health Insurance System (NHIS) is tasked with collecting medical records, insurance statuses, and other pertinent data through these enrollment methods, and providing corresponding services [14].

Measures
MHO and MUO are defined based on the metabolic status of the population with obesity. Obesity is characterized by an abnormal or excessive accumulation of adipose tissue, and BMI is primarily used to define it [1]. According to evidence regarding the high risk of type 2 diabetes or cardiovascular diseases with lower BMI standards, the World Health Organization (WHO) proposed each country make decisions about obesity definitions [15,16]. The Korean Society for the Study of Obesity defined obesity as a BMI (weight in kg divided by the square of height in meters) ≥25 kg/m² [17]. Indicators for assessing metabolic status include waist circumference, fasting blood glucose levels, blood pressure, high-density lipoprotein (HDL) cholesterol levels, and fasting triglyceride levels. This study used the cutoff points ≥90 cm in male participants and ≥85 cm in female participants for waist circumference (a Korean-specific cutoff point) [17]; ≥130/85 mmHg or using antihypertensive medications for blood pressure; ≥100 mg/dL or using antidiabetic medications for fasting blood glucose; ≥150 mg/dL for fasting triglycerides; and <40 mg/dL in male participants and <50 mg/dL in female participants for HDL cholesterol [18]. An individual is defined as metabolically unhealthy if ≥3 of these 5 indicators yield abnormal results. Therefore, MUO is a state of obesity with 3 or more abnormal indicators of metabolic status. Those who do not meet these criteria are defined as MHO [5].

In this study, healthy behaviors (smoking, alcohol consumption, and regular exercise), which are representative factors influencing MetS and obesity, were examined. Participants were classified according to their smoking status. Alcohol consumption was assessed by categorizing participants into individuals who consume alcohol heavily (heavy; ≥50 g/day), moderate alcohol (moderate; 30-49 g/day), and do not drink alcohol (none) [19]. Regular exercise was defined as engaging in high-intensity exercise for at least 20 minutes, ≥3 times per week, or moderate-intensity exercise for at least 30 minutes, ≥5 times per week [20].

Study Population
We identified participants (aged ≥20 years) who had health checkups in 2009-2010 and 2013-2014. A total of 3,351,989 individuals were obese (BMI ≥25 kg/m²) during both examinations. From this group, we excluded those who had MUO during the first examination and selected those aged between 20 and 44 years during the second examination. We also excluded individuals diagnosed with cancer or cardiovascular diseases (myocardial infarction and stroke) before
the second examination. After excluding missing values, the final sample size included in the analysis was determined to be 562,765 individuals.

Data Analysis

Young adults’ general characteristics, health behaviors, and metabolic statuses were summarized using the chi-square test and 2-tailed t tests. We used logistic regression analysis to examine the association between the transition to MUO and general characteristics, health behaviors, and metabolic states. Based on the results of the logistic regression, factors associated with the transition to MUO were identified, and scores were assigned ranging from 0 to 100. The prediction model for calculating the risk of the transition to MUO was depicted using a nomogram. P values <.05 were considered statistically significant. Statistical analysis was conducted using SAS (version 9.4; SAS Institute Inc).

Ethical Considerations

The project was submitted for evaluation to the institutional review board of Chung-Ang University (1041078-202203-HR-080). Since the study was a secondary analysis using the NHIS data, the institutional review board exempted the study from formal ethical approval. All participants who underwent health checkups through NHIS consented to provide the results and answers for research purposes. The data were deidentified by the NHIS and provided with permission.

Results

General Characteristics and Metabolic Status of Participants According to Metabolic Status After 4 Years

The general characteristics of participants who transitioned from MHO to MUO over a 4-year period, compared to those who did not, are outlined in Table 1. A higher proportion of male participants was observed in the group that transitioned to MUO, with a statistically significant difference (P<.001). Age, income level, and health behaviors (smoking status, alcohol consumption, and regular exercise) showed statistically significant differences as well (P<.001). Both groups showed significant differences regarding metabolic status based on the screening time point, even at the initial screening. Mainly, the group that transitioned to MUO exhibited remarkable changes in metabolic status, with variations in triglyceride, fasting glucose, waist circumference, blood pressure, and HDL.
Table 1. General characteristics of participants according to metabolic status after 4 years.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=562,765)</th>
<th>Metabolically healthy obesity (n=391,593)</th>
<th>Metabolically unhealthy obesity (n=171,172)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sex (male), n (%)</td>
<td>460,022 (81.74)</td>
<td>311,859 (79.64)</td>
<td>148,163 (86.56)</td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>37.39 (4.87)</td>
<td>37.34 (4.91)</td>
<td>37.52 (4.78)</td>
<td></td>
</tr>
<tr>
<td>Income (Q1&lt;sub&gt;a&lt;/sub&gt;), n (%)</td>
<td>45,302 (8.05)</td>
<td>32,297 (8.25)</td>
<td>13,005 (7.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Smoking status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>224,106 (39.82)</td>
<td>167,087 (42.67)</td>
<td>57,019 (33.31)</td>
<td></td>
</tr>
<tr>
<td>Past</td>
<td>112,365 (19.97)</td>
<td>77,603 (19.82)</td>
<td>34,762 (20.31)</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>226,294 (40.21)</td>
<td>146,903 (37.51)</td>
<td>79,391 (46.38)</td>
<td></td>
</tr>
<tr>
<td><strong>Alcohol consumption, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>None</td>
<td>166,112 (29.52)</td>
<td>120,089 (30.67)</td>
<td>46,023 (26.89)</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>329,432 (58.54)</td>
<td>229,660 (58.65)</td>
<td>99,772 (58.29)</td>
<td></td>
</tr>
<tr>
<td>Heavy</td>
<td>67,221 (11.94)</td>
<td>41,844 (10.69)</td>
<td>25,377 (14.83)</td>
<td></td>
</tr>
<tr>
<td><strong>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), mean (SD)</strong></td>
<td>88.94 (7.04)</td>
<td>87.34 (6.55)</td>
<td>92.61 (6.75)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Waist circumference (cm), mean (SD)</td>
<td>96.02 (16.79)</td>
<td>92.83 (12.5)</td>
<td>103.32 (22.21)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fasting glucose (mg/dL), mean (SD)</td>
<td>124.1 (12.52)</td>
<td>121.48 (11.67)</td>
<td>130.08 (12.35)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg), mean (SD)</td>
<td>78.4 (9.34)</td>
<td>76.64 (8.73)</td>
<td>82.43 (9.46)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Diastolic blood pressure (mm Hg), mean (SD)</td>
<td>51.04 (13.6)</td>
<td>53.04 (13.64)</td>
<td>46.45 (12.32)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>HDL&lt;sub&gt;b&lt;/sub&gt;-cholesterol (mg/dL), mean (SD)</td>
<td>134.9 (134.7-135.1)</td>
<td>115.6 (115.4-115.8)</td>
<td>192.1 (191.7-192.6)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fasting glucose (mg/dL), geometric mean (95% CI)</td>
<td>27.77 (2.33)</td>
<td>27.36 (2.05)</td>
<td>28.71 (2.65)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Metabolic syndrome component (initial exam), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal waist circumference</td>
<td>167,920 (29.84)</td>
<td>104,932 (26.8)</td>
<td>62,988 (36.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal fasting glucose</td>
<td>85,539 (15.2)</td>
<td>56,577 (14.45)</td>
<td>28,962 (16.92)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal blood pressure</td>
<td>180,849 (32.14)</td>
<td>117,449 (29.99)</td>
<td>63,400 (37.04)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal HDL-cholesterol</td>
<td>79,524 (14.13)</td>
<td>51,530 (13.16)</td>
<td>27,994 (16.35)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal triglycerides</td>
<td>167,215 (29.71)</td>
<td>99,100 (25.31)</td>
<td>68,115 (39.79)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Metabolic syndrome component (after 4 years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal waist circumference</td>
<td>268,350 (47.68)</td>
<td>136,694 (34.91)</td>
<td>131,656 (76.91)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal fasting glucose</td>
<td>168,896 (30.01)</td>
<td>70,083 (17.9)</td>
<td>98,813 (57.73)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal blood pressure</td>
<td>246,816 (43.86)</td>
<td>120,013 (30.65)</td>
<td>126,803 (74.08)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal HDL-cholesterol</td>
<td>136,472 (24.25)</td>
<td>53,000 (13.53)</td>
<td>83,472 (48.76)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abnormal triglycerides</td>
<td>242,884 (43.16)</td>
<td>104,619 (26.72)</td>
<td>138,265 (80.78)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>Participants who got medical aid or was included in the lowest quartile of income.

<sup>b</sup>HDL: high-density lipoprotein.

Logistic Regression Analyses of the Transition From MHO to MUO

We identified factors contributing to the transition from MHO to MUO through logistic regression analysis (Table 2). Adjusting for all significant characteristics identified in Table 1, we constructed model 2. The model demonstrated that transitioning from MHO to MUO was independently associated with an increase in age.
# Table 2. Logistic regression analysis of the transition from metabolically healthy obesity to metabolically unhealthy obesity.

| Variables                                | Participants, n | Metabolically unhealthy obesity, n (%) | Model 1, odds ratio (95% CI)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>General characteristics</td>
<td></td>
<td></td>
<td>Model 2, odds ratio (95% CI)</td>
</tr>
<tr>
<td><strong>Age (per 1 year)</strong></td>
<td>__b</td>
<td>—</td>
<td>1.008 (1.006-1.009)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>460,022</td>
<td>148,163 (32.21)</td>
<td>1.65 (1.62-1.67)</td>
</tr>
<tr>
<td>Female</td>
<td>102,743</td>
<td>23,009 (22.39)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>45,302</td>
<td>13,005 (28.71)</td>
<td>0.92 (0.9-0.93)</td>
</tr>
<tr>
<td>Q2-Q4</td>
<td>517,463</td>
<td>158,167 (30.57)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>224,106</td>
<td>57,019 (25.44)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Past</td>
<td>112,365</td>
<td>34,762 (30.94)</td>
<td>1.31 (1.29-1.33)</td>
</tr>
<tr>
<td>Current</td>
<td>226,294</td>
<td>79,391 (35.08)</td>
<td>1.58 (1.56-1.6)</td>
</tr>
<tr>
<td><strong>Alcohol consumption</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>166,112</td>
<td>46,023 (27.71)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Moderate</td>
<td>329,432</td>
<td>99,772 (30.29)</td>
<td>1.13 (1.12-1.15)</td>
</tr>
<tr>
<td>Heavy</td>
<td>67,221</td>
<td>25,377 (37.75)</td>
<td>1.58 (1.55-1.61)</td>
</tr>
<tr>
<td><strong>Regular exercise</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>452,103</td>
<td>143,511 (31.74)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Yes</td>
<td>110,662</td>
<td>27,661 (25)</td>
<td>0.72 (0.71-0.73)</td>
</tr>
<tr>
<td><strong>Metabolic syndrome component (first exam)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Waist circumstance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>394,845</td>
<td>108,184 (27.40)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Yes</td>
<td>167,920</td>
<td>62,988 (37.51)</td>
<td>1.59 (1.57-1.61)</td>
</tr>
<tr>
<td><strong>Fasting glucose</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>477,226</td>
<td>142,210 (29.8)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Yes</td>
<td>85,539</td>
<td>28,962 (33.86)</td>
<td>1.21 (1.19-1.23)</td>
</tr>
<tr>
<td><strong>Blood pressure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>381,916</td>
<td>107,772 (28.22)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Yes</td>
<td>180,849</td>
<td>63,400 (35.06)</td>
<td>1.37 (1.36-1.39)</td>
</tr>
<tr>
<td><strong>HDL-d-cholesterol</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>483,241</td>
<td>143,178 (29.63)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Yes</td>
<td>79,524</td>
<td>27,994 (35.2)</td>
<td>1.29 (1.27-1.31)</td>
</tr>
<tr>
<td><strong>Triglycerides</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>395,550</td>
<td>103,057 (26.05)</td>
<td>1 (Reference)</td>
</tr>
<tr>
<td>Yes</td>
<td>167,215</td>
<td>68,115 (40.73)</td>
<td>1.95 (1.93-1.98)</td>
</tr>
</tbody>
</table>

aModel 1 was not adjusted, and model 2 was adjusted for age, sex, income, smoking, drinking, regular exercise, waist circumstance, fasting glucose, blood pressure, HDL-cholesterol, and triglycerides; all variables had a statistically significant association with the transition to metabolically unhealthy obesity (P<.001).

bNot applicable.

cQ1-Q4, first through fourth quartiles.

dHDL: high-density lipoprotein.
For each additional year of age, the odds ratio (OR) of transitioning to MUO was 1.04 times higher (95% CI 1.002-1.005). Male participants showed a 1.30 times higher OR of transitioning to MUO than female participants (95% CI 1.27-1.32). People belonging to the lowest economic group showed a 1.08 higher OR of transitioning to MUO than other economic status groups (95% CI 1.05-1.1). Regarding healthy behaviors, having a history of smoking and being an individual who currently smoke showed 1.1 (95% CI 1.09-1.12) and 1.27 (95% CI 1.25-1.29) higher OR of transitioning to MUO than people without a smoking history. Individuals consuming alcohol heavily had a 1.26 higher OR (95% CI 1.23-1.28) of transitioning to MUO compared to people not drinking alcohol. People participating in regular exercise had a 0.72 lower OR (95% CI 0.7-0.73) of transitioning to MUO than people not regularly exercising. Abnormal waist circumference at the initial screening (OR 2.11, 95% CI 2.08-2.14), impaired fasting glucose or related medication use (OR 1.69), hypertension or related medication use (OR 1.76, 95% CI 1.76-1.78), low HDL cholesterol or related medication use (OR 1.89, 95% CI 1.86-1.92), and elevated triglyceride or related medication use (OR 2.25, 95% CI 2.22-2.28) were all statistically significantly associated with the transition to MUO.

Prediction of the Transition From MHO to MUO

A nomogram (Figure 1) was constructed to assist clinical predictions, providing a quantitative estimation of the probability of transitioning to MUO. This is a screening tool for predicting the transition from MHO to MUO with 66% discriminative ability (Figure 2). According to individual characteristics, each value is located based on each variable in the nomogram. At each variable point, a vertical line is drawn up to the scale of “points,” followed by identifying the point of the variable. After marking all values based on variables, all points of each variable are summarized, and the point on the scale of “total points” is marked. At the total point, after drawing a vertical line to the scale of “predicted value,” the matched point will be the individual’s predicted value of transitioning from MHO to MUO. Using multivariate logistic regression, we identified 11 independent predictors, assigning the highest score in the nomogram to each variable (eg, 44 years: 9 points; male participant: 32 points; first income level quartile: 9 points; individual who smokes: 29 points; individual who drinks: 29 points; no regular physical activity: 41 points; abnormal waist circumference: 92 points; abnormal fasting glucose or taking diabetes medications: 65 points; abnormal blood pressure or taking hypertension medication: 70 points; abnormal HDL level or taking dyslipidemia medications: 78 points; abnormal serum triglyceride level or taking dyslipidemia medication: 100 points; and total points: 344 points). When the total points reached 295, the prediction rate for transitioning from MHO to MUO was determined to be higher than 50%. As depicted in Figure 3, the nomogram showed close agreement between predicted and observed outcomes in the study population, with a slight tendency for overestimation in the 0.1-0.2 and 0.5-0.6 intervals. Figure 2 shows the receiver operating characteristic curve used to validate the nomogram, yielding an area under the curve of 0.66 (95% CI 0.658-0.661).

Figure 1. The nomogram for predicting the transition from metabolically healthy obesity to metabolically unhealthy obesity.
Figure 2. A receiver operating characteristic (ROC) curve of the nomogram (model 2).

Figure 3. The calibration curves for the nomogram for predicting the transition from metabolically healthy obesity to metabolically unhealthy obesity. The dotted line represents the ideal concordance between the predicted and actual probabilities. The solid line represents the performance of the nomogram.
Discussion

Principal Findings

The prevalence of obesity is escalating globally, accompanied by MetS, increasing the risk of NCDs. This upward trend among young adults can be attributed to lifestyle alterations during adulthood [10] and, more recently, decreased physical activity or changed dietary habits due to the COVID-19 pandemic [11]. Moreover, individuals with MUO face substantially elevated risks of NCDs compared to those with MHO [4]. Considering that MHO might merely represent a “honeymoon phase” preceding the shift to MUO [6], the urgency of managing MHO among young adults is underscored.

This study investigated the influence of sociodemographic factors, metabolic status, and health behaviors on MetS development. By comparing young adults who transitioned from MHO to MUO with those who maintained their health status, we identified associations between the transition and factors with sex, low socioeconomic status, 5 metabolic variables, and health-related behaviors (drinking, smoking, and irregular physical activity). The nomogram was constructed to predict the transition from MHO to MUO using the data from the results.

Transitioning to MUO was higher in male participants than in female participants. In a previous study, MetS prevalence based on sex showed various results according to age. A cross-sectional study using data from the Third National Health and Nutrition Examination Survey in America showed a similar result to this study: a higher prevalence of MetS in male individuals [21]. However, they identified that women showed a higher MetS prevalence than men in older age. Using a similar perspective, a cohort study from Switzerland reported that the sex-based prevalence of MetS could be different according to the age range [22]. In Korea, the MetS fact sheet for 2010 using data from the Korean National Health and Nutrition Examination Surveys from 2007 to 2018 reported a higher prevalence in the male population but increasing trends in the female population aged >70 years [23]. Thus, our participants were constrained to young adults. Additionally, the results of this study were similar to those of previous studies. Regarding economic status, individuals with lower income levels showed a higher risk of transitioning to MUO, aligning with previous studies on risk factors for MetS development [24,25]. Our findings align with the reality that obesity is more prevalent than underweight, especially in low-income countries [1]. Furthermore, dietary issues in these regions correlate with diet-related NCDs [1]. Consequently, the findings emphasize the need to consider sex and the importance of health care provision for those in low socioeconomic positions within the community in the management of individuals with MHO to prevent transitioning to MUO.

The associations between obesity and health behaviors, that is, smoking status [26], alcohol consumption [27], and physical activity [28], have been widely investigated and linked to MetS [29]. This study further validated the significance of health-related behavior patterns in the transition from MHO to MUO in young adults with obesity. We found that moderate alcohol consumption (<30 g of alcohol daily) was associated with a decreased transition to MUO. Similar to the result, previous research [30] indicated a positive correlation between moderate alcohol consumption and a reduced risk of MetS. However, a study targeting Asian populations indicated that even moderate alcohol intake might affect metabolic status [31]. Recognizing that alcohol consumption itself could affect dietary habits in the population with obesity, as well as the possibility that the frequency of alcohol consumption might influence the occurrence of MetS more than the quantity consumed, is crucial [32]. Therefore, a particular caution, including restricting the frequency of consumption, is required regarding alcohol consumption to prevent the onset of MetS in patients with obesity.

We found that abnormal metabolic factors played a crucial role in the transition to MUO over a 4-year span. Specifically, abnormal waist circumference and triglyceride levels showed nearly ≥2 higher ORs for the transition to MUO. Waist circumference was recognized as a more pertinent obesity measure than BMI [33] and is frequently used to forecast cardiometabolic disorders [34]. Waist circumference is strongly linked to abnormal fasting glucose [35] and lipid profiles [36], both affecting metabolic states. Our findings underscore the importance of waist circumference control by demonstrating its association with the development of MUO, with an OR >1.6 for other abnormal metabolic variables. Unlike other metabolic conditions that might demand pharmacological management, waist circumference is a physical attribute that can be reduced through lifestyle changes, such as exercise, dietary adjustments, and curbed alcohol consumption [37]. Managing waist circumference is not only a strategy for combating obesity but also for improving other abnormal metabolic disorders [37]. Therefore, tailored interventions aimed at trimming waist circumference are essential. Additionally, our findings revealed a pronounced association between triglyceride levels and the transition to MUO, connecting high triglyceride levels with prevalent issues such as type 2 diabetes and abnormal HDL in people with obesity [38]. Approximately 15 years ago, a cross-sectional study in America revealed that abnormal triglyceride and HDL levels were one of the most prevalent MetS combination factors in young male participants and female participants [39]. Still, the research even proposed triglyceride levels and waist circumference as specific indicators for predicting MetS in African American people [40]. For individuals with MHO with abnormal metabolic factors, proactive management, possibly including medication control and lifestyle modifications, is vital. More intensive interventions might be warranted in abnormal waist circumference and serum triglyceride cases.

The nomogram was developed with identified factors, predicting the transition from MHO to MUO. As a simplified tool [13], the nomogram allows young adults with MHO to directly calculate their risk of transitioning to MUO with their health examination results. Created using data from the NHIS, it can serve as a valuable resource for adults undergoing regular health checkups, enabling them to assess their risk and formulate personalized health plans. Individuals can easily evaluate their risks without relying on health care professionals or medical
experts. People with MHO can efficiently use this tool to tailor their interventions by entering their sex, health behaviors, and metabolic status scores into the nomogram.

**Limitations**

The study had some limitations as a secondary analysis of Korean health examination data. Although dietary habits or nutrition consumption are critical factors for obesity or MetS, we could not include them as variables since we used publicized NHS data. Second, while clinical measurements were accurately performed in a hospital setting, health behavior results—obtained through self-response surveys—might be subjective. Third, further validation for other racial groups is needed. Differences in obesity [41] and MetS [42] prevalence among various races or ethnicities necessitate cautious adaptation and validation for broader applicability.

The other limitation of the nomogram is that prediction variables are multifaceted, comprising several modifiable factors that interact. Since changes in one variable can influence other variables, predicting outcomes based on single-variable changes can be intricate. However, enhancing these factors and embracing healthy behaviors can lead to improved metabolic status and potentially aid in obesity mitigation.

Despite the limitations, a comprehensive analysis was conducted using the data of all participants with obesity in South Korea over 4 years. The study and the developed nomogram can be applied generally to Koreans to predict the transition to MUO. For the next step, the research can be extended to people without obesity, finding associations between MHO and MUO and comparing risk factors among them. Since the nomogram can show intuitive results of predictions and be interpreted easily by any reader [13], it can be broadly used as a screening tool, leveraging national-level examination results. Coupled with advancements in digital health, telehealth, and wearable devices [43], the screening tool targeting metabolic status will play an essential role in directing appropriate interventions to those in need.

**Conclusions**

The study explored the dynamics of young adults with obesity in South Korea, pinpointing factors that correlate with the transition from MHO to MUO and creating the predictive nomogram. In this study, the scoring arrangement for transitioning to MUO revealed that the major variables were triglycerides, waist circumference, HDL cholesterol, blood pressure, and fasting glucose. Although these risk factors were highlighted in previous studies [33-36], the significance of this study lies in the development of a nomogram that enables even the general population to easily and intuitively assess their likelihood of transitioning to MUO.

With the rising prevalence of obesity and increasing intervention strategies, our screening tool simplifies the prediction using easily identifiable factors for the general population. However, this study serves as a starting point, and further research should broaden models to include racial characteristics, thereby improving the relevance of the nomogram across diverse populations.

**Acknowledgments**

The study was supported by the National Research Foundation of Korea (NRF-2022R1A2C100274). The study was presented as a poster at the 14th International Nursing Conference 2023.

**Data Availability**

The data sets generated and/or analyzed during this study are not publicly available because of the National Health Insurance Service regulation [44]. The customized database used in this study was formatted by the National Health Insurance Service in Korea, which strictly prohibits sharing the data publicly. Only permitted researchers can access the data set, and analysis can only be performed in the designated data center.

**Conflicts of Interest**

None declared.

**References**

1. Obesity and overweight. WHO. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-06-09]


44. National Health Insurance Data Sharing Service. NHISS. URL: https://nhiss.nhis.or.kr/bd/ab/bdaba000eng.do [accessed 2023-09-01]

Abbreviations

- **HDL**: high-density lipoprotein
- **MetS**: metabolic syndrome
- **MHO**: metabolically healthy obesity
- **MUO**: metabolically unhealthy obesity
- **NCD**: noncommunicable disease
- **NHIS**: National Health Insurance System
- **OR**: odds ratio
- **WHO**: World Health Organization
Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Abstract

Background: The prevalence of type 2 diabetes mellitus (DM) and pre–diabetes mellitus (pre-DM) has been increasing among youth in recent decades in the United States, prompting an urgent need for understanding and identifying their associated risk factors. Such efforts, however, have been hindered by the lack of easily accessible youth pre-DM/DM data.

Objective: We aimed to first build a high-quality, comprehensive epidemiological data set focused on youth pre-DM/DM. Subsequently, we aimed to make these data accessible by creating a user-friendly web portal to share them and the corresponding codes. Through this, we hope to address this significant gap and facilitate youth pre-DM/DM research.

Methods: Building on data from the National Health and Nutrition Examination Survey (NHANES) from 1999 to 2018, we cleaned and harmonized hundreds of variables relevant to pre-DM/DM (fasting plasma glucose level ≥100 mg/dL or glycated hemoglobin ≥5.7%) for youth aged 12-19 years (N=15,149). We identified individual factors associated with pre-DM/DM risk using bivariate statistical analyses and predicted pre-DM/DM status using our Ensemble Integration (EI) framework for multidomain machine learning. We then developed a user-friendly web portal named Prediabetes/diabetes in youth Online Dashboard (POND) to share the data and codes.

Results: We extracted 95 variables potentially relevant to pre-DM/DM risk organized into 4 domains (sociodemographic, health status, diet, and other lifestyle behaviors). The bivariate analyses identified 27 significant correlates of pre-DM/DM (P<.001, Bonferroni adjusted), including race or ethnicity, health insurance, BMI, added sugar intake, and screen time. Among these factors, 16 factors were also identified based on the EI methodology (Fisher P of overlap=7.06×10^6). In addition to those, the EI approach identified 11 additional predictive variables, including some known (eg, meat and fruit intake and family income) and less recognized factors (eg, number of rooms in homes). The factors identified in both analyses spanned across all 4 of the domains mentioned. These data and results, as well as other exploratory tools, can be accessed on POND.

Conclusions: Using NHANES data, we built one of the largest public epidemiological data sets for studying youth pre-DM/DM and identified potential risk factors using complementary analytical approaches. Our results align with the multifactorial nature of pre-DM/DM with correlates across several domains. Also, our data-sharing platform, POND, facilitates a wide range of applications to inform future youth pre-DM/DM studies.

(JMIR Public Health Surveill 2024;10:e53330) doi:10.2196/53330

KEYWORDS

youth prediabetes and diabetes; public data set; NHANES; web portal; epidemiology; biostatistics; machine learning; National Health and Nutrition Examination Survey
**Introduction**

Type 2 diabetes mellitus (DM) is a complex disease influenced by several biological and epidemiological factors [1, 2], such as obesity [3], family history [4], diet [1, 5], physical activity level [1, 6-8], and socioeconomic status [9-11]. Prediabetes, characterized by elevated blood glucose levels below the diabetes threshold, is a precursor condition to DM [12]. There has been an alarming increasing trend in the prevalence of youth with pre–diabetes mellitus (pre-DM) and DM both in the United States [13-19] and worldwide [20, 21], and the numbers of newly diagnosed youth living with pre-DM/DM are also expected to increase [14, 20, 22]. The latest estimate based on nationally representative data showed that the prevalence of pre-DM among youth increased from 11.6% in 1999-2002 to 28.2% in 2015-2018 in the United States [13]. This growth is particularly concerning because pre-DM/DM disproportionately affects racial and ethnic minority groups and those with low socioeconomic status [9-11, 22-24], leading to significant health disparities. Having pre-DM/DM at a younger age also confers a higher health and economic burden resulting from living with the condition for more years and a higher risk of developing other cardiometabolic diseases [25-30]. This serious challenge calls for increased translational research into factors associated with pre-DM/DM among youth and how they can collectively affect disease risk and inform prevention strategies.

In particular, the most critically needed research in this direction is exploring the collective impact of various risk factors across multiple health-related domains. While clinical factors, such as obesity, have been mechanistically linked to insulin resistance [31], it is important to consider the broader perspective. There is an increasing recognition that social determinants of health (SDoH) play a significant role in amplifying the risk of pre-DM/DM and their related disparities. For example, factors such as limited access to health care, food and housing insecurity, and the neighborhood-built environment have been identified as influential contributors [9-11, 32]. However, to gain a comprehensive understanding, it is essential to delve into other less studied variables, such as screen time, acculturation, or frequency of eating out, and examine how they interact to increase the risk of pre-DM/DM among youth [2].

One of the major challenges that has limited translational research into youth pre-DM/DM risk factors is that there are not publicly available, easily accessible data comprehensively profiling interrelated epidemiological factors for young individuals [2]. Specifically, most available public diabetes data portals focus on providing aggregated descriptive trends, such as pre-DM/DM prevalence for the entire population or subgroups stratified by race and ethnicity [33-36], which does not allow in-depth examination of the relationships between multiple risk factors and pre-DM/DM risk using individual-level data. While there do exist a few individual-level public diabetes data sets [37-41], they include mainly clinical measurements, while other important risk factors such as those related to diet, physical activity, and SDoH are limited. In addition, these data sets are not available for youth populations, as they focus exclusively on adult populations and not on youth specifically [37, 39-41]. Furthermore, these data sets are not accompanied by any user-friendly web-based portals that can help explore or analyze these data to reveal interesting knowledge about youth pre-DM/DM. This shows that there is a lack of a comprehensive data set that includes multiple epidemiological variables to study youth pre-DM/DM and easily usable functionalities to explore and analyze data.

To directly address this data gap, we turned to the National Health and Nutrition Examination Survey (NHANES), which offers a promising path for examining pre-DM/DM among the US youth population by providing a rich source of individual- and household-level epidemiological factors. As a result, NHANES has been a prominent data source for studying youth pre-DM/DM trends and associated factors [18, 42-45]. However, the use of NHANES data requires extensive data processing that is laborious and time-intensive [46]. This represents a major challenge for the widespread use of these high-quality and extensive data for studying youth pre-DM/DM.

In this work, we directly addressed the above challenges by processing NHANES data from 1999 to 2018 into a large-scale, youth diabetes–focused data set that covers a variety of relevant variable domains, namely, sociodemographic factors, health status indicators, diet, and other lifestyle behaviors. We also provided public access to this high-quality comprehensive youth pre-DM/DM data set, as well as functionalities to explore and analyze it, through the user-friendly Prediabetes/diabetes in youth Online Dashboard (POND) [47]. We demonstrated the data set’s use and potential through 2 case studies that used statistical analyses and machine learning (ML) approaches, respectively, to identify important epidemiological factors that are associated with youth pre-DM/DM.

Through this work, we aim to advance youth diabetes research by providing the most comprehensive epidemiological data set available through a public web portal and illustrating the value of these resources through our example case studies based on statistical analyses and ML. Our overarching goal is to enable researchers to investigate the multifactorial variables associated with youth pre-DM/DM, which may drive translational advances in prevention and management strategies.

**Methods**

**Overview**

Figure 1 [48] shows the overall study design and workflow. In the following subsections, we detail the components of the workflow.
Figure 1. Study design and workflow. We processed data from 10 survey cycles (1999-2018) from the National Health and Nutrition Examination Survey (NHANES), which yielded 15,149 youths with known pre-DM/DM status. We extracted 95 variables that were relevant to pre-DM/DM and organized them into 4 domains: sociodemographic, health status, diet, and other lifestyle behaviors. We made the data set easily accessible to the public through the user-friendly POND (Prediabetes/diabetes in youth Online Dashboard) web portal, enabling users to navigate, visualize, and download the data. In addition, we conducted 2 case studies with complementary statistical and machine learning methods that are designed to illustrate the translation potential of our data set and point. Both analyses identified predictive variables associated with youth diabetes, and the results can be explored in POND (some images in this figure were obtained from an open-source collection). DM: diabetes mellitus.

Data Source and Study Population

We built the youth pre-DM/DM data set based on publicly available NHANES data [49] spanning the years from 1999 to 2018. Developed by the Centers for Disease Control and Prevention, NHANES is a serial cross-sectional survey that gathers comprehensive health-related information from nationally representative samples of the noninstitutionalized population in the United States. The survey uses a multistage probability sampling method and collects data through questionnaires, physical examinations, and biomarker analysis. Each year, approximately 5000 individuals are included in the survey, and the data are publicly released in 2-year cycles. Figure 2 details the process used to define our study population. Briefly, of the total 101,316 participants in 1999-2018 NHANES, we excluded individuals who (1) were not within the 12-19 years age range, (2) did not have either of the biomarkers used to define pre-DM/DM status, and (3) answered “Yes” to “Have you ever been told by a doctor or health professional that you have diabetes?” The youth pre-DM/DM outcome of this work was derived as follows: youth were considered at risk of pre-DM/DM if their fasting plasma glucose (FPG) was at or greater than 100 mg/dL, or their glycated hemoglobin (HbA1c) was at or greater than 5.7%, according to the current American Diabetes Association (ADA) pediatric clinical guidelines [2].
**Validation of the Study Population**

We estimated pre-DM/DM prevalence across the 10 survey cycles (1999-2018) by incorporating the NHANES design elements in the analysis and compared the general trend with those reported in the literature [18,19]. We also specifically applied the analytical methods reported in a recent study [13] based on NHANES data to our study population to replicate the trends in pre-DM among youth in the United States from 1999 to 2018 reported in that analysis. Specifically, that study selected a youth population from 12-19 years of age with positive sampling weight from the fasting subsample (i.e., nonzero and nonmissing Fasting Subsample 2 Year Mobile Examination Centers Weight ['WTSAF2YR']; personal communication) without a self-reported physician-diagnosed DM. In addition, that study focused only on pre-DM, which was defined as an HbA1c level between 5.7% and 6.4% or an FPG level between 100 mg/dL and 125 mg/dL [13].
Development of Youth Pre-DM/DM Data Set

Based on the most recent ADA standard of care recommendations including factors related to pre-DM/DM risk and management [2], we selected 27 potentially relevant NHANES questionnaires and grouped them into 4 domains: sociodemographic, health status, diet, and other lifestyle behaviors. For example, under the health status domain, BMI was included as a potential risk factor for youth pre-DM/DM [2]. Similarly, lifestyle and behavioral variables included factors, such as diet and physical activity, that have been shown to be critical for pre-DM/DM prevention in both observational studies and randomized clinical trials [50-52]. Our sociodemographic domain included demographic, socioeconomic, and SDoH variables (eg, age, gender, poverty status, and food security). Except for commonly available clinical measurements, such as blood pressure and total cholesterol, we did not include laboratory data (eg, triglycerides, transferrin, C-reactive protein, interleukin-6, and white blood cells), since these measurements were not collected for all NHANES participants and were not commonly accessible for the general population.

From the selected questionnaires, we identified a list of 95 variables based on the aforementioned methodology. The complete list of variables is provided in Table S1 in Section S1 of Multimedia Appendix 1 [13,49,53-62] and on our POND web portal [47]. All the code developed, processed data, and detailed description of variables are also available on the web portal [47]. The process of extracting these variables involved extensive examination of the questions that were asked, consultation of the literature, and discussions to reach consensus within the study team. The details of this process are provided in Figure S1 and Section S2 of Multimedia Appendix 1. We used SAS (version 9.4; SAS Institute) and R (version 4.2.2; R Core Team, 2022) in R Studio (version 4.2.2; R Core Team, 2022) for data processing and data set development.

Building the POND

To facilitate other researchers’ use of our youth pre-DM/DM data set and make our methodology transparent and reproducible, we developed POND to share our processed data set and enable users to understand and explore the data on their own. The web portal was developed using R markdown and the flexdashboard package [63] and was published as a Shiny application [64]. Table S2 and Section S3 in Multimedia Appendix 1 provide details of all the R packages used to develop POND, and the related code is available on the portal’s download page.

Case Studies in Using the Data Set to Better Understand Youth Pre-DM/DM

Overview

To examine the validity and use of our data set for advancing translational research on youth pre-DM/DM, we conducted 2 complementary data analyses. We first conducted bivariate analyses to identify variables associated with pre-DM/DM status. We used both the aforementioned capabilities of EI to build and interpret a predictive model of youth pre-DM/DM status using NHANES data [67-69], and we have previously used these algorithms to predict pre-DM/DM status specifically among youth in a subsample of our current study population [42]. We expanded these existing analyses by taking into account the multidomain nature of our data set with the goal of building an effective and interpretable predictive model of youth pre-DM/DM. To that end, we leveraged our recently developed ML framework, Ensemble Integration (EI) [53,54], with all 4 domains and their variables in our data set. EI incorporates both consensus and complementarity in our data set by first inferring local predictive models from the individual domains, that is, sociodemographic, health status, diet, and other lifestyle behaviors, that are expected to capture information and interactions specific to the domains. These local models and information are then integrated into a global pre-DM/DM, comprehensive pre-DM/DM prediction model using heterogeneous ensemble algorithms [70] (Figure S2, Table S3, and Table S4 under Section S5 in Multimedia Appendix 1). These algorithms, such as stacking, allow the integration of an unrestricted number and variety of local models into the global predictive model, thus offering improved performance and robustness. EI also enables the identification of the most predictive variables in the final model, thus offering deeper insights into the outcome being predicted.

We used both the aforementioned capabilities of EI to build and interpret a predictive model of youth pre-DM/DM status based on our data set. We also compared the predictive performance of the model with three alternative approaches: (1) a modified form of the ADA screening guideline [55], which is based on BMI, total cholesterol level, hypertension, and race or ethnicity, to assess the use of data-driven screening for youth pre-DM/DM
(Table S5 in Multimedia Appendix 1); (2) EI applied to individual variable domains, namely, sociodemographic, health status, diet and other lifestyle behaviors, to assess the value of multidomain data for youth pre-DM/DM prediction; and (3) extreme gradient boosting (XGBoost) [71] applied to our combined multidomain data set as a representative alternate ML algorithm. This alternative was chosen as XGBoost is considered the most effective classification algorithm for tabular data [72], since it can potentially capture feature interactions across different domains [73,74]. The prediction performance of EI and all the alternative approaches were assessed in terms of the commonly used area under the receiver operating characteristic curve (AUROC) [75] and balanced accuracy (BA; average of specificity and sensitivity) [76] measures. The performance of the ML-based prediction approaches, namely, multi- and single-domain EI and XGBoost, was evaluated in a 5-fold cross-validation setting repeated 10 times [77]. These performance scores were statistically compared using the Wilcoxon rank sum test, and the resultant P values were corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure to yield false discovery rates (FDRs) [78]. More details of ML model building; the alternative approaches; and the evaluation methodology, including cross-validation, model selection, and comparison, are available in section S5 in Multimedia Appendix 1. Finally, we used EI’s interpretation capabilities [53,54] to identify the variables in our data set that were the most predictive of youth pre-DM/DM status and compare them with the variables identified from the bivariate analyses described in the above subsection.

Ethical Considerations

This study used existing deidentified and anonymized data in the public domain directly downloadable from the NHANES website and thus, according to the Common Rule, was exempt from institutional review board review and the informed consent requirement. NHANES was conducted by the Centers for Disease Control and Prevention National Center for Health Statistics. NHANES survey procedures and protocol were approved by the National Center for Health Statistics ethics review board for each survey cycle [79].

Results

Study Population Derived From NHANES

Our study population consisted of 15,149 youths aged 12-19 years who participated in the 1999-2018 NHANES cycles and met our selection criteria (Figure 2). Approximately 13.3% (2010/15,149) of US youth were at risk of pre-DM/DM according to the clinically standard criteria for defining pre-DM/DM per ADA guidelines (FPG $\geq 100$ mg/dL and HbA$_1c$ $\geq 5.7$%); Table 1).
<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall (N=15,149)</th>
<th>With pre-DM/DM(b) (n=2010; unweighted %=13.3)(b)</th>
<th>With no pre-DM/DM (n=13,139)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>15 (13-17)</td>
<td>15 (13-17)</td>
<td>16 (14-17)</td>
</tr>
<tr>
<td>Female sex, n (%)</td>
<td>7430 (49)</td>
<td>691 (34.4)</td>
<td>6739 (51.3)</td>
</tr>
<tr>
<td><strong>Race or ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black, non-Hispanic</td>
<td>4292 (28.3)</td>
<td>676 (33.6)</td>
<td>3616 (27.5)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5565 (36.7)</td>
<td>711 (35.4)</td>
<td>4854 (36.9)</td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>4033 (26.6)</td>
<td>431 (21.4)</td>
<td>3602 (27.4)</td>
</tr>
<tr>
<td>Other</td>
<td>1259 (8.3)</td>
<td>192 (9.6)</td>
<td>1067 (8.1)</td>
</tr>
<tr>
<td><strong>Insurance, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>6392 (43)</td>
<td>744 (37.7)</td>
<td>5648 (43.8)</td>
</tr>
<tr>
<td>Medicare, government, or single service</td>
<td>2026 (13.6)</td>
<td>268 (13.6)</td>
<td>1758 (13.6)</td>
</tr>
<tr>
<td>Medicaid or CHIP(c)</td>
<td>3637 (24.4)</td>
<td>564 (28.6)</td>
<td>3073 (23.8)</td>
</tr>
<tr>
<td>No insurance</td>
<td>2821 (19)</td>
<td>395 (20)</td>
<td>2426 (18.8)</td>
</tr>
<tr>
<td>Authorized for food stamps</td>
<td>7833 (69.4)</td>
<td>1037 (61.1)</td>
<td>6796 (70.8)</td>
</tr>
<tr>
<td><strong>Health status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI percentile, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight (BMI percentile &lt; 5th), n (%)</td>
<td>462 (3.1)</td>
<td>40 (2.0)</td>
<td>422 (3.2)</td>
</tr>
<tr>
<td>Normal weight (5th ≤ BMI percentile &lt; 85th), n (%)</td>
<td>8516 (56.8)</td>
<td>933 (46.8)</td>
<td>7583 (58.4)</td>
</tr>
<tr>
<td>Overweight (85th ≤ BMI percentile &lt; 95th), n (%)</td>
<td>2788 (18.6)</td>
<td>356 (17.9)</td>
<td>2432 (18.7)</td>
</tr>
<tr>
<td>Obese (95th ≤ BMI percentile), n (%)</td>
<td>3214 (21.5)</td>
<td>663 (33.3)</td>
<td>2551 (19.6)</td>
</tr>
<tr>
<td>Hypertensive(d), n (%)</td>
<td>2552 (17.4)</td>
<td>502 (26.1)</td>
<td>2050 (16.1)</td>
</tr>
<tr>
<td>High total cholesterol (≥170 mg/dL), n (%)</td>
<td>4951 (33.2)</td>
<td>707 (35.6)</td>
<td>4244 (32.8)</td>
</tr>
<tr>
<td>Fasting plasma glucose (mg/dL), median (IQR)</td>
<td>93 (88-98)</td>
<td>102 (100-106)</td>
<td>91 (86-95)</td>
</tr>
<tr>
<td>Hemoglobin A(_1c)% (%), median (IQR)</td>
<td>5.2 (5.0-5.4)</td>
<td>5.5 (5.2-5.7)</td>
<td>5.2 (5.0-5.3)</td>
</tr>
<tr>
<td><strong>Diet, median (IQR)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meals eaten out per week</td>
<td>2 (1-3)</td>
<td>2 (1-3)</td>
<td>2 (1-3)</td>
</tr>
<tr>
<td>Total grain (oz eq(e)) intake 24 hours prior</td>
<td>6.55 (4.24-9.66)</td>
<td>6.43 (4.19-9.58)</td>
<td>6.57 (4.25-9.67)</td>
</tr>
<tr>
<td>Total fruits (cup eq) intake 24 hours prior</td>
<td>0.38 (0.00-1.44)</td>
<td>0.26 (0.00-1.37)</td>
<td>0.40 (0.00-1.45)</td>
</tr>
<tr>
<td>Total vegetable (cup eq) intake 24 hours prior</td>
<td>0.88 (0.39-1.58)</td>
<td>0.84 (0.37-1.54)</td>
<td>0.89 (0.39-1.59)</td>
</tr>
<tr>
<td>Total protein (oz eq) intake 24 hours prior</td>
<td>5.29 (2.71-9.15)</td>
<td>4.73 (2.46-8.37)</td>
<td>5.38 (2.76-9.34)</td>
</tr>
<tr>
<td>Added sugar (tsp eq) intake 24 hours prior</td>
<td>20.42 (11.49-32.49)</td>
<td>20.09 (11.15-31.89)</td>
<td>20.48 (11.57-32.59)</td>
</tr>
<tr>
<td><strong>Other lifestyle behavior</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity minutes per week, median (IQR)</td>
<td>209 (45-488)</td>
<td>210 (49-476)</td>
<td>209 (45-491)</td>
</tr>
<tr>
<td>Screen time hours per day, median (IQR)</td>
<td>5 (3-8)</td>
<td>5 (3-8)</td>
<td>5 (2-7)</td>
</tr>
<tr>
<td>Exposed to secondhand smoke at home, n (%)</td>
<td>3297 (21.9)</td>
<td>469 (23.6)</td>
<td>2828 (21.7)</td>
</tr>
</tbody>
</table>

\(a\) Unweighted statistics of some key variables describing the study population in the youth pre-DM/DM data set overall and by pre-DM/DM status. More detailed statistics for all the variables in our data set can be found in the Data Exploration section of POND.

\(b\) Pre-DM/DM: pre–diabetes mellitus and diabetes mellitus.

\(c\) CHIP: child health insurance program.

\(d\) Hypertensive was defined by blood pressure ≥90th percentile or ≥120/80 mm Hg for children 13 years of age and older [2].

\(e\) eq: equivalent.
Validation of the Study Population

We estimated that the survey-weighted prevalence of pre-DM/DM in our study population rose substantially from 4.1% (95% CI 2.8-5.4) in 1999 to 22% (95% CI 18.5-25.6) in 2018 (Figure S3 and section S6 in Multimedia Appendix 1). This increasing trend of pre-DM/DM prevalence was consistent with that reported in other NHANES-based studies, which had pre-DM/DM prevalence ranging from 17.7% to 18% [18,19]. We also applied the study population and pre-DM definition criteria reported in a recent study [13] to NHANES data and derived a similarly sized study population (n=6656 vs n=6598 in the current vs previous analysis [13]) and youth pre-DM prevalence, which ranged from 11.1% (95% CI 8.9-13.3) to 37.3% (95% CI 31.0-43.6) in our analysis compared with from 11.6% (95% CI 9.5-14.1) to 28.2% (95% CI 23.3-33.6) in the study by Liu et al [13] (Table S6 in Multimedia Appendix 1).

Youth Pre-DM/DM-Focused Data Set

We extracted 95 epidemiological variables from NHANES and organized them into 4 pre-DM/DM-related domains, namely, sociodemographic, health status, diet, and other lifestyle behaviors (Table S1 in Multimedia Appendix 1). Table 1 shows the unweighted statistics of some key study population characteristics. Among youth with pre-DM/DM (n=2010), the proportion of youth who were non-Hispanic Black, non-Hispanic White, Hispanic, and other race or ethnicity (including non-Hispanic persons who reported races other than Black or White and non-Hispanic Asian) were 33.6% (n=676), 21.4% (n=431), 35.4% (n=711), and 9.6% (n=192), respectively.

Approximately, half (7719/15,149, 51%) of the population were male, and they represented 65.6% (1319/2010) of those with pre-DM/DM. Approximately 32.4% (4528/15,149) of the youth had a family income below poverty level, and 69.4% (7833/15,149) were from households receiving food stamps. The proportion of youth covered by private insurance was higher among those with than with no pre-DM/DM (5648/13,139, 43.8% vs 744/2010, 37.7%). Overall, 21.5% (3214/15,149) of the youth were obese as defined by having a BMI at or above the 95th percentile based on age and gender, and the proportion was 33.3% (663/2010) among youth with pre-DM/DM. Youth with pre-DM/DM tended to have less fruit and vegetable intake and ate lower amounts of protein and total grains than those with no pre-DM/DM. Youth with and with no pre-DM/DM showed similar amounts of physical activity with 209 and 210 minutes per week, respectively (Table 1).

Pre-DM/DM in Youth Online Dashboard

To facilitate other researchers’ use of our youth pre-DM/DM data set and make our methodology transparent and reproducible, we developed POND, which is available on [47]. Users can navigate POND through its built-in functionalities. For example, users are able to explore the details of the 95 individual variables (Figure 3A) and their distributions by pre-DM/DM status (Figure 3B), examine the risk factors of youth pre-DM/DM identified from the case studies described below (Figure 3C), as well as download the data for customized analysis and the analytical code to replicate our findings (Figure 3D). In addition, we make available all the code used to develop the data set, our case studies, and POND itself.

Figure 3. Screenshots of different functionalities available in POND (Prediabetes/diabetes in youth Online Dashboard). (A) Detailed dictionary of the 95 variables included in our youth pre-DM/DM database organized by 4 domains. (B) Data exploration section showing the distribution of user-selectable variables by pre-DM/DM status. (C) Case study section detailing the results of bivariate association analyses and the prediction of youth pre-DM/DM status from machine learning approaches. (D) Download section, where the data set and the code used in this study are publicly available to facilitate reproducibility and further exploration for interested users. ADA: American Diabetes Association; DM: diabetes mellitus; NHANES: National Health and Nutrition Examination Survey.
Case Studies Using Our Data Set to Better Understand Youth Pre-DM/DM

Overview

We examined the validity and use of our processed multidomain data set for translational studies on youth pre-DM/DM by the following 2 complementary types of data analyses.

Identifying Individual Variables Associated With Pre-DM/DM Status

In our bivariate analyses, we found 27 variables to be significantly ($P < .001$, Bonferroni adjusted) associated with pre-DM/DM status (Figure 4 and Table S7 in Multimedia Appendix 1). These variables spanned all 4 domains and included gender, race or ethnicity, use of food stamps, health insurance status, BMI, total protein intake, and screen time. Similar results were found when repeating these bivariate association tests after accounting for NHANES survey design elements (Table S7 in Multimedia Appendix 1).

Predicting Youth Pre-DM/DM Status With ML

We used an ML framework, EI [53,54], to leverage the multidomain nature of our data set and predict youth pre-DM/DM status. We also compared EI’s performance with alternative prediction approaches, most prominently the widely used XGBoost algorithm [71].

The best-performing multidomain EI methodology, stacking [75] using logistic regression, predicted youth pre-DM/DM status (AUROC=0.67; BA=0.62) more accurately than all the alternative approaches (Figure 5), namely, XGBoost (AUROC=0.64; BA=0.60; Wilcoxon rank sum FDR=1.7×10^4 and 1.8×10^4, respectively), the ADA pediatric screening guidelines (AUROC=0.57, BA=0.57; Wilcoxon rank sum FDR=1.7×10^4 and 1.8×10^4, respectively), and 4 single-domain EI (AUROC=0.63-0.54; BA=0.60-0.53; FDR <1.7×10^4 and 1.8×10^4, respectively).

The multidomain EI also identified 27 variables (the same as the number of significant variables from bivariate analyses) that contributed the most to predicting youth pre-DM/DM status. Among these variables, 16 overlapped with those identified from the bivariate statistical analyses (Figure 6; Fisher $P$ of overlap=7.06×10^-6). These variables identified by both approaches included some established pre-DM/DM risk factors such as BMI and high total cholesterol, as well as some less-recognized ones such as screen time and taking prescription drugs [2].
Figure 5. Comparison of the performance of multiple approaches for predicting youth pre-DM/DM status based on machine learning approaches. We compared the performance of the multidomain Ensemble Integration (EI) approach with 3 alternative prediction approaches. The alternative approaches were (1) a modified form of the American Diabetes Association (ADA) screening guideline (vertical red line), (2) single-domain EI-based prediction based on each of the 4 individual domains, and (3) the commonly used extreme gradient boosting (XGBoost) algorithm applied to our whole data set. Performance was measured in terms of the area under the receiver operating characteristic curve and balanced accuracy (average of sensitivity and specificity) measures. For each machine learning approach, the horizontal bar shows the average of the corresponding scores and the error bar indicates the corresponding standard error measured over 10 rounds of 5-fold cross-validation. AUROC: area under the receiver operating characteristic curve; EI: Ensemble Integration.

Figure 6. Variables associated with youth pre-DM/DM selected by bivariate analyses and the multidomain Ensemble Integration (EI) approaches. Venn diagram summarizing the overlap between the 27 significant variables identified in the bivariate analyses and the 27 most predictive variables identified from the multidomain EI model. We found that 16 variables overlapped between the 2 methods (Fisher $P=7.60\times10^{-6}$) and were drawn from all 4 domains (shown in different colors), indicating the multifactorial nature of youth pre-DM/DM. DM: diabetes mellitus; HH: household; ML: machine learning.
Discussion

Principal Findings

Leveraging the rich information in NHANES spanning nearly 20 years, we built the most comprehensive epidemiological data set for studying youth pre-DM/DM. We accomplished this by selecting and harmonizing variables relevant to youth pre-DM/DM from sociodemographic, health status, diet, and other lifestyle behaviors domains. This youth pre-DM/DM data set, as well as several functionalities to explore and analyze it, is publicly available in our user-friendly web portal, POND. We also conducted case studies using the data set with both traditional statistical methods and ML approaches to demonstrate the potential of using this data set to identify factors relevant to youth pre-DM/DM. The combination of the comprehensive public data set and POND provides avenues for more informed investigations of youth pre-DM/DM.

The future translational impact of pre-DM/DM research, facilitated by comprehensive data sets such as the one developed in this study, holds significant promise for advancing our understanding of the disease and its risk factors among youth. By enabling researchers to investigate multifactorial variables associated with pre-DM/DM, this data set contributes to several areas of research and has a broader impact on the scientific community. First, the data set’s comprehensive nature allows researchers to explore the collective impact of various risk factors across multiple health domains. By incorporating sociodemographic factors, health status indicators, diet, and lifestyle behaviors, researchers can gain a holistic understanding of the interplay between these factors and pre-DM/DM risk among youth. This knowledge can be used to generate hypotheses for further studies and inform the development of targeted interventions and prevention strategies that address the specific needs of at-risk populations. Furthermore, the data set provides an opportunity to delve into less-studied variables and their interactions in relation to pre-DM/DM risk. Variables such as screen time, acculturation, or frequency of eating out, which are often overlooked in traditional research, can be examined to uncover their potential influence on pre-DM/DM risk among youth. This expands the scope of translational research and enhances our understanding of the multifaceted nature of the disease.

One of the major contributions of our work was POND, our publicly available web portal, which provided access to all materials related to our data set and analyses, thus enabling transparency and reproducibility. Although several such portals are available in other biomedical areas, such as genomics [76-78], there is a general lack of such tools in epidemiology and public health. We hope that, in addition to facilitating studies into pre-DM/DM, POND illustrates the use of such portals for population and epidemiological studies as well.

The results of the case studies and validation exercises we conducted were also consistent with existing literature. The case studies identified known pre-DM/DM risk factors, such as gender [15,17,19], race and ethnicity [2,9,10,24], health measures (BMI, hypertension, and cholesterol) [2,55], income [9,11], insurance status [9,10], and health care availability [9,10], thus affirming the validity of the data set. In addition, our analyses revealed some less studied variables, such as screen time, home ownership status, self-reported health status, soy and nut consumption, and frequency of school meal intake, which may influence youth pre-DM/DM risk. Further study of these variables may reveal new knowledge about pre-DM/DM among youth. More generally, such novel findings further demonstrate the use of our data set and data-driven methods for further translational discoveries about this complex disorder.

Limitations

Although our work has several strengths and high potential use for youth pre-DM/DM studies, it is not without limitations. First, as our data set was derived from NHANES, we adopt limitations to the survey in our data set. Since NHANES is a cross-sectional survey, the pre-DM/DM status and its related variables provide only consecutive snapshots of youth in the United States over time across the available survey cycles. Thus, the associations identified are better suited for hypothesis generation purposes and require in-depth investigation using prospective longitudinal and randomized trial designs. In addition, we modified the ADA guideline for determining pre-DM/DM status according to variable availability. Due to the high missingness of 45% in family history (DIQ70) and the complete missingness of maternal history (DIQ175S) from 1999 to 2010 in the raw NHANES data, we were unable to include family history of diabetes in the data set. Similarly, NHANES does not provide data regarding every condition associated with insulin resistance. Therefore, we used hypertension and high cholesterol as proxies for insulin resistance. On the other hand, as our main purpose is to use POND as a conduit between this comprehensive youth pre-DM/DM database and interested researchers, our method can be adopted to longitudinal data sets should they become available in the future. Second, for the prediction of pre-DM/DM status, EI’s performance was found to be significantly better than the alternative approaches, including a modified form of the suggested guideline [45]. However, this performance assessment was based only on cross-validation, which is no substitute for validation on external data sets that is necessary for rigorous assessment. Finally, while our preliminary case study analyses identified a wide range of variables associated with youth prediabetes and diabetes, other known risk factors, such as current asthma status [80-82], added sugar consumption [83-85], sugary fruit and juice intake [83-86], and physical activity per week [6-8,50], were not identified. This limitation can be addressed by using other data analysis methods beyond our bivariate testing and ML approaches, highlighting more potential use cases of our data set.

Conclusions

Overall, the future impact of translational pre-DM/DM research facilitated by comprehensive data sets and web servers like ours extends beyond individual studies. It creates opportunities for interdisciplinary collaboration and reproducibility, strengthens evidence-based decision-making, and supports the development of targeted interventions for the prevention and management of pre-DM/DM among youth. By providing rich resources, our work can enable researchers to build upon existing knowledge.
and push the boundaries of translational pre-DM/DM research, ultimately leading to improved health outcomes for at-risk populations.

Acknowledgments
This study was enabled in part by computational resources provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai. The Ensemble Integration used in this work was implemented by Jamie JR Bennett. This work was funded by National Institutes of Health grants R21DK131555 and R01HG011407.

Data Availability
The data set and code used in this study are available at Zenodo [87] and our web portal POND [47].

Authors' Contributions
BL and GP contributed equally as cosenior and cosupervisory authors. NV, BL, and GP conceptualized the project. CM, YCL, NV, BL, and GP designed the methodology. CM and BL implemented the data curation and bivariate analyses. YCL implemented the ML case study and POND. CM and YCL conducted formal analysis and visualization. CM, YCL, NV, BL, and GP wrote the manuscript. NV, BL, and GP supervised the project.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental materials.

References


36. U.S. Chronic Disease Indicators: Diabetes | Chronic Disease and Health Promotion Data & Indicators. URL: https://vitalsigns.cdc.gov/Chronic-Disease-Indicators/UI-S-Chronic-Disease-Indicators-Diabetes/f8ti-h92k/ [accessed 2023-05-17]


38. Type 2 Diabetes Knowledge Portal. URL: https://td.hugeamp.org/ [accessed 2023-05-17]


62. Benjamini Y, Hochberg Y. Controlling the false discovery rate. a practical and powerful approach to multiple testing 1995;57(1):289-300 [FREE Full text]

63. R Markdown Format for flexible dashboards. URL: https://pgr.srutos.io/flexdashboard/ [accessed 2023-05-18]

64. Shiny. Welcome to shiny. URL: https://shiny.posit.co/r/getstarted/shiny-basics/lesson1/index.html [accessed 2023-05-18]


Abbreviations

ADA: American Diabetes Association
AUROC: area under the receiver operating characteristic curve
BA: balanced accuracy
DM: diabetes mellitus
EI: Ensemble Integration
FDR: false discovery rate
FPG: fasting plasma glucose
HbA1c: glycated hemoglobin
ML: machine learning
NHANES: National Health and Nutrition Examination Survey
POND: Prediabetes/diabetes in youth Online Dashboard
pre-DM: pre–diabetes
SDoH: social determinants of health
XGBoost: extreme gradient boosting

©Catherine McDonough, Yan Chak Li, Nita Vangeepuram, Biao Liu, Gaurav Pandey. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 02.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Review

Correlation of Socioeconomic and Environmental Factors With Incidence of Crohn Disease in Children and Adolescents: Systematic Review and Meta-Regression

Jens Weidner¹, MPH; Ingmar Glauche¹, PD, Dr rer med; Ulf Manuwald², Prof Dr; Ivana Kern³, Dr rer medic; Ines Reinecke¹, Dr rer medic; Franziska Bathelt¹⁴, Dr rer nat; Makan Amin⁵, MPH; Fan Dong³, MPH; Ulrike Rothe⁶, Prof Dr Med; Joachim Kugler³, Prof Dr Med

¹Institute for Medical Informatics and Biometry, Medical Faculty Carl Gustav Carus, TU Dresden, Dresden, Germany
²Faculty of Applied Social Sciences, University of Applied Sciences (FHD), Dresden, Germany
³Institute and Policlinic for Occupational and Social Medicine, Department of Health Sciences/Public Health, Medical Faculty Carl Gustav Carus, TU Dresden, Dresden, Germany
⁴Thiem-Research GmbH, Cottbus, Germany
⁵Department for Trauma Surgery and Orthopaedics, Park-Klinik Weissensee, Berlin, Germany
⁶GWT of TUD, Dresden, Germany
* these authors contributed equally

Corresponding Author:
Jens Weidner, MPH
Institute for Medical Informatics and Biometry
Medical Faculty Carl Gustav Carus
TU Dresden
Fetscherstrasse 74
Dresden, 01307
Germany
Phone: 49 351458 ext 89910
Email: jens.weidner@tu-dresden.de

Abstract

Background: The worldwide incidence of Crohn disease (CD) in childhood and adolescence has an increasing trend, with significant differences between different geographic regions and individual countries. This includes an increase in the incidence of CD in countries and geographic regions where CD was not previously prevalent. In response to the increasing incidence, the pediatric care landscape is facing growing challenges.

Objective: This systematic review and meta-analysis were undertaken to comprehensively delineate the incidence rates of CD in pediatric populations across different countries and to explore potential influencing factors.

Methods: We performed a systematic review of PubMed and Embase (via Ovid) for studies from January 1, 1970, to December 31, 2019. In addition, a manual search was performed in relevant and previously published reviews. The results were evaluated quantitatively. For this purpose, random effects meta-analyses and meta-regressions were performed to investigate the overall incidence rate and possible factors influencing the incidence.

Results: A qualitative synthesis of 74 studies was performed, with 72 studies included in the meta-analyses and 52 in the meta-regressions. The results of our meta-analysis showed significant heterogeneity between the individual studies, which cannot be explained by a sample effect alone. Our findings showed geographical differences in incidence rates, which increased with increasing distance from the equator, although no global temporal trend was apparent. The meta-regression analysis also identified geographic location, UV index, and Human Development Index as significant moderators associated with CD incidence.

Conclusions: Our results suggest that pediatric CD incidence has increased in many countries since 1970 but varies widely with geographic location, which may pose challenges to the respective health care systems. We identified geographic, environmental, and socioeconomic factors that contribute to the observed heterogeneity in incidence rates. These results can serve as a basis for future research. To this end, implementations of internationally standardized and interoperable registries combined with the dissemination of health data through federated networks based on a common data model, such as the Observational Medical
Outcomes Partnership, would be beneficial. This would deepen the understanding of CD and promote evidence-based approaches to preventive and interventional strategies as well as inform public health policies aimed at addressing the increasing burden of CD in children and adolescents.

**Trial Registration:** PROSPERO International prospective register of systematic reviews CRD42020168644; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=168644

**International Registered Report Identifier (IRRID):** RR2-10.1136/bmjopen-2020-037669

*(JMIR Public Health Surveill 2024;10:e48682)* doi:10.2196/48682

**KEYWORDS**
Crohn’s disease; inflammatory bowel disease; pediatric; children; adolescents; environmental factors; Crohn disease; gastroenteritis; inflammatory bowel diseases; bowel disease; digestive system; gastrointestinal disease

## Introduction

Crohn disease (CD), ulcerative colitis (UC), and indeterminate colitis are chronic inflammations of the gastrointestinal tract and are summarized under inflammatory bowel disease (IBD). Since the beginning of the 21st century, a progression in incidence, mainly due to CD, has been observed in both industrialized and emerging countries [1,2]. IBD is an immune-mediated disease that can affect people of all ages. However, about 20% of IBD cases are diagnosed before one is 20 years old, with an adverse shift in the age of diagnosis to early childhood years. Approximately 25% of children and adolescents are younger than 10 years at diagnosis and 4% are younger than 5 years [2-5]. International epidemiologic data on CD vary considerably concerning the country and calendar year, and temporal trends are also controversial [1,6]. With an annual incidence of IBD of 5 to 11 per 100,000 children and adolescents, pediatrics face growing problems [3,5,7]. The incidence of CD is significantly higher compared to UC [1,8]; therefore, the following will focus on CD.

The etiology of CD is still not definitively understood. However, the etiology seems to be multifactorial and consists of an interaction of genetic, environmental, and lifestyle factors [9-11]. For IBD, the Western lifestyle has been discussed as the cause of CD for some time [12,13]. A similar international progression of incidence has also been observed for other immune-mediated chronic diseases, and differences have been made about the influence of the Western lifestyle as measured by socioeconomic factors. For example, in their meta-analysis of diabetic ketoacidosis in type 1 diabetes, Große et al [14] identified an association between incidence and geographic as well as socioeconomic factors. Several studies also reported variations between incidence and geographic latitude for IBD [10,15]. The increase in CD incidence with latitude supports the hypothesis that higher residential sun exposure is associated with a lower risk of IBD. The results of these studies have been interpreted to suggest that low vitamin D status may be a risk factor for IBD [16]. The prevalence of vitamin D deficiency is global. Available data suggest that it occurs regardless of the development of the respective countries or the geographic latitude. Accordingly, consistent evidence indicates that the prevalence of vitamin D deficiency is highest in Asia, the Middle East, Africa, and countries with higher latitudes [17,18]. The medical and health-economic relevance of treating children and adolescents with IBD continues and is based on observations in several international studies, with the result that the number of pediatric IBD has increased and the onset of the disease seems to be shifting to early childhood. The impact of this shift in new cases is associated with a high individual as well as the societal burden of disease and will place a heavy burden on the respective health care systems [5,13,19,20]. This study aims to describe global trends in the incidence of CD since 1970 and to identify possible factors influencing the increasing incidence.

## Methods

### Study Design

A systematic review was conducted for IBD disorders. Studies were initially included from 1970 to 2019. To improve transparency in methodology, the study protocol for this review was published as “Study Protocol Epidemiology of Inflammatory Bowel Disease in Childhood and Adolescence: a Systematic Review” [21].

This systematic literature search was performed in the PubMed and Embase databases via Ovid. In addition, a manual search was performed in bibliographies of previously published and relevant systematic reviews. For detailed methodology and screening of this systematic review, we refer to the published study protocol [21].

For this study, we updated the previous systematic review from 2019 and 2020 to 2022. In this update, which was carried out until August 2022, we used the same search term as before but also included studies published up to December 31, 2021, which covered the observation period from 1970 to 2019. The complete search strategy can be viewed in Multimedia Appendix 1. The inclusion and exclusion criteria shown in Table 1 were defined for this study.
In addition, we used the mean latitude to calculate the absolute distance to the equator irrespective of the northern or southern location [15]. We extracted the UVI from the United Nations Sustainable Development Goals data from the World Health Organization database [93].

The second dimension of possible moderators included socioeconomic factors. For this purpose, we used the percentage of gross domestic product (GDP) spent on health, which we extracted from the Organisation for Economic Co-operation and Development database “Health expenditure and financing” [94]. The Human Development Index (HDI) was included in the analysis as another possible moderator. The HDI assesses a country’s developmental state and combines life expectancy at birth, expected years of schooling, and gross national income per capita [95]. The values relevant to this study were extracted from the United Nations Development Programme’s Human Development Reports [95] from 1990 onward and averaged for statistical analysis. In addition, data on the GDP of the respective included countries from the GENESIS database of the Federal Statistical Office were used for further moderator analysis [96]. Furthermore, the universal health coverage (UHC) service coverage index Sustainable Development Goals 3.8.1 was extracted from the World Health Organization database [93]. UHC quantifies coverage of essential health services and is defined as the average coverage based on tracer interventions that include reproductive, maternal, newborn, child health, infectious diseases, noncommunicable diseases, service capacity, and access among the population [97].

**Statistical Analysis**

We performed random-effects meta-analyses and meta-regressions to assess the variability of incidence rates. Analysis was performed with R (version 4.2.1.; R Foundation for Statistical Computing) software using the `metafor` package version 3.8-1 [98]. Meta-analysis was performed on a log scale (log incidence rates) using the general inverse variance method. Random effects and the extent of heterogeneity were estimated using the restricted maximum likelihood estimator. For the meta-regression, a multivariate model was constructed to control for multiple factors and to account for potential confounders.

### Table 1. Inclusion and exclusion criteria for this systematic review and meta-analysis according to the PICOS scheme.

<table>
<thead>
<tr>
<th>Population</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children and adolescents aged 0-18 years</td>
<td>Aged &gt;18 years</td>
<td>No incidence</td>
</tr>
<tr>
<td>Intervention</td>
<td>Incidence</td>
<td>No incidence</td>
</tr>
<tr>
<td>Comparison</td>
<td>Geographical characteristics, environmental factors, and economic factors</td>
<td>_b</td>
</tr>
<tr>
<td>Outcome</td>
<td>CD</td>
<td>No IBD, UC, or IBD-U studies with unreported diagnostic criteria</td>
</tr>
<tr>
<td>Type of study</td>
<td>Cohort studies and register studies, prevalence studies and cross-sectional studies (population based)</td>
<td>Case-control studies, systematic reviews, meta-analyses, and case studies</td>
</tr>
<tr>
<td>Language</td>
<td>English, French, and German</td>
<td>All others</td>
</tr>
</tbody>
</table>

*aPICOS: Population, Intervention, Comparison, Outcome, Study design.
*bNot available.

cCD: Crohn disease.
dIBD: inflammatory bowel disease.
eUC: ulcerative colitis.
fIBD-U: indeterminate colitis.

The title and abstract screening and the full-text screening were carried out independently by 2 project participants. The extraction of the data and the corresponding consistency checks were also carried out by 2 project participants. In case of disagreement, a third project participant was consulted for mediation. All included studies were evaluated for study quality. The critical appraisal tools Critical Assessment of Structure Prediction and Scottish Intercollegiate Guidelines Network were used for this purpose. In addition, a risk of bias analysis, following the procedure described in the Cochrane Handbook [22], was performed (see Multimedia Appendix 2 [5-7,23-92]). Studies of poor quality were not excluded from the quantitative synthesis to avoid loss of information.

**Ethical Considerations**

An ethics vote was not required for this systematic review because patients were not directly involved in this study.

**Data Extraction**

All included studies were independently screened for incidence rates and study characteristics using a standardized table summary of findings. In case of missing data, contact with the authors was made. The data were exported to a database and processed for statistical analysis. For studies by 1 author that reported multiple incidence rates for children and adolescents, the mean values of incidence rates and study sizes were calculated for the respective observation period.

For the planned meta-regression, we classified possible moderators of heterogeneity into 2 dimensions: geographic and environmental factors on the one hand, and socioeconomic factors on the other. Longitude and latitude as well as exposure to UV sunlight (UV radiation index [UVI]) were assigned to the first dimension of geographical and environmental factors. The geographic data were extracted from Geoplaner (version 3.1; Martin Nathensen). When studies are nationwide or involve multiple centers within a country, the mean latitude value applied to the corresponding country or area was considered. In addition, we used the mean latitude to calculate the absolute

https://publichealth.jmir.org/2024/1/e48682
identify further moderators of heterogeneity in incidence rates. The pooled incidence rates for each observation period formed the dependent variable. The observation period for each study was averaged and assigned as the starting time of the given study. The absolute distance of the included countries from the equator, UVI, HDI, health expenditure as a percentage of GDP, GDP, and UHC were included as additional independent variables in the regression model. In addition to the estimate of $\tau^2$, the Q-test for heterogeneity and the $I^2$ statistic are reported. The $I^2$ value was interpreted according to Higgins and Thompson [99] as follows: 0% to 40%, possibly insignificant; 30% to 60%, moderate heterogeneity; 50% to 90%, substantial heterogeneity; and 75% to 100%, considerable heterogeneity [99]. The influence of the moderators was evaluated using the $R^2$ statistic as a measure of the explained heterogeneity. An a priori significance level of 0.05 (5%) was set for all statistical methods. To control the risk of publication bias, statistical methods such as the Egger regression test and the rank correlation test were applied to quantitatively assess the risk of publication bias. In addition, we applied the trim-and-fill analysis and the fail-safe N-analysis (Rosenberg method) to consider and control the potential risk of publication bias.

Results

Data Basis and General Assessment of Studies

A total of 3153 studies were found from the previous systematic search conducted in 2019. The update of the systematic literature search yielded another 83 records. After removing duplicates, 77 studies were screened in the systematic literature search update. Another 5 studies from the update were included in the qualitative and quantitative synthesis. In total, the systematic literature research resulted in 81 findings from 29 countries with the search terms CD, UC, and indeterminate colitis. CD was the subject of a total of 74 studies, which were included in the qualitative synthesis of this work. Further, 2 studies had to be excluded retrospectively due to the lack of population reference. The meta-analysis included 72 studies from 26 countries and the meta-regression included 52 studies (Figure 1 [100]). The references of the included studies and a table summary of included studies can be viewed in Multimedia Appendices 3 [6,7,23-75,77-92] and 4 [7,23-75,77-92,101].

In the 72 studies eligible for meta-analysis, the incidence rates on the linear scale varied from 0.14/10^5 to 11.40/10^5. Performing a random-effects meta-analysis revealed that the overall mean incidence rate was 2.64/10^5 (95% CI 2.09 to 3.34; on log-scale –10.54, 95% CI –10.78 to –10.31), whereas the $I^2$ value of 97.88% suggests that the substantial heterogeneity of this study’s results cannot be explained by a sampling effect alone (Multimedia Appendix 5 [5-7,23-92]). In the following, we set out to identify factors that can account for the substantial dispersion in study results. Interestingly, the individual weights for each study were largely dominated by the contribution of the between-study variance while the study-specific variance (ie, the sampling effect) had a smaller effect. Consequently, the
studies in the random-effects meta-analysis have rather similar relative weights.

**Time as a Moderator of CD Incidence**

We included studies from a 50-year observation period, from 1970 to 2019. To assess whether the time point of this study influenced the CD incidence, we performed a meta-regression in which time is considered the continuous variable, whereas incidence rates are the dependent variable. Figure 2 confirms that the moderator time has no significant effect on incidence rates for CD. Moreover, time as a moderator cannot explain the heterogeneity, so the remaining heterogeneity remains substantial (test of moderators $P=0.39; I^2=97.85\%; R^2=0.00$; see also Multimedia Appendix 6-11 [5-7,23-92] subgroup analysis incidence CD in 10-year steps). These results suggest that there must be other moderators to explain the observed heterogeneity. Figure 2 also displays a slight negative trend with a simultaneous increase in heterogeneity. Some of the studies with low incidence values (depicted in the lower right corner) are from Taiwan, Finland, Saudi Arabia, Mexico, and Argentina, reinforcing the impression of greater geographical division. In the next step, we specifically examine the influence of the geographic component on CD incidence rates.

**Figure 2.** Meta-regression: incidence over time: dependent variable incidence CD, independent variable time, bubbles symbolize the studies that have been included, with each bubble’s size corresponding to the weighting assigned to the respective study (k=72, estimator: REML): test of moderators $F_{\text{test}}=0.71; P=0.39; I^2=97.85\%$; variance explanation via $R^2 \%$. CD: Crohn disease; REML: restricted maximum likelihood.

**Geographical and Environmental Factors as Moderators of CD Incidence Rates**

It is interesting to see that the highest mean incidence rates per 100,000 children and adolescents during the observation period from 1970 to 2019 were observed in Australia (11.12 new cases/10⁵), Finland (6.31/10⁵), Canada (7.12/10⁵), Germany (6.15/10⁵), and New Zealand (6.07/10⁵). The lowest incidence rates were reported in studies from countries in Asia and South America. Strikingly, an incidence of CD almost twice as high was reported in Australia compared to the other included countries (Figure 3 and forest plot geographic variation in incidence rates of CD in Multimedia Appendix 12). These data suggest geographic heterogeneity, which we first consider at the continental level.
Indeed, a meta-regression with the variable *continents* indicated that 41.34% of the heterogeneity can be explained. The test for moderators is significant ($P<.001$). Comparing this regression model with a more complex model in which we added the factor *time* to the moderator *continents*, an ANOVA showed no significant model improvement, confirming the notion that time does not act as a major moderator. In a further step, we examined the development of the incidence of CD over time for each continent individually. The results of this analysis suggest that the incidence of CD has developed differently in relation to the continents. Increasing trends were noted for North America, Europe, and Asia. For South America and Australia or the Pacific region, we found no confirmation of increasing incidence because of too few data points (see Multimedia Appendices 13 and 14).

The results also suggest that distance from the equator may affect the incidence of CD. A corresponding meta-regression, which included *absolute distance from the equator* as a moderator, showed that CD incidence increased significantly with increasing distance from the equator (Figure 4). Extrapolated to 1000 km, the incidence rate increased by 0.36%. The test for moderators yielded a significant result ($P<.001$). However, given the considerable heterogeneity in study results, distance from the equator formally contributed only moderately to better explain this variance ($R^2=29.14\%$; Table 1 and Figure 4). We found similar results when we recalculated the analysis for the *country-specific UVI* instead of the *absolute distance from the equator*. The results show that incidence rates decrease with increasing UV exposure. The results with this factor were significant in the test of moderators ($P<.001$) and 18.57% heterogeneity was resolved (Table 1 and Figure 4). Given the correlation between the moderators of absolute distance from the equator and UVI ($r=-0.87$, $P<.001$), we refrained from a joint regression model to avoid problems of collinearity and unreliable coefficient estimates.
Figure 4. Meta-regression, (A) increasing incidence with increasing distance from the equator; variable incidence CD, independent variable absolute distance from the equator (k=52, estimator: REML): test of moderators F test=18.78; P<.001; I²=96.29%; variance explanation via R² 29.14%. (B) Decreasing incidence with increasing UVI; variable incidence CD, independent variable UVI (k=52, estimator: REML): test of moderators F test=11.35; P<.001; I²=96.94%; variance explanation via R² 18.57%; bubbles symbolize the studies that have been included, with each bubble’s size corresponding to the weighting assigned to the respective study. CD: Crohn disease; REML: restricted maximum likelihood; UV: ultraviolet; UVI: ultraviolet radiation index.

Socioeconomic Factors as Moderators of CD Incidence Rates

In the next step, we investigated the extent to which socioeconomic factors could be considered moderators of heterogeneity. The results of the corresponding meta-regression showed that the HDI, health expenditure as a percent of GDP, and the UHC index acted as moderators. Accordingly, the frequency of CD increases with increasing values of each moderator (Table 2 and Figure 5). To avoid issues with collinearity and unreliable coefficient estimates resulting from the correlations between the socioeconomic factors, we decided not to use a joint regression model.

Table 2. Meta-regression results. Dependent variable: incidences CD⁵; independent variables: absolute distance to the equator and UVI⁶. ME⁷ model (k=52, estimator: REML⁸).

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
<th>P value</th>
<th>95% CI</th>
<th>I² (%)</th>
<th>R² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute distance to equator</td>
<td>0.0003</td>
<td>0.45</td>
<td>-27.06</td>
<td>&lt;.001</td>
<td>0.0001 to 0.0005</td>
<td>96.29</td>
<td>29.14</td>
</tr>
<tr>
<td>UVI</td>
<td>-0.474</td>
<td>0.141</td>
<td>-3.37</td>
<td>&lt;.001</td>
<td>-0.75 to -0.19</td>
<td>96.94</td>
<td>18.57</td>
</tr>
</tbody>
</table>

⁵CD: Crohn disease.
⁶UVI: ultraviolet radiation index.
⁷ME: mixed effects.
⁸REML: restricted maximum likelihood.
Figure 5. Meta-regression: (A) increasing incidence with increasing HDI; variable incidence CD, independent variable HDI (k=52, estimator: REML); test of moderators F test=26.4, P<.001; I²=95.87%; variance explanation via R²=40.8%. (B) Increasing incidence with increasing CHE-GDP%; variable incidence CD, independent variable CHE-GDP% (k=52, estimator: REML); test of moderators F test=18.78, P<.001; I²=96.53%; variance explanation via R²=29.4%. (C) Increasing incidence with increasing universal health coverage index SDG 3.8.1; variable incidence CD, independent variable universal health coverage (k=52, estimator: REML); test of moderators F test=17.27, P<.001; I²=96.33%; variance explanation via R²=28.86; bubbles symbolize the studies that have been included, with each bubble’s size corresponding to the weighting assigned to the respective study. CD: Crohn disease; CHE-GDP%: health expenditure as a percentage of gross domestic product; HDI: Human Development Index; REML: restricted maximum likelihood; UHC: universal health coverage; SDG: Sustainable Development Goals.

**Multifactorial Regression Model to Explain CD Incidence Rates**

In our analysis, we identified different widely independent factors - study timing, absolute distance from the equator, and HDI or UHC - as univariate moderators of CD incidence rates. To explain the high degree of heterogeneity between studies that we observed during the analyzed study period, we used a multifactorial meta-regression model that accounted for these complementary moderators as the final step of our investigation. As a result, the corresponding model showed a joint R² of 62.5%, indicating that almost two-thirds of the heterogeneity can be explained by these 3 moderators. The test for moderators was significant at P<.001 (see Tables 3 and 4).

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
<th>P value</th>
<th>95% CI</th>
<th>I² (%)</th>
<th>R² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDI</td>
<td>9.98</td>
<td>1.94</td>
<td>5.14</td>
<td>&lt;.001</td>
<td>6.18 to 13.79</td>
<td>95.87</td>
<td>40.8</td>
</tr>
<tr>
<td>CHE-GDP%</td>
<td>0.26</td>
<td>0.06</td>
<td>4.33</td>
<td>&lt;.001</td>
<td>0.15 to 0.39</td>
<td>96.53</td>
<td>29.4</td>
</tr>
<tr>
<td>GDP</td>
<td>&lt;0.00</td>
<td>0.0004</td>
<td>-0.02</td>
<td>.98</td>
<td>-0.0001 to 0.0001</td>
<td>97.57</td>
<td>&lt;0.00</td>
</tr>
<tr>
<td>Universal health coverage</td>
<td>0.1</td>
<td>0.02</td>
<td>4.12</td>
<td>&lt;.001</td>
<td>0.05 to 0.15</td>
<td>96.33</td>
<td>28.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Note</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CD: Crohn disease.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDI: Human Development Index.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHE-GDP%: health expenditure as a percentage of gross domestic product.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDP: gross domestic product.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHC: universal health coverage.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDG: Sustainable Development Goals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME: mixed effects.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REML: restricted maximum likelihood.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Meta-regression results. Dependent variable: incidences CD; independent variables: HDI, CHE-GDP%, GDP, and UHC service coverage index SDG 3.8.1. ME model (k=52, estimator: REML).
Table 4. Meta-regression results. Dependent variable: incidences CD\(^a\); independent variables: time, absolute distance from equator, and Human Development Index (HDI\(^b\))/ME\(^c\) model (k=52, estimator: REML\(^d\)). Test of moderators F test=24.57; P<.001\(^e\).

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Estimate</th>
<th>SE</th>
<th>z value</th>
<th>P value</th>
<th>95% CI</th>
<th>R(^2) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.030</td>
<td>0.01</td>
<td>3.48</td>
<td>&lt;.001</td>
<td>0.013 to 0.047</td>
<td>92.99</td>
</tr>
<tr>
<td>Absolute distance to the equator</td>
<td>0.0003</td>
<td>0.0001</td>
<td>3.69</td>
<td>&lt;.001</td>
<td>0.0001 to 0.0004</td>
<td>92.99</td>
</tr>
<tr>
<td>HDI</td>
<td>9.50</td>
<td>1.84</td>
<td>5.15</td>
<td>&lt;.001</td>
<td>1.914 to 10.153</td>
<td>92.99</td>
</tr>
</tbody>
</table>

\(^a\)CD: Crohn disease.
\(^b\)HDI: Human Development Index.
\(^c\)ME: mixed effects.
\(^d\)REML: restricted maximum likelihood.
\(^e\)multifactorial model: incidence rate ~ (time + absolute distance to the equator + Human Development Index).

Discussion

Our systematic review with meta-analysis and meta-regression examined global trends in the incidence of CD. Although several individual studies reported an increase in incidence rates for CD in a certain (national) cohort, few high-quality studies were able to substantiate and quantify such an increase (risk of bias analysis in Multimedia Appendix 2). Furthermore, some of the studies reporting temporal trends in CD incidence rates were controversially discussed [1,6]. Different study designs also made it difficult to compare incidence rates over time, which may further contribute to the substantial heterogeneity in incidence rates.

In this systematic review, we evaluated a total of 72 studies from 26 countries on the incidence of CD over a 50-year observation period. We found substantial heterogeneity in incidence rates, which was confirmed by meta-analysis using a random effects model (Cochrane Q=3349.38; P<.001; R\(^2\)=97.88%). Despite the large heterogeneity of the data, we obtained several interesting results. First, we found no clear evidence of a general global trend toward increasing CD incidence rates over time. While incidence rates might increase within individual countries, it rather appears that the inclusion of studies from a broader range of countries also increases the overall between-study heterogeneity to the extent that a global temporal trend is not identifiable. While incidence rates might increase within individual countries, it rather appears that the inclusion of studies from a broader range of countries also increases the overall between-study heterogeneity to the extent that a global temporal trend is not identifiable. This might be a result of differences in methodology and how individual studies reported incidence rates over time and needs further investigation in future research. The fact that little data were available in certain regions may also have contributed to the fact that a global trend over time was not discernible from our analyses. Kuenzig et al [102] reported similar issues in this regard. Due to the different reporting of incidence rates of IBD in childhood and adolescence, they also had difficulties in describing a clear temporal trend. Second, we observed a dependency of the incidence rates on the geographic location, with increasing incidence for countries that are further away from the equator. Third, we observed a similar effect for several socioeconomic factors, in which higher scores correlated significantly with higher CD incidence rates.

Regarding the geographic differences in incidence rates of CD, several studies reported a north-south gradient. For example, Nerich et al [15] reported the effects of latitude on the geographic distribution of CD. However, quantification of the gradient by latitude was not performed. Armitage et al [23] similarly reported a significant north-south gradient for CD in Scotland. Since recent epidemiologic studies have reported an increasing incidence of IBD worldwide, including in southern countries, particularly also in the southern hemisphere [2,103], we decided to use absolute distance from the equator as a factor to represent and quantify a relationship between incidence and geographic location. The result of our meta-regression showed that incidence rates increased with increasing distance from the equator. This result corresponds with the results of our further analysis of possible moderators of incidence rates. We found that countries with a high UVI, other than Australia, have a lower incidence of CD. Our results correlate with findings from other studies showing that higher exposure to UV radiation, or sunlight, is associated with a lower risk of CD and IBD [16,104]. In our results, Australia stood out with a high incidence. Although Australia is considered sun-rich, 17% of Australian adolescents have vitamin D deficiency [104]. Vitamin D is formed in the skin when exposed to UV radiation. We therefore suspect a correlation between vitamin D status and the incidence of CD. Further studies should therefore examine in particular whether a low vitamin D status is a risk factor for CD or IBD or a consequence. However, studies show that patients with IBD and especially CD also show a low vitamin D status and indicate a correlation with disease activity. Unfortunately, disentangling causality and correlation is an unresolved challenge in the ongoing debate about the interplay between vitamin D and IBD [105-107].

Although latitude or absolute distance from the equator and UV exposure are correlated, they cannot be fully replaced in our statistical analysis.

Our results concerning the socioeconomic factors contribute to the hypothesis that CD might correlate with industrialized, urbanized societies, largely due to a Western lifestyle and other associated environmental factors [13] which themselves go along with higher socioeconomic scores. It is also known that...
the incidence and prevalence of CD vary between countries with different HDIs [108-110]. Although there are few epidemiologic studies of CD in low- and middle-income countries, the incidence of CD is increasing significantly worldwide, affecting even countries previously considered to be at low risk [103,111]. It has been observed that the incidence and prevalence in middle-income countries are also increasing in children and adolescents, which has been attributed to the rapid modernization and Westernization of the population [13]. Our findings seem to follow a global pattern, namely that the process of industrialization has an impact on the incidence of CD. In this regard, we also follow the view of Takahashi et al [109] and Ananthakrishnan et al [103] that the level of development of countries and Western lifestyles are related to the level of incidence. However, causality cannot be inferred from our study. Further research is needed for this purpose. In recent times, there has been a growing significance attributed to observational research conducted on real-world data, leading to the establishment of global research networks, exemplified by the Observational Health Data Sciences and Informatics community. These networks aim to facilitate large-scale studies grounded in the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). Consequently, the use of observational studies that use real-world data is a valuable way to study CD and IBD in the future [112].

Limitations
This study is limited by the use of 2 databases for systematic literature search. Another limitation of this study is the exclusion of studies that were not published in English, Spanish, French, or German. Given that the included studies were mainly from countries with good access to the health care system (UHC >70), the underreporting of countries with poorer access to the health care system should be discussed. We controlled the risk of publication bias using the Eggers regression test, rank correlation test, trim and fill analysis, and fail-safe N analysis (Rosenberg method). Although these methods did not statistically indicate a bias due to publication bias, a small bias cannot be completely ruled out.

Conclusions
Based on available study data from 1970 to 2019, we could not identify a global, temporal trend toward increasing CD incidence rates, although these effects are reported for individual countries or regions. Instead, we could demonstrate that a substantial part of the observed heterogeneity between the published study results can be explained by geographic location and socioeconomic factors. Our study can be used to provide quantitative estimates of these trends for CD in childhood and adolescence. However, to establish causal relationships regarding potential risk factors, further studies are necessary, including those conducted in countries with lower levels of development. Nevertheless, our analysis provides valuable information to drive future research and health policies aiming to reduce the incidence of CD among children and adolescents. This needs continuous global monitoring of the incidence of IBD in childhood and adolescence to fully understand the trends in IBD incidence [102]. To this end, the implementation of internationally standardized and interoperable registries, coupled with the dissemination of health data via federated networks grounded on a CDM, such as the OMOP CDM, is deemed advantageous. The OMOP CDM aligns most closely with the requisites conducive to expediting data exchange within longitudinal studies [112,113]. The usage of such registries and data networks holds the potential to streamline the exhaustive and standardized accumulation as well as dissemination of data. This, in turn, would enhance our comprehension of CD and foster evidence-based approaches for preventive and interventional strategies.

Acknowledgments
We gratefully acknowledge support from the Sächsische Landesbibliothek—Staats- und Universitätsbibliothek/Technische Universität Dresden Open Access Publication Fund. The funders for the publication did not influence this study’s design, data collection, analysis, decision to publish, or preparation of this paper. This paper was written and finalized by the authors without the help of artificial intelligence such as ChatGPT or other generative language models. A vote of the ethics committee was not required for this study because it was a systematic literature review and no patients were directly involved. This study was conducted without third-party funding or support.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors’ Contributions
Conceptualization was done by JW, UR, and UM. JW performed on the methodology, writing of and preparation of the original draft, and administration. Systematic literature search and extraction were by JW and UM. Data analysis was performed by JW and IG. The writing, review, and editing of this paper were done by IG, UM, IR, FB, JK, UR, FD, and MA. JW, FB, and IR worked on the visualization. Supervision was handled by JK and IG. IG handled the statistical supervision. All authors have read and agreed to the published version of this paper.

Conflicts of Interest
None declared.
Multimedia Appendix 13
Meta-regression, subgroups continents, dependent variable incidence CD (Crohn disease), independent variable: time, bubbles symbolize the studies that have been included, with each bubble's size corresponding to the weighting assigned to the respective study.

[**PNG File**, 409 KB - publichealth_v10i1e48682_app13.png]

Multimedia Appendix 14
Meta-regression results: dependent variable incidences CD (Crohn disease), independent variables: time, 1 too few observations for regression analysis, pooled IR (incidence rate) 0.23/105 (on log scale –12.96), a multiplicative change factor >1 increasing IR, <1 decreasing IR.

[**PDF File (Adobe PDF File)**, 63 KB - publichealth_v10i1e48682_app14.pdf]

Multimedia Appendix 15
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.

[**PDF File (Adobe PDF File)**, 85 KB - publichealth_v10i1e48682_app15.pdf]

References


34. Cosgrove M, Al-Atia RF, Jenkins HR. The epidemiology of paediatric inflammatory bowel disease. Arch Dis Child 1996;74(5):460-461 [FREE Full text] [doi: 10.1136/adc.74.5.460] [Medline: 8669668]


Abbreviations

CD: Crohn disease
CDM: common data model
GDP: gross domestic product
HDI: Human Development Index
IBD: inflammatory bowel disease
OMOP: Observational Medical Outcomes Partnership
UC: ulcerative colitis
UHC: universal health coverage
UV: ultraviolet
UVI: ultraviolet radiation index

Edited by A Mavragani; submitted 03.05.23; peer-reviewed by J Luo, J Walldorf, B Elsner; comments to author 07.12.23; revised version received 28.12.23; accepted 23.01.24; published 25.03.24.

Please cite as:

URL: https://publichealth.jmir.org/2024/1/e48682
doi:10.2196/48682
PMID:38526534
Original Paper

Effectiveness, Safety, and Acceptability of Primaquine Mass Drug Administration in Low-Endemicity Areas in Southern Thailand: Proof-of-Concept Study

Jaranit Kaewkungwal¹, PhD; Wanlapa Roobsoong², PhD; Saranath Lawpoolsri¹, PhD; Wang Nguitragool³, PhD; Suwich Thammaphalo⁴, PhD; Pathomporn Prikchoo⁴, MSc; Amnatt Khamsiriwatchara⁵, MSc; Rungrawee Pawarana⁵, BSc; Pawinee Jarujaree⁶, MSc; Daniel M Parker⁶,⁷, PhD; Piyrarut Sripoorote², MSc; Mondha Kengkanpanich⁸, PhD; Chetta Ngamjarus⁹, PhD; Jetsumon Sattabongkot², PhD; Liwang Cui¹⁰, PhD

¹Department of Tropical Hygiene, Mahidol University, Bangkok, Thailand
²Mahidol Vivax Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand
³Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand
⁴Ministry of Public Health, Nonthaburi, Thailand
⁵Department of Population Health and Disease Prevention, University of California, Irvine, Irvine, CA, United States
⁶Department of Epidemiology and Biostatistics, University of California, Irvine, Irvine, CA, United States
⁷Department of Health Education and Behavioral Sciences, Faculty of Public Health, Mahidol University, Bangkok, Thailand
⁸Department of Epidemiology and Biostatistics, Faculty of Public Health, Khon Kaen University, Khon Kaen, Thailand
⁹Department of Infectious Diseases and Internal Medicine, Department of Internal Medicine, University of South Florida, Tampa, FL, United States

Corresponding Author:
Jaranit Kaewkungwal, PhD
Department of Tropical Hygiene
Mahidol University
Rajavithi Road
Bangkok, 10400
Thailand
Phone: 66 0871001951
Email: jaranitk@biophics.org

Abstract

Background: A challenge in achieving the malaria-elimination target in the Greater Mekong Subregion, including Thailand, is the predominance of Plasmodium vivax malaria, which has shown extreme resilience to control measures.

Objective: This proof-of-concept study aimed to provide evidence for implementing primaquine mass drug administration (pMDA) as a strategy for P. vivax elimination in low-endemicity settings.

Methods: The study employed a mixed-methods trial to thoroughly evaluate the effectiveness, safety, acceptability, and community engagement of pMDA. The quantitative part was designed as a 2-period cluster-crossover randomized controlled trial. The intervention was pMDA augmented to the national prevention and control standards with directly observed treatment (DOT) by village health volunteers. The qualitative part employed in-depth interviews and brainstorming discussions. The study involved 7 clusters in 2 districts of 2 southern provinces in Thailand with persistently low P. vivax transmission. In the quantitative part, 5 cross-sectional blood surveys were conducted in both the pMDA and control groups before and 3 months after pMDA. The effectiveness of pMDA was determined by comparing the proportions of P. vivax infections per 1000 population between the 2 groups, with a multilevel zero-inflated negative binomial model adjusted for cluster and time as covariates and the interaction. The safety data comprised adverse events after drug administration. Thematic content analysis was used to assess the acceptability and engagement of stakeholders.

Results: In the pre-pMDA period, the proportions of P. vivax infections in the pMDA (n=1536) and control (n=1577) groups were 13.0 (95% CI 8.2-20.4) and 12.0 (95% CI 7.5-19.1), respectively. At month 3 post-pMDA, these proportions in the pMDA (n=1430) and control (n=1420) groups were 8.4 (95% CI 4.6-15.1) and 5.6 (95% CI 2.6-11.5), respectively. No statistically significant differences were found between the groups. The number of malaria cases reduced in all clusters in both groups, and
Conclusions: pMDA was associated with high adherence, safety, and tolerability, but it may not significantly impact \( P. vivax \) transmission. As this was a proof-of-concept study, we decided not to scale up the intervention with larger clusters and samples. An alternative approach involving a targeted primaquine treatment strategy with primaquine and DOT is currently being implemented. We experienced success regarding effective health care workforces at point-of-care centers, effective collaborations in the community, and commitment from authorities at local and national levels. Our efforts boosted the acceptability of the malaria-elimination initiative. Community engagement is recommended to achieve elimination targets.

Trial Registration: Thai Clinical Trials Registry TCTR20190806004; https://www.thaichronictrials.org/show/TCTR20190806004

(JMIR Public Health Surveill 2024;10:e51993) doi:10.2196/51993

KEYWORDS

mass drug administration; cluster-crossover randomized controlled trial; community-based trial; participatory epidemiology; \( P. vivax \); primaquine

Introduction

Background

As part of the World Health Organization (WHO) campaign for “zero malaria,” the Greater Mekong Subregion (GMS) of Southeast Asia has developed a strategic plan for regional malaria elimination by 2030, while Thailand aims to achieve this goal by 2024 [1,2]. One major challenge facing the GMS is the predominance of \( P. vivax \) malaria [3,4], which has shown extreme resilience to control measures [5,6]. Initiatives to eliminate malaria have an excellent impact on \( P. falciparum \) but not on \( P. vivax \) worldwide due to the various unique aspects of \( P. vivax \) biology. One challenge is asymptomatic infections with \( P. vivax \), which are undetected and untreated, potentially contributing to transmission over several weeks or months [7]. Asymptomatic infections are especially common in low-endemicity areas where control tools have reduced the malaria burden [8-10]. A study along the Thailand-Myanmar border noted that while the proportion of severe \( P. vivax \) malaria varied across different geographic regions and transmission settings, a significant proportion of the community had asymptomatic parasitemia, even in low-transmission areas [9].

The identification and appropriate treatment of asymptomatic infected individuals, who are typically missed by clinical case-based surveillance, have become critical for interrupting malaria transmission in the final elimination phase [11,12]. Several approaches have been proposed to accelerate \( P. vivax \) elimination, for example, novel serological test-and-treat interventions, radical cure strategies, case-centered surveillance and response systems, and mass drug administration (MDA) [13-15]. To deal with low blood parasitemia and the formation of hypnozoites associated with \( P. vivax \) infection that evade conventional diagnosis, presumptive preventive treatment of an endemic population by MDA using a hypnozoiticidal drug, such as primaquine, is often the chosen strategy to eliminate residual \( P. vivax \) transmission [15-17]. Large-scale MDA with pyrimethamine and PQ was associated with decreased \( P. vivax \) transmission in central and southern China [14].

The WHO mentioned the lessons learned from MDA implementation (so-called “mass primaquine preventive treatment [MPPT]”) in several temperate countries, for example, MPPT combined with vector control and other preventive measures resulted in the rapid containment of \( P. vivax \) epidemics and may have contributed to the interruption of transmission in low-transmission settings [18]. Prior to 2019, there were no data on the implementation of MPPT in tropical and subtropical areas. In 2019, when this project started in Thailand, the WHO reported 2 MDA studies in the GMS with differing results, with one study demonstrating only a short-term reduction in \( P. vivax \) transmission and the other study finding no effect [18]. It appears that MDA with PQ has been successful in eliminating \( P. vivax \) in the temperate zone [11,15,19,20], but its applicability and effectiveness for eliminating \( P. vivax \) in tropical countries remain to be evaluated.

Therefore, as part of the International Center of Excellence for Malaria Research (ICEMR) project, supported by the US National Institutes of Health (U19 AI089672), this study was conducted to evaluate the effectiveness, safety, and acceptability of MDA with PQ in low-endemicity areas in a tropical country. Particularly, this study was considered a proof-of-concept evaluation of MDA with PQ (termed primaquine mass drug administration [pMDA]) using a cluster-crossover randomized trial design to provide the evidence base for designing and implementing a pMDA strategy in low-malaria-endemic settings in Thailand. With the planning of the pMDA strategy by the National Malaria Control Program in Thailand, we also sought to critically assess the acceptability and engagement of stakeholders at various levels. According to the main ICEMR project, if pMDA (phase I) was proven to be effective, safe, and acceptable, the pMDA intervention would be scaled up to cover over 100,000 villagers in 200 clusters in 8 provinces (phase II) using a stepped-wedge design to provide a statistically robust evaluation of pMDA.

Objectives

The goal of this proof-of-concept study was to thoroughly assess the effectiveness, safety, acceptability, and community engagement of pMDA as a strategy for malaria elimination in Thailand. The specific aims were as follows: (1) to compare the...
proportions of P. vivax infections between communities under the national standard of prevention and control (SPC) program and communities under the SPC program augmented with pMDA, (2) to assess the safety of PQ implemented in the communities, and (3) to assess the acceptability and stakeholder engagement of the pMDA intervention for possible scale-up to phase II.

Methods

Trial Design
The study employed a mixed-method design comprising quantitative and qualitative approaches. The quantitative part, which was used for assessing the effectiveness and safety of the pMDA intervention, adopted a cluster-crossover randomized controlled design, a modified cluster randomized design, which is particularly feasible in pragmatic clinical trials in health care systems [21,22]. This design was selected as it is suitable for determining the effectiveness of a routinely used intervention in health care practice, in which the intervention is randomized (turned on and off) at the community level instead of the individual level [23]. In this study, clusters were randomized to a sequence of treatment conditions. A group of clusters received pMDA in the first period (year 1) and the SPC program in the second period (year 2), while the other group of clusters received these preventive activities in reverse order.

The qualitative part, which was used for assessing the acceptability and stakeholder engagement of the pMDA intervention, used in-depth interviews (IDIs) and brainstorming sessions among stakeholders. Stakeholders were primarily sensitized before the study started and informed regularly during the study period.

Settings and Locations
The study was conducted in 2 districts of 2 southern provinces in Thailand (Yala and Narathiwat), which had persistently low P. vivax malaria transmission. The study sites comprised 7 purposively selected villages, the smallest administrative unit in Thailand, with each typically having about 200-400 residents. The 7 villages were selected according to their reported malaria incidences in 2018 from routine passive case detection in the malaria surveillance system of Thailand (unpublished data). Five villages (#1, 2, 4, 5, and 6) were in Yala Province, with P. vivax incidence rates of 4%-7%, and 2 had unexpectedly high rates (over 30%) in 2018. Two villages (#3 and 7) were in Narathiwat Province, with P. vivax incidence rates of 1%-2%. For the study design, one group of villages served as the pMDA treatment group, while the other served as the control group. In the following year, the intervention was swapped between the 2 groups of villages. A map showing locations and distances among the 2 sets of 7 clusters (clusters 1-3 vs clusters 4-7) in the 2 provinces is presented in Figure 1. The nearest distances between 2 sets of clusters were 3.7-4.5 km. It should be noted that primary health care management in each village was independent. On average, a village health volunteer (VHV) is responsible for providing health care services to 10 households in the village. Each VHV assists the local health workers in promoting health, preventing diseases, and providing basic health services to local communities. In this study, VHVs were assigned to implement pMDA and collect data from the study participants in the nonoverlapping households for which they were responsible.
Study Participants

For the quantitative study, all villagers living in the selected clusters were invited to join the study if they were aged 1 year or older. For PQ administration, the exclusion criteria were as follows: (1) pregnancy and lactation, (2) age <7 years, (3) glucose-6-phosphate dehydrogenase (G6PD) deficiency, (4) hemoglobin level <8 g/dL, (5) body weight <15 kg, (6) history of allergy to PQ, and (7) history of drug reactions, such as hemolysis or dark urine, after PQ intake.

For the qualitative study, key informants included stakeholders representing the following 4 levels of engagement and participation: national level (malaria experts or consultants, Thailand Ministry of Public Health [MOPH]), regional and provincial level (health officers or authorities), district level (personnel at operational units of district hospitals), and local level (community leaders, religion gatekeepers, and VHVs). Villagers in the community were also invited, regardless of whether they participated in pMDA activities.

Intervention

The control group was exposed to the SPC program, a routine malaria prevention and control program, implemented by the MOPH. The SPC program includes routine case reports, case investigations, and disease- or vector-control activities at the village level. Besides routine activities, eligible subjects within the pMDA group received a dose of 0.25 mg/kg of PQ daily for 14 days. The 0.25 mg/kg dose was selected according to the WHO recommendation and evidence supported by previous studies, as it is well tolerated in G6PD-normal individuals [11,24]. Those who were G6PD deficient were excluded from PQ administration. Directly observed treatment (DOT) was performed to ensure compliance.

Outcomes and Study Procedures

For the quantitative study, the primary outcome was the proportion of P. vivax infections among the study participants within each cluster before and after 2 rounds of pMDA. The secondary outcome was the safety of the study participants who took PQ, which was closely monitored during the 14-day treatment. The study procedure is shown in Figure 2. Baseline demographic information was collected using a structured questionnaire at the start of the study. One-time testing for G6PD was performed for all villagers, using the qualitative CareStart G6PD rapid diagnostic test, before pMDA to assess the eligibility of study participants to receive PQ. For each round of pMDA, an initial cross-sectional blood survey (CSS) was conducted in both the pMDA and control groups before pMDA implementation, and a follow-up CSS was performed 3 months after pMDA. As an additional postintervention follow-up, a CSS was performed for both groups 6 months after the second round of pMDA. At each survey, finger-prick blood was collected from each participant to prepare dried filter-paper blood spots, which were later used for Plasmodium detection by quantitative polymerase chain reaction (qPCR) to identify asymptomatic infection cases in the CSS population. At the
time of each survey period, those in the CSS population who had clinical symptoms and were detected by the microscopic method in the routine MOPH malaria surveillance system were also subsequently verified by qPCR and counted as confirmed \(P. \text{vivax}\)–infected cases. The prevalence, defined as the proportion of confirmed \(P. \text{vivax}\) infections at the community level in different periods, was calculated from the number of PCR-confirmed \(P. \text{vivax}\) cases (both asymptomatic and symptomatic cases) divided by the total number of cases in each round of the CSS.

**Figure 2.** Study procedures to assess the effect of primaquine mass drug administration (pMDA) for reducing \(P. \text{vivax}\) infection in the community in a 2-period cluster-crossover randomized controlled trial during the period 2019-2022. CSS: cross-sectional blood survey; DOT: directly observed treatment; FU: follow-up; G6PD: glucose-6-phosphate dehydrogenase.

For the qualitative study, the primary outcome was information about the acceptability and stakeholder engagement of the pMDA intervention implemented in the localities. Information was collected using IDIs at the study participants’ homes and community meeting places. A brainstorming meeting at the regional health office was also arranged for collective opinions among representative authorities at the national, regional, and provincial levels. Data collection was performed after round 2 at the conclusion of the study.

**Sample Size**

For the quantitative study, the sample size was calculated with the notation that it was a proof-of-concept study to assess the potential impact of pMDA implementation. A go or no-go decision to scale up the study with a larger sample size would be made based on the potential effectiveness and feasibility of the intervention according to the interclass cluster correlation effect obtained from this study. We planned for a sample size of 1500 study participants per group. With this sample size, the power to detect differences between the 2 groups varied according to the baseline proportions of \(P. \text{vivax}\) in the study areas, with a 2-tailed type I error of 5%. With a sample size of 1500 per group, when the baseline proportion of \(P. \text{vivax}\) in the cluster is low at 3% (30 in 1000) and the effect sizes (difference between the pMDA and control groups) are 50% and 75%, powers of 79.1% and 99.5%, respectively, would be achieved. In areas with a very low baseline at 1% (10 in 1000) and effect sizes of 50% and 75%, the study would have detection powers of 35.5% and 74.1%, respectively.

For the qualitative study, IDIs were planned to include study participants from 24 families and at least six key informants from 2 villages of each province, including those who completed PQ administration, those who did not complete pMDA, and those who did not participate in pMDA. IDIs were planned for stakeholders in pMDA areas, including 12 VHV’s (at least six per province), 4 health care authorities (district and provincial levels), and 4 local leaders or gatekeepers (village heads and religious masters). A brainstorming meeting at the conclusion of the study was planned with 15 representatives, including MOPH consultative experts, authorities from regional and district vector-borne units, authorities from regional and district health offices, staff from hospitals at the subdistrict level, and staff who worked in the selected areas.

**Randomization**

To plan and evaluate the intervention, we subdivided the selected study areas into villages, which are the smallest administrative units in Thailand. A village is considered a cluster and treated as the unit of randomization. The sizes of the purposively selected clusters by authorities in the study areas were about 150-250, with a larger size of around 600 residents. Considering the cluster sizes and the distances between clusters, we purposively allocated 2 out of 5 villages in Yala and 1 out of 2 villages in Narathiwat to receive the pMDA intervention in round 1 (year 1) and switch to the control group in round 2 (year 2).

**Statistical Methods**

The potential effectiveness of pMDA was explored in terms of reductions in the prevalences and proportions of confirmed \(P. \text{vivax}\) infections on qPCR in the CSS population, with a comparison between the intervention and control groups. This was a 2-period cluster-crossover study involving 2 cluster periods and multiple CSSs before and after pMDA in each period. As suggested in the literature, the data should be analyzed with hierarchical models with random effects in order
to allow for different outcome probabilities in each period, cluster, and cluster period [25,26]. With a cluster-level randomization design, it is statistically more efficient to employ a model adjusted for the cluster effect [22]. As a proof-of-concept for a pragmatic trial in the community setting, the potential effectiveness of pMDA was thus determined by comparing the proportions of P. vivax infections between the 2 groups at each period of CSS, with the 2 periods combined representing the overall impact of the intervention. An additional comparison of the proportions of P. vivax infections between the 2 groups at 6 months after round 2 was also performed to explore the longer-term effect. As the study was conducted in low-endemic areas, it is more likely that most study participants would not be infected. Therefore, the analysis used the R package NBZIMM, which provides functions for setting up a multilevel zero-inflated negative binomial model adjusted for random intercepts of the clusters [27,28]. The model for the overall impact of the intervention in the 2-period crossover was adjusted for the cluster effect plus time period and the interaction between the time period and intervention. The prevalence ratio and its 95% CI from the model have been reported, and statistical significance was set at \( P < .05 \).

The safety of the pMDA strategy was monitored in terms of adverse events that occurred after drug administration throughout the follow-up period. All events have been presented descriptively. The prevalence of G6PD deficiency has been described with its 95% CI.

The acceptability and stakeholder engagement of the pMDA strategy were examined using qualitative analysis. IDIs and brainstorming sessions were led by trained facilitators, including an experienced lead facilitator, 4 co-facilitators and note-takers, and 2 local health care staff. All team members had been trained on ethical considerations for human research subjects and had read the study protocol and data collection methodologies. They discussed the important points to be explored. Audio recordings of the IDIs and meetings were transcribed, and the transcripts were reviewed and classified into key themes according to the study objectives and themes that emerged during the reviews. Thematic analyses were conducted on the notes of participant responses and the determined themes. The content was analyzed to identify themes by manually exploring, interpreting, and categorizing the data via consensus among facilitators. The priority themes set according to the study objectives included the following: perceptions, expectations, engagement, factors influencing the decision to participate or not to participate, factors influencing noncompletion of the 14-day regimen, perceived achievements, blockage and solutions in pMDA, and challenges in pMDA implementation.

### Ethical Considerations

This study was approved by the Ethics Committee of the Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand (approval number: MUTM 2019-031-01). All participants consented to participate in this study before enrollment. Participants younger than 18 years were consented or assented along with consent from their parents. Community leaders and authorities in national and local health facilities were informed. Moreover, they provided consent and were involved as part of the community sensitization prior to the study initiation. All staff responsible for research activities were trained in human subject research protection. The identifiable information of study participants was treated as confidential information, and the participants' identification numbers were coded. To treat pMDA as an additional part of the routine malaria prevention activities by health care personnel and VHVs in natural settings, no compensation was provided to the consented study participants. Generative AI was not used in any portion of manuscript preparation.

### Results

#### Study Participant Characteristics and Intervention Implementation and Coverage

A total of 2550 individuals resided in 7 clusters. In round 1, 1624 participants consented to the CSS, with 925 in clusters 1-3 and 699 in clusters 4-7 being allocated to the pMDA and control groups, respectively. In the 3 villages allocated to pMDA, 70.8% (655/925) of individuals were eligible according to the inclusion and exclusion criteria, and 87.8% (575/655) received pMDA. The follow-up CSS at 3 months after PQ administration was performed among 1423 study participants, with 92.0% (851/925) in the pMDA clusters versus 81.8% (572/699) in the control clusters.

In round 2, the clusters were crossed over, and 1489 participants consented to the CSS, with 878 in clusters 1-3 and 611 in clusters 4-7 being switched to the control and pMDA groups, respectively. In the 4 villages that switched to pMDA, 66.8% (408/611) of individuals were eligible according to the inclusion and exclusion criteria and 91.9% (375/408) received pMDA. The follow-up CSS at 3 months after pMDA in round 2 was conducted among 1427 study participants, with 94.8% (579/611) in the pMDA clusters and 96.6% (848/878) in the control clusters. The additional postintervention CSS at 6 months after the second PQ administration was performed among 1401 study participants, with 92.9% (568/611) in the pMDA group and 94.9% (833/878) in the control group. Figure 3 presents the CONSORT (Consolidated Standards of Reporting Trials) diagram showing the flow of participants throughout the trial [29,30]. The CONSORT checklist is presented in Multimedia Appendix 1.

The demographic attributes and G6PD deficiency statuses of the study participants in the 7 clusters are shown in Tables 1 and 2. Based on the census data, there were slightly fewer male participants than female participants, with an age range of 1 to 96 years in both rounds of pMDA. Testing for G6PD deficiency revealed an overall prevalence of 5.4% (95% CI 4.3%-6.6%) in the 7 clusters. G6PD-deficient individuals were excluded from PQ administration.
Figure 3. CONSORT (Consolidated Standards of Reporting Trials) diagram showing the flow of participants throughout the 2-period cluster-crossover trial. CSS: cross-sectional blood survey; G6PD: glucose-6-phosphate dehydrogenase; G6PDd: glucose-6-phosphate dehydrogenase deficiency; HB: hemoglobin; Hx: history; pMDA: primaquine mass drug administration; PQ: primaquine.
Table 1. Baseline characteristics of the participants for round 1 of the primaquine mass drug administration and control clusters during the 2-period cluster-crossover sequence.

<table>
<thead>
<tr>
<th>Variable</th>
<th>pMDA&lt;sup&gt;a&lt;/sup&gt; clusters</th>
<th>Control clusters</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
</tr>
<tr>
<td>Pre-MDA&lt;sup&gt;b&lt;/sup&gt; CSS&lt;sup&gt;c&lt;/sup&gt;, n</td>
<td>219</td>
<td>175</td>
<td>531</td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>82 (37.4)</td>
<td>81 (46.3)</td>
<td>242 (45.6)</td>
</tr>
<tr>
<td>Age (years), median (range)</td>
<td>32 (1-74)</td>
<td>29 (2-80)</td>
<td>40 (1-91)</td>
</tr>
<tr>
<td>G6PD&lt;sup&gt;d&lt;/sup&gt; deficiency, n (%)</td>
<td>12 (5.5)</td>
<td>7 (4.0)</td>
<td>34 (6.4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>pMDA: primaquine mass drug administration.  
<sup>b</sup>MDA: mass drug administration.  
<sup>c</sup>CSS: cross-sectional blood survey.  
<sup>d</sup>G6PD: glucose-6-phosphate dehydrogenase.  
<sup>e</sup>Not applicable.

Table 2. Baseline characteristics of the participants for round 2 of the primaquine mass drug administration and control clusters during the 2-period cluster-crossover sequence.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control clusters</th>
<th>pMDA&lt;sup&gt;a&lt;/sup&gt; clusters</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
</tr>
<tr>
<td>Pre-MDA&lt;sup&gt;b&lt;/sup&gt; CSS&lt;sup&gt;c&lt;/sup&gt;, n</td>
<td>226</td>
<td>189</td>
<td>463</td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>88 (38.9)</td>
<td>88 (46.6)</td>
<td>204 (44.1)</td>
</tr>
<tr>
<td>Age (years), median (range)</td>
<td>32 (3-75)</td>
<td>28 (1-81)</td>
<td>42 (2-92)</td>
</tr>
<tr>
<td>G6PD&lt;sup&gt;d&lt;/sup&gt; deficiency, n (%)</td>
<td><em>e</em></td>
<td><em>e</em></td>
<td><em>e</em></td>
</tr>
</tbody>
</table>

<sup>a</sup>pMDA: primaquine mass drug administration.  
<sup>b</sup>MDA: mass drug administration.  
<sup>c</sup>CSS: cross-sectional blood survey.  
<sup>d</sup>G6PD: glucose-6-phosphate dehydrogenase.  
<sup>e</sup>Not applicable.

**Potential Effectiveness of pMDA**

To explore the potential effectiveness of pMDA, we compared the prevalences of *P. vivax* infections between the 2 study groups in each CSS period and the combined 2-period crossover. With the very low number of cases found for each CSS period, the prevalences were reported as the proportions of confirmed *P. vivax* infections per 1000 population. Among the CSS population in the pre-pMDA period of round 1 (July 2019), the proportions of *P. vivax* infections in the 2 groups were similar at around 14-15 per 1000 population. Among the CSS population in the 3-month post-pMDA period (November 2019), the proportions in the pMDA and control clusters reduced to 12.9 and 8.7 per 1000 population, respectively. Comparisons of the *P. vivax* proportions per 1000 population between the 2 groups (pMDA vs control) at each time period of round 1 showed no statistically significant differences (Table 3).

Among the CSS population in the pre-pMDA period of round 2 (July 2020), the proportions of *P. vivax* infections in the cross-over pMDA (clusters 4-7) and control (clusters 1-3) clusters were slightly different at 11.5 and 9.1 per 1000 population, respectively. Among the CSS population in the 3-month post-pMDA period (November 2020), the proportions in the pMDA and control clusters reduced to approximately 1.7 and 3.5 per 1000 population, respectively. Similarly, no statistically significant differences were found between the 2 groups in the round 2 study (Table 4). During the postintervention follow-up period, at 6 months after round 2 (March 2021), the proportions of *P. vivax* infections in the pMDA group (clusters 1-3) and control group (clusters 4-7) were 2.4 and 1.8 per 1000 population. Comparisons between the 2 groups also showed no significant differences (Table 4).

The inclusive effectiveness of pMDA for the 2-period crossover is shown in Table 5. In the pre-pMDA period, the inclusive numbers of participants across the 7 clusters (clusters 1-7), who
received the pMDA intervention and control intervention, were 1536 (round 1 + round 2: 925+611) and 1577 (878+699), respectively. The proportions of *P. vivax* infections per 1000 population in the pMDA and control groups before pMDA were slightly different at 13.0 and 12.0 per 1000 population, respectively, with no statistically significant difference. In the 3-month post-pMDA period, the inclusive numbers of participants across the 7 clusters, who underwent the pMDA intervention and the control intervention, were 1430 (round 1 + round 2: 851+579) and 1420 (848+572), respectively. The proportions of *P. vivax* infections per 1000 population in the pMDA and control groups at 3 months after pMDA were 8.4 and 5.6, respectively, with no statistically significant difference.

**Table 3.** Comparisons of the proportions of *P. vivax* infections for round 1 of the primaquine mass drug administration and control clusters during the 2-period cluster-crossover sequence.

<table>
<thead>
<tr>
<th>Variable</th>
<th>pMDA&lt;sup&gt;a&lt;/sup&gt; clusters</th>
<th>Control clusters</th>
<th>PR&lt;sup&gt;b&lt;/sup&gt; (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
<td>Total of 1-3</td>
</tr>
<tr>
<td>Pre-pMDA CSS&lt;sup&gt;d&lt;/sup&gt;, n</td>
<td>219</td>
<td>175</td>
<td>531</td>
<td>925</td>
</tr>
<tr>
<td>Pre-pMDA, n (infections/1000 population)</td>
<td>2 (9.1)</td>
<td>8 (45.7)</td>
<td>3 (5.7)</td>
<td>13 (14.1, 7.8-24.6)</td>
</tr>
<tr>
<td>3 months post-pMDA CSS, n</td>
<td>206</td>
<td>163</td>
<td>482</td>
<td>851</td>
</tr>
<tr>
<td>3 months post-pMDA, n (infections/1000 population)</td>
<td>1 (4.9)</td>
<td>7 (42.9)</td>
<td>3 (6.2)</td>
<td>11 (12.9, 6.8-23.7)</td>
</tr>
</tbody>
</table>

<sup>a</sup>pMDA: primaquine mass drug administration.
<sup>b</sup>PR: prevalence ratio of the pMDA vs control groups based on a zero-inflated negative binomial mixed model (adjusted for cluster, without time period and interaction).
<sup>c</sup>For the total, the value of infections/1000 population is provided along with the 95% CI.
<sup>d</sup>CSS: cross-sectional blood survey.
<sup>e</sup>Not applicable.
Table 4. Comparisons of the proportions of *P. vivax* infections for round 2 of the primaquine mass drug administration and control clusters during the 2-period cluster-crossover sequence.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control clusters</th>
<th>pMDA(^a) clusters</th>
<th>PR(^b) (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
<td>Total of 1-3</td>
</tr>
<tr>
<td>Pre-pMDA CSS(^d), n</td>
<td>226</td>
<td>189</td>
<td>463</td>
<td>878</td>
</tr>
<tr>
<td>Pre-pMDA, n (infections/1000 population)</td>
<td>1 (4.4)</td>
<td>3 (15.9)</td>
<td>4 (8.6)</td>
<td>8 (9.1, 4.2-18.6)</td>
</tr>
<tr>
<td>3 months post-pMDA CSS, n</td>
<td>226</td>
<td>188</td>
<td>434</td>
<td>848</td>
</tr>
<tr>
<td>Pre-pMDA, n (infections/1000 population)</td>
<td>0 (0.0)</td>
<td>1 (5.3)</td>
<td>2 (4.6)</td>
<td>3 (3.5, 0.9-11.2)</td>
</tr>
<tr>
<td>6 months post-pMDA CSS, n</td>
<td>219</td>
<td>191</td>
<td>423</td>
<td>833</td>
</tr>
<tr>
<td>Pre-pMDA, n (infections/1000 population)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>2 (4.7)</td>
<td>2 (2.4, 0.4-9.6)</td>
</tr>
</tbody>
</table>

\(^a\)pMDA: primaquine mass drug administration.
\(^b\)PR: prevalence ratio of the pMDA vs control groups based on a zero-inflated negative binomial mixed model (adjusted for cluster, without time period and interaction).
\(^c\)For the total, the value of infections/1000 population is provided along with the 95% CI.
\(^d\)CSS: cross-sectional blood survey.
\(^e\)Not applicable.

Table 5. Comparisons of the proportions of *P. vivax* infections before and after primaquine mass drug administration between the intervention and control groups combined over the 2-period cluster-crossover sequence.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intervention group (clusters 1-7)</th>
<th>Control group (clusters 1-7)</th>
<th>PR(^a) (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-pMDA(^b) CSS(^c), n</td>
<td>1536</td>
<td>1577</td>
<td><em>d</em></td>
<td>—</td>
</tr>
<tr>
<td>Pre-pMDA, n (infections/1000 population, 95% CI)</td>
<td>20 (13.0, 8.2-20.4)</td>
<td>19 (12.0, 7.5-19.1)</td>
<td>2.8 (0.3-28.4)</td>
<td>.37</td>
</tr>
<tr>
<td>3 months post-pMDA CSS, n</td>
<td>1430</td>
<td>1420</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Pre-pMDA, n (infections/1000 population, 95% CI)</td>
<td>12 (8.4, 4.6-15.1)</td>
<td>8 (5.6, 2.6-11.5)</td>
<td>2.1 (0.3-13.7)</td>
<td>.46</td>
</tr>
</tbody>
</table>

\(^a\)PR: prevalence ratio of the pMDA vs control groups based on a zero-inflated negative binomial mixed model (adjusted for cluster, with time period as a covariate and interaction with the pMDA intervention).
\(^b\)pMDA: primaquine mass drug administration.
\(^c\)CSS: cross-sectional blood survey.
\(^d\)Not applicable.

Safety and Adverse Drug Reactions of PQ During pMDA

Adverse events potentially due to PQ intake were recorded in approximately 5% of participants in each round (30/575, 5.2% in round 1 and 20/375, 5.3% in round 2) (Table 6). Although study participants with G6PD deficiency were excluded, about 1% of the participants who received PQ showed symptoms suggesting acute hemolysis. Six participants had hemoglobin levels of <8 g/dL, and another 4 participants had dark urine. Major chief complaints after taking PQ varied and included headache, weakness, muscle ache and pain, and dry throat. All
study participants with safety concerns were stopped from further PQ intake.

Table 6. Safety and adverse drug reactions during the 2-period cluster-crossover sequence.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Round 1 (n=575\textsuperscript{a}), n (%)</th>
<th>Round 2 (n=375\textsuperscript{a}), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population with adverse effects</td>
<td>30 (5.2)</td>
<td>20 (5.3)</td>
</tr>
<tr>
<td>Hemoglobin level &lt;8 g/dL</td>
<td>2 (0.3)</td>
<td>4 (1.1)</td>
</tr>
<tr>
<td>Dark urine</td>
<td>2 (0.3)</td>
<td>2 (0.5)</td>
</tr>
<tr>
<td>Other adverse events (chief complaints)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headache</td>
<td>8 (1.4)</td>
<td>5 (1.3)</td>
</tr>
<tr>
<td>Weakness</td>
<td>7 (1.2)</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Dry throat</td>
<td>5 (0.9)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Muscle ache and pain</td>
<td>1 (0.1)</td>
<td>4 (1.1)</td>
</tr>
<tr>
<td>Tachycardia</td>
<td>2 (0.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Chest tightness</td>
<td>1 (0.1)</td>
<td>2 (0.5)</td>
</tr>
<tr>
<td>Constipation</td>
<td>2 (0.3)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>0 (0.0)</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Vomiting</td>
<td>1 (0.1)</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Itching</td>
<td>0 (0.0)</td>
<td>2 (0.5)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Total population with at least 1-day drug administration.

Acceptability of the pMDA Program

Information was collected from community representatives, including 12 from 2 villages in Yala Province and 12 from 2 villages in Narathiwat Province. Key informants also included 11 VHWs in the selected villages (7 from Yala and 4 from Narathiwat) and 18 representatives of health care personnel who worked in the study areas and MOPH consultative experts. Among the 53 key informants, 34 were male and 19 were female, and their age ranged from 18 to 86 years. Some of the study participants were community and religious leaders. Most VHWs and their family members participated in the pMDA activities. In assessing the acceptance of the pMDA program and stakeholder engagement, 5 themes were explored, namely, perceptions, expectations, and engagement; factors influencing the decision to participate or not participate; factors influencing noncompletion of the 14-day regimen; perceived achievements, blockages, and solutions in pMDA; and challenges in pMDA implementation.

Perceptions, Expectations, and Engagement

Study participants informed us that they were willing to participate in pMDA because they recognized that malaria was a major problem around their residential areas. Many of them, particularly the older generation, had experienced malaria. All villagers who agreed to the CSS, but could not participate in pMDA due to the exclusion criteria, perceived the program’s benefits. On the other hand, all those who rejected pMDA stated that they or their families had never been infected with malaria, that they would seek treatment should they be infected, and that prevention was not necessary based on these factors. VHWs, who were key players in pMDA implementation, perceived the program’s benefits. Almost all of them indicated that it would not be a burden as they have to perform home visits and other activities in the villages as part of their routine job anyway. All local health care officers at the district and subdistrict levels noted that they expected the program to help reduce malaria cases in their areas. They said that they had no worries about program implementation, as they could easily collaborate with community leaders and VHWs.

There were many malaria cases in our province – in the top ranking in Thailand. After the pMDA project, the malaria cases were reduced to none in our village… Some VHWs from other villages asked us why the project only came to our village. [VHV]

Factors Influencing the Decision to Participate or Not Participate

Many participants said that they took a decision after going to one of several community engagement meetings arranged by the research team, community and regional leaders, and VHWs. Being well-informed about drug safety and the G6PD deficiency survey, most participants felt safe participating in the CSS and pMDA program.

Both of us took the drug for 14 days without any side effects. We decided to participate after attending the community meeting. We had no worries about taking the drug and the blood draw because we both used to get malaria. If we have to take the drug once a year, we still want to do so. We will ask our children to do so next time. [pMDA-compliant husband and wife]

In contrast, all VHWs and local health care staff indicated that most teenagers and small children did not participate in the program. With the drug problem (including illegal herbs, amphetamine, cocaine, and other drugs), particularly among...
teenagers, they feared that the CSS would reveal their drug status. Some mentioned that taking medication for 14 days was too long and that they might participate in a shorter regimen. Cultural beliefs also affected the idea of taking medication. One community leader stated that some local people believed that eating durian (a local fruit) with malaria drugs might affect people’s health.

Two of us in the family did not participate even though we used to get malaria – this is because we had to go work in the forest and had no time for DOT. ... My son aged 18 years old did not participate in the project because he had to go and study out of town... [pMDA nonparticipant]

Factors Influencing Noncompletion of the 14-Day Regimen
All pMDA participants who did not complete the 14-day regimen indicated that they had to stop taking PQ because they had adverse drug reactions. A few had to withdraw via the VHVs owing to safety issues (eg, dark urine and hemoglobin <8 g/dL).

There are 5 of us in the family – only 3 participated in drug administration, the other 2 did not because of their pregnancies. I took the drug for only 4 days and stopped because I had a headache – if not, I would take it for the whole 14-days. [pMDA noncompliant person]

Perceived Achievements, Blockages, and Solutions in pMDA
The brainstorming meeting reached a consensus that pMDA had an important impact on identifying asymptomatic infections because routine qPCR was not performed at local sites. The screening test for G6PD deficiency also contributed to significant implications, as there was little knowledge about G6PD deficiency prevalence in this region. Knowing the G6PD deficiency status instilled confidence in the local staff for PQ delivery during pMDA.

When asked about pMDA activities that required improvement, government officers recognized that community engagement still did not reach all target groups, and this was confirmed by villagers and VHVs. All stakeholders suggested that there should be greater coverage and more frequent community sensitization and engagement, which should be specific to the target groups (nonparticipant populations). Though willing to perform home visits, some VHVs noted that there should be some ways to handle intensive 14-day DOT. The brainstorming session also indicated a lack of resources (both workforce and financial support) that could obstruct the success of prevention and control measures.

At the beginning, we felt worried about our skills and the heavy workload of performing 14-day follow-ups. We also worried about getting villagers to understand the project, about taking the drug and the blood draw... The situation was better when we worked together with community leaders. [Local health care officer]

Challenges in pMDA Implementation
As noted during the brainstorming session, human resources would be a major challenge when moving forward and upsizing pMDA. A few health authorities mentioned that in order to have a successful program, it must be a top-down approach, which means the policy must be initiated by authoritative bodies at higher levels and delegated to local operational entities. On the other hand, another MOPH authority noted that the local level should initiate the idea and propose it to the upper level. There were concerns about not only the human workforce but also budget allocation.

Taking a 14-day regimen requires a G6PDd screening test. This test is rather expensive – but if you do not do it, the villagers may not want to take the drug. This is important and we need to communicate well with villagers to have them take the drug. Importantly, we also need the full endorsement of higher-level authorities, i.e., MOPH. If the MOPH had such a policy, the local offices would do it. [MOPH authority at the local level]

Discussion
Overview
The main goals of this study were to evaluate the effectiveness, safety, acceptability, and stakeholder engagement of pMDA to accelerate P. vivax elimination in Thailand and to provide the information needed by the Thai MOPH for evidence-based decision making. The study employed a mixed-methods approach. The quantitative part of the study employed a 2-period cluster-crossover randomized trial design assessing pMDA in addition to standard prevention and control measures. The qualitative part was performed by IDIs and brainstorming discussions.

Effectiveness and Safety of the pMDA Intervention
As a proof-of-concept study, the potential effectiveness of the pMDA intervention was assessed by comparing the proportions of P. vivax infections between clusters under pMDA and clusters under SPC in each period and after the 2-period crossover. The study results indicated no statistically significant differences between the 2 groups both before and after pMDA implementation. However, there were reducing trends in P. vivax prevalence after pMDA implementation in both the treatment and control groups. In a systematic review in the Cochrane database, it was found that studies on MDA in Cambodia, Laos, Myanmar, and Vietnam in very low- to low-endemicity settings showed varying degrees of reductions in P. vivax prevalence immediately following the intervention, but the effects were not sustained [31]. Similarly, WHO guidelines for malaria in 2023 noted that MDA for P. vivax conducted in 8 countries showed rapid reductions in transmission with immediate- to short-term benefits 1-3 months after the last round of MDA, but long-term benefits at 12-24 months were not apparent [32]. In this study, the comparable reduction trends in the pMDA and control groups may reflect that pMDA was not as effective as anticipated. This might be due to the Hawthorne effect (the alteration of behavior by the
study subjects due to their awareness of being observed), since both study participants and health care workers in the study areas were aware of the additional pMDA activities in their vicinity. It is also important to note that since the start of the study in 2019, drastically decreasing incidences have been observed across the country and not only in the study clusters. Moreover, people’s mobility was limited and malaria risk behaviors were less frequent during the COVID-19 pandemic, which coincidently started in 2019. Thus, it is difficult to reach definitive conclusions about the effectiveness of the pMDA intervention.

Consistent with the findings in other studies [15,20,33], pMDA was safe and well-tolerated. Testing for G6PD deficiency at the point of care before PQ administration is a precondition for safe administration [19]. Several previous studies noted that the tolerability of PQ has been good, with a low frequency of adverse events reported even with heterogeneous levels of G6PD deficiency [15,18]. This study, however, excluded individuals with inborn G6PD deficiency. As a safety monitoring measure, VHV’s performed the intensive 14-day DOT of PQ receivers residing in households for which they were responsible. Other studies also reported that pMDA under supervision with good monitoring mechanisms for adverse events in the population would result in less severe adverse events related to PQ [15,18,19]. This study and other pMDA studies involving dihydroartemisinin-piperquine and PQ showed similar common adverse events, including gastrointestinal disturbances (diarrhea, vomiting, abdominal pain, and nausea), dizziness, headache, and general body weakness [18,20,34]. Severe adverse events suggesting acute hemolysis occurred in 1.1% of PQ receivers, resulting in treatment cessation.

Acceptability and Stakeholder Engagement in pMDA Activities

In the WHO manual for antimalarial MDA implementation, high coverage and adherence of the target population (ie, >80%) must be ensured, preferably implementing centralized distribution at a fixed site and performing DOT. In this study, only about 60% of participants consented to the initial CSS, but over 90% of them were retained in the 3-year study. Among those, only 70% were eligible for MDA, with 90% taking PQ and 90% completing the 14-day regimen under DOT. The qualitative part of this study suggested that acceptance of pMDA among study participants was predominantly due to their trust in health care representatives (VHVs) who actively performed home visits to provide health care support. This study confirmed that DOT, although labor-intensive, could maintain full adherence and reassure study participants about the purposes and safety of the MDA activities, pMDA and DOT heightened the morale and relationships between villagers and health care personnel in the study areas. Similarly, another study on reactive drug administration for *P. vivax* elimination in Thailand suggested that good acceptance of the program was related to education and sensitization campaigns on the purpose and rationale of the intervention [35]. Moreover, the WHO guidelines 2023 noted that a systematic review of 18 studies reported that the most common barrier to the acceptability of MDA for *P. vivax* was fear of adverse events, and some studies mentioned that sensitization on the benefits of MDA helped reduce concerns about adverse effects [32]. Community engagement is critical to the success of MDA for *P. vivax* infection as it affects participation rates and full treatment compliance, while lack of engagement with local health care providers limits treatment adherence [32]. A systematic review of published, unpublished, and gray literature documenting past MDA experiences identified the importance of operational implementation and community engagement, with drug distribution and DOT being mainly performed by community volunteers and local health workers [36]. A review of previous studies noted in the WHO guidelines 2023 also reflects that the impact of MDA on *P. vivax* infection, whether positive or negative, is likely related to the level of acceptance of the intervention among the malaria program staff, and there have been no surveys of these key stakeholders regarding this issue [32]. As suggested in the literature, all aspects of community engagement in MDA must be tailored to local (social, cultural, and political) circumstances [37,38]. This study involved community members as part of the intervention team and considered their customs and opinions. Study participants and public health care providers at local and national levels willingly accepted and believed that the preventive and control activities in this study (regardless of being in the pMDA or control groups) had boosted awareness in the community and led to personal changes in malaria-preventive behaviors. The qualitative part of this study confirmed the key determinants of pMDA, and it was shown that the feasibility of maintaining or upscaling the MDA intervention was related to the existence of active and continuous activities for community engagement, community sensitization, and maintaining collaborations with those from the point-of-care level up to authorities at local and national levels.

Limitations of the Study

This study had several limitations. The number and size of clusters were small, and the study may not have observed heterogeneities among different groups. Limited mobility due to the COVID-19 pandemic coinciding with pMDA activities during the study follow-up period might have caused the malaria prevalence in all study clusters to decline drastically. Such circumstances complicate and hinder the interpretation of the results. Moreover, there were significant heterogeneities among the different clusters, making comparisons between the pMDA and control groups difficult. Regarding the study design, this study employed a cluster-crossover randomized trial design, which is particularly feasible for pragmatic clinical trial in health care systems [21,22]. The design is highly efficient and should be considered as it breaks up the total trial duration into a series of repeated measures. Thus, the required number of clusters could be substantially reduced while potentially providing generalizable, robust, and internally valid evidence for the evaluation of effects across settings or clusters [21,39,40]. Although the cluster-crossover design is robust when the number of clusters is small [26], this study involved only 7 clusters and the crossover sample size was around 1500 per group. Given the predictably large variation in malaria prevalence within each cluster, a larger sample size would have been needed. Thus, the results of this study may not have external validity and may not
be generalizable to other settings with different natures of target populations and degrees of transmission.

Conclusions

pMDA under DOT showed high adherence, safety, and tolerability, but it may not significantly impact *P. vivax* transmission, particularly in low-transmission areas. Although the impact of pMDA was inconclusive, the results were consistent. Malaria cases reduced in all clusters, regardless of whether they were in the pMDA group. The Hawthorne effect may reflect the trigger or accelerator of elimination by involving significant political, logistical, and financial commitments of coordinating bodies and collaborative efforts among various levels of stakeholders. We experienced success regarding effective health care workforces at point-of-care centers, effective collaborations in the community, and commitment from authorities at local and national levels. Such community engagement efforts boosted the acceptability of the malaria-elimination initiative.

Despite the safety and exemplary acceptance of the intervention, we cannot proclaim the effectiveness of pMDA. As we planned this study as a proof-of-concept study, the results were considered as a basis for decision making. A conditional recommendation in the WHO guidelines 2023 suggested the use of MDA for *P. vivax* infection when there is evidence of the acceptability, feasibility, impact (incidence or prevalence of malaria infection at the community level), and potential harm of MDA (including testing for G6PD deficiency) [32]. Since we did not observe the impact of pMDA on *P. vivax* transmission during the study period and the malaria incidence in Thailand increased in 2023 after the study period, we decided not to scale up the study with larger clusters and samples. Instead of implementing pMDA in the population within the intervention clusters, we are working on an on-going project with an alternative approach involving a targeted PQ treatment strategy that provides PQ via DOT only to the targeted population in households around each index case living in the intervention cluster. The effectiveness of such an alternative approach remains to be determined.

Acknowledgments

We sincerely appreciate the study participants, village health volunteers, and local health care personnel in the communities for their kind contributions to testing the study concept. We gratefully acknowledge all staff from the Thai Ministry of Public Health for their active roles in primaquine mass drug administration implementation and all staff who worked under the International Centers of Excellence for Malaria Research (ICEMR) project for performing their assigned activities and assisting in data collection. We would like to thank Paul Adams for English editing and proofreading. This work was supported by the National Institute for Allergy and Infectious Diseases, National Institutes of Health (NIH) (U19 AI089672).

Data Availability

All data and codes used in this analysis are presented in the manuscript.

Authors' Contributions

JK, SL, JS, WR, and WN designed the study. WR, WN, ST, PS, and PP implemented the intervention in the communities. WR, WN, AK, RP, PJ, PS, and MK collected the data, accessed and verified the data, and performed initial data analysis. JK, CN, and SL designed the statistical analysis. JK and MK led the qualitative data analysis. JK and WR wrote the first draft of the manuscript. JS, DMP, and LC provided important comments on the draft manuscript. All authors read and approved the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.

[PDF File (Adobe PDF File), 66 KB - publichealth_v10i1e51993_app1.pdf ]

References


Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CSS: cross-sectional blood survey
DOT: directly observed treatment
G6PD: glucose-6-phosphate dehydrogenase
GMS: Greater Mekong Subregion
ICEMR: International Center of Excellence for Malaria Research
IDI: in-depth interview
MDA: mass drug administration
MOPH: Ministry of Public Health
MPPT: mass primaquine preventive treatment
pMDA: primaquine mass drug administration
PQ: primaquine
SPC: standard of prevention and control
VHV: village health volunteer
WHO: World Health Organization
Effectiveness, Safety, and Acceptability of Primaquine Mass Drug Administration in Low-Endemicity Areas in Southern Thailand: Proof-of-Concept Study


Please cite as:


URL: https://publichealth.jmir.org/2024/1/e51993
doi: 10.2196/51993
PMID: 38922648

©Jaranit Kaewkungwal, Wanlapa Roobsoong, Saranath Lawpoolsri, Wang Nguitragool, Suwich Thammapalo, Pathomporn Prikchoo, Amnat Khamisiriwatchara, Rungrawee Pawarana, Pawinee Jarujareet, Daniel M Parker, Piyarat Sripoorote, Mondha Kengjanpanich, Chetta Ngamjarus, Jetsumon Sattabongkot, Liwang Cui. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 26.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Participatory Disease Surveillance for the Early Detection of Cholera-Like Diarrheal Disease Outbreaks in Rural Villages in Malawi: Prospective Cohort Study

Mariana Gasparoto Pereira Valerio, MSc, MD; Beverly Laher, BSc, MSc; John Phuka, PhD; Guilherme Lichand, PhD; Daniela Paolotti, PhD; Onicio Leal Neto, PhD

1Global Studies Institute, University of Geneva, Geneva, Switzerland
2Kamuzu University of Health Sciences, Lilongwe, Malawi
3Graduate School of Education, Stanford University, Stanford, CA, United States
4ISI Foundation, Turin, Italy
5Department of Epidemiology and Biostatistics, Mel & Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

Corresponding Author:
Onicio Leal Neto, PhD
Department of Epidemiology and Biostatistics
Mel & Enid Zuckerman College of Public Health
University of Arizona
1295 N Martin Ave
Tucson, AZ, 85724
United States
Phone: 1 5206263200
Email: onicio@gmail.com

Abstract

Background: Cholera-like diarrheal disease (CLDD) outbreaks are complex and influenced by environmental factors, socioeconomic conditions, and population dynamics, leading to limitations in traditional surveillance methods. In Malawi, cholera is considered an endemic disease. Its epidemiological profile is characterized by seasonal patterns, often coinciding with the rainy season when contamination of water sources is more likely. However, the outbreak that began in March 2022 has extended to the dry season, with deaths reported in all 29 districts. It is considered the worst outbreak in the past 10 years.

Objective: This study aims to evaluate the feasibility and outcomes of participatory surveillance (PS) using interactive voice response (IVR) technology for the early detection of CLDD outbreaks in Malawi.

Methods: This longitudinal cohort study followed 740 households in rural settings in Malawi for 24 weeks. The survey tool was designed to have 10 symptom questions collected every week. The proxies’ rationale was related to exanthematic, ictero-hemorrhagica for endemic diseases or events, diarrhea and respiratory/targeting acute diseases or events, and diarrhea and respiratory/targeting seasonal diseases or events. This work will focus only on the CLDD as a proxy for gastroenteritis and cholera. In this study, CLDD was defined as cases where reports indicated diarrhea combined with either fever or vomiting/nausea.

Results: During the study period, our data comprised 16,280 observations, with an average weekly participation rate of 35%. Maganga TA had the highest average of completed calls, at 144.83 (SD 10.587), while Ndindi TA had an average of 123.66 (SD 13.176) completed calls. Our findings demonstrate that this method might be effective in identifying CLDD with a notable and consistent signal captured over time ($R^2=0.681404$). Participation rates were slightly higher at the beginning of the study and decreased over time, thanks to the sensitization activities rolled out at the CBCCs level. In terms of the attack rates for CLDD, we observed similar rates between Maganga TA and Ndindi TA, at 16% and 15%, respectively.

Conclusions: PS has proven to be valuable for the early detection of epidemics. IVR technology is a promising approach for disease surveillance in rural villages in Africa, where access to health care and traditional disease surveillance methods may be limited. This study highlights the feasibility and potential of IVR technology for the timely and comprehensive reporting of disease incidence, symptoms, and behaviors in resource-limited settings.

(JMIR Public Health Surveill 2024;10:e49539) doi:10.2196/49539
KEYWORDS
participatory surveillance; digital epidemiology; interactive voice response; cholera; public health; outbreak; cohort study; public health surveillance; health technology; digital surveillance

Introduction
Cholera-like diarrheal diseases (CLDDs) are a group of diarrheal illnesses that share similar pathogenesis and symptoms with cholera but are caused by agents other than the *Vibrio cholerae*, such as *Escherichia coli* and some strains of *Campylobacter*, *Yersinia*, *Aeromonas*, and other pathogens [1]. Typically, CLDDs are responsible for milder and shorter episodes of diarrhea than cholera, but a comprehensive clinical assessment, including laboratory testing, is essential for accurate diagnosis, especially when occurring in cholera-endemic areas like Malawi [1-3]. The epidemiological profile of CLDD is characterized by seasonal patterns, often coinciding with the rainy season when contamination of water sources is more likely [4-6]. Therefore, the global burden of diarrheal disease lies heavily in regions that face disruptive environmental and socioeconomic conditions that affect the water supply and sanitation infrastructure. Most of these regions are in Sub-Saharan Africa, where cross-border cholera outbreaks, lack of testing capacity, and surveillance system limitations add an extra layer of complexity to the situation [7-14]. On December 5, 2022, cholera was declared a public health emergency, and it has been considered the country’s worst outbreak of the past 10 years [15]. The rise in cases expected during the wet season has extended to the dry season, with deaths reported in all 29 districts since March 2022 [15,16].

The situation grew even more alarming after the flooding caused by Tropical Storm Ana and Cyclone Gombe in January and March 2022, respectively [17]. These phenomena led to population displacement and water, sanitation, and hygiene (WASH) infrastructure disruption, exacerbated by longstanding issues such as the inadequate preparedness for disease control, the absence of a reliable and geographically distributed early detection system, and shortages in the oral cholera vaccine and other treatments due to multiple outbreaks within the African region [2,5,15,17-19]. As of April 4, 2023, the World Health Organization (WHO) has reported 160,756 cases of suspected cholera in the African region with a case fatality ratio (CFR) of 2.1%. Malawi has been significantly affected, accounting for 35% (n=56,763) of the total cases and 52% (n=1722) of the total fatalities, demonstrating a consistently high CFR above 3% [16]. Unfortunately, the true burden of cholera is likely higher than the reported figures due to underreporting and lack of access to health care in many areas [7,8].

In line with the urgent need for comprehensive action, it is crucial to focus on viable strategies that can complement traditional epidemiological surveillance and foster community engagement [20]. In this field, participatory surveillance (PS) has proven to be valuable for the early detection of epidemics. This system engages the community in a bidirectional manner, capturing strategic data from the community, processing the acquired knowledge, and providing nearly real-time information back [21-23]. This approach involves the active participation of community members, health workers, and other stakeholders in the surveillance process, including reporting and monitoring of disease incidence, symptoms, and behaviors. PS has demonstrated applicability in different scenarios and diseases, such as influenza, cholera, COVID-19, Zika virus, and others [21,24-31]. One of the main advantages of PS is that it overcomes several challenges associated with traditional disease surveillance methods. PS empowers communities to take ownership of their health and well-being, allowing for more comprehensive and timelier case reporting [32,33].

PS using interactive voice response (IVR) technology is a promising approach for disease surveillance in rural villages in Africa for collecting timely and comprehensive data on disease incidence, symptoms, and behaviors in resource-limited settings [34,35]. Additionally, IVR technology is cost-effective and has the potential to improve disease surveillance and control efforts in low-income settings [35-37]. Furthermore, it is an easy-to-use technology that requires minimal training and can reach a wide range of people, including those in remote and rural areas, where access to health care and traditional disease surveillance methods may be limited.

This study aims to evaluate the feasibility and outcomes of low-cost, high-frequency, and high-quality data collection through PS for the early detection of CLDD outbreaks in Malawi, implementing a system to identify its early signals.

Methods
Overview
This longitudinal prospective cohort study followed 740 rural households in Malawi for up to 24 weeks. It was rolled out in 2 Traditional Authorities (TAs) located in Salima District, the central region of Malawi (Figure 1). The choice of the TAs took into consideration the broader context of a larger study—the Child Development Study, an initiative to leverage high-frequency data collection and novel technologies for understanding child development in low-income settings. Therefore, logistical considerations and resource availability guided the choice of these TAs as suitable candidates for inclusion.

A total of 4051 households were enrolled in the Child Development Study. From this total, a sample randomization was performed, selecting 740 households from both TA areas, representing 3743 household members, including 2393 children. For the purpose of this study, it is imperative to highlight the definition of a household in Malawi as “one or more persons, related or unrelated, who make common provisions for food and who regularly take their food from the same pot and/or share the same grain house (nkhowwe) or pool their incomes together for the purpose of purchasing food” [38].

The recruitment strategy was implemented by the Kamuzu University of Health Sciences (KUHEN) local team, together with trained enumerators responsible for collecting the households’ phone numbers, conducting a baseline survey, and
gathering metadata. All these interactions, including the sensitization campaigns promoted throughout the study, took place at the community-based childcare centers (CBCCs), as the inclusion criteria for the main study required at least 1 child attending a CBCC.

Figure 1. Study area highlighting the locations of the Salima district, Maganga Traditional Authority (TA), Ndindi TA, and the community-based childcare centers (CBCCs).

Ethical Considerations
The College of Medicine Research and Ethics from Malawi (COMREC; reference no P_11_20_3202) and the Human Subjects Committee of the Faculty of Economics, Business Administration, and Information Technology at the University of Zurich (2018-046) approved this study. All participants signed the informed consent prior to enrollment and were compensated for their time spent reading and responding to messages in accordance with COMREC guidelines. Additionally, all phone costs were covered by the project.

Data Collection via IVR Technology
The survey instrument was designed with 10 questions about symptoms to be answered weekly by the same household member over a period of 24 weeks (Table 1). Additionally, the design allowed all types of phones to be included, with participants required to simply type, press, or dial the numbers indicated by the voice message to respond with “yes” or “no.” The IVR was live from July 18, 2022, until January 8, 2023. Calls were conducted in Chichewa, the primary language spoken in Malawi. The survey content was pretested with in-country KUHES representatives to reduce events of loss in translation and misinterpretation. The survey asked whether anyone in the household had experienced specific symptoms in the past 7 days, including fever, headache, joint pain, vomiting/nausea, jaundice, chills, body ache, bleeding, and diarrhea. Additionally, participants were asked if anyone in the household had undergone a malaria test during the same period, with response options including “took a test with a negative result,” “did not take a test,” and “took a test with a positive result.”

The proxies, and consequently, the questions used in the survey were selected based on the literature [39,40]. Special attention was given to designing questions capable of gathering comprehensive information while minimizing participant fatigue and enhancing adherence, as the same questions were repeated weekly. For this reason, the survey used proxies for malaria, rash, respiratory diseases, and CLDD. The rationale for these proxies was based on different disease characteristics: (1) exanthematic and ictero-hemorrhagic for endemic diseases or events and (2) diarrhea and respiratory for targeting both acute and seasonal conditions. This study specifically focused on reports related to CLDD, defined by the occurrence of diarrhea AND fever OR vomiting/nausea. Other proxies and their associated syndromes will be explored in future studies.

The household heads received calls from Monday to Friday, between 8 AM and 10 AM (a retry pattern of 2 calls with 2-hour intervals) and follow-up calls in the evening between 5:30 PM and 7:30 PM, running with this pattern throughout the week. An extract, transform, and load (ETL) process was performed weekly to feed data into a dashboard, allowing for near-real-time data visualization of the compiled results.
Sensitization Campaigns and Participant Engagement

Sensitization campaigns were conducted to raise awareness among participants before the calls officially started. Additionally, during the development of the study, we organized meetings at each CBCC used as a reference point for the study (Figure 1). Based on the eligibility criteria, all households involved in the study had at least 1 child who regularly attended the CBCCs. To ensure widespread coverage and engagement, we conducted sensitization campaigns at the CBCC locations. This approach made it easier for participants to access the information and encouraged their active involvement. The campaigns were held at various times and on different days to maximize participation and reach as many people as possible. They were facilitated by an Assistant Environmental Health Officer, a social welfare representative, and the Association of Early Childhood Development Personnel in Malawi. These campaigns aimed to raise awareness and inform participants about the process of calls and interactions. They addressed any concerns or questions about technical issues related to using mobile phones in this specific context, reinforced that the repetitive calls were intentional and not spam or mistakes, and discussed the importance of preventive health measures. Table 2 describes the framework of the sensitization campaigns.

Table 2. Sensitization campaigns framework for engaging participants during the prospective cohort.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>To introduce the IVR system to potential participants by giving them a live demonstration of how it works. This allows participants to become familiar with the technology and understand how to use it.</td>
<td>Demonstrative live call</td>
</tr>
<tr>
<td>To explain the purpose of the study and the importance of the data being collected. This helps participants understand how their participation will contribute to the success of the project and how it will benefit child development and health.</td>
<td>Presentation explaining in simple words how the data would be collected and the relevance of the study.</td>
</tr>
<tr>
<td>To ensure that participants understand the frequency and consistency of the calls they will receive. This reinforces the importance of their participation and helps to build trust and rapport with the participants.</td>
<td>Clarification and reinforcement that the calls would come weekly from the same number, addressing the same questions.</td>
</tr>
<tr>
<td>To motivate and empower participants to take an active role in the project. By understanding the importance of their participation in preventing and detecting outbreaks, the participants are more likely to engage fully with the IVR system and provide accurate information.</td>
<td>Presentation explaining the importance of preventive and early detection of outbreaks and empowering the participants as key collaborators to the success of such initiatives</td>
</tr>
<tr>
<td>To ensure that participants understand that they can still participate even if they miss a call. Because we provide clear instructions on this and reinforce that there is no cost involved, participants are more likely to remain engaged with the project.</td>
<td>Reinforcement about the possibility of calling back in case the participant missed a call.</td>
</tr>
<tr>
<td>To allow participants to choose a convenient time for the weekly calls. By accommodating the schedules of participants, the project is more likely to receive accurate and consistent data.</td>
<td>Enquirement for the preferable time spot to receive the calls.</td>
</tr>
<tr>
<td>To provide participants with an opportunity to ask questions and seek further clarifications about the project. This helps to build trust and rapport with the participants and ensures that they are fully informed about the purpose and methodology of the study.</td>
<td>Opening time for participant’s questions and further requests for clarifications.</td>
</tr>
</tbody>
</table>

Table 1. List of questions and answers collected. Adjustments and translations were made in Chichewa, keeping the meaning of each symptom.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you or anybody in your house experienced fever in the last 7 days?</td>
<td>If yes, dial 1; if, no dial 2.</td>
</tr>
<tr>
<td>Have you or anybody in your house experienced a headache in the last 7 days?</td>
<td>If yes, dial 1; if, no dial 2.</td>
</tr>
<tr>
<td>Have you or anybody in your house experienced joint pain in the last 7 days?</td>
<td>If yes, dial 1; if, no dial 2.</td>
</tr>
<tr>
<td>Have you or anybody in your house experienced vomiting/nausea in the last 7 days?</td>
<td>If yes, dial 1; if, no dial 2.</td>
</tr>
<tr>
<td>Have you or anybody in your house experienced jaundice in the last 7 days?</td>
<td>If yes, dial 1; if, no dial 2.</td>
</tr>
<tr>
<td>Have you or anybody in your house experienced body aches in the last 7 days?</td>
<td>If yes, dial 1; if, no dial 2.</td>
</tr>
<tr>
<td>Have you or anybody in your house experienced bleeding in the last 7 days?</td>
<td>If yes, dial 1; if, no dial 2.</td>
</tr>
<tr>
<td>Have you or anybody in your house experienced diarrhea in the last 7 days?</td>
<td>If yes, dial 1; if, no dial 2.</td>
</tr>
<tr>
<td>Did you or anybody in your house take a malaria test in the last 7 days?</td>
<td>If you took the test with a negative result, dial 1; If you did not take a test, dial 2; and if you took a test with a positive result, dial 3.</td>
</tr>
</tbody>
</table>
Statistical Analysis

We computed the household attack rate as an indication of how many households were at risk over the studied period. We assumed household reports as a proxy for the risk of presenting any of those symptoms related to CLDD. The attack rate is expressed as follows:

\[
\text{Attack Rate} = \frac{N_{hp}}{T_{hh}}
\]

where \(N_{hp}\) is the household positive reports (number of true reports on the given proxy), and \(T_{hh}\) is the total number of households in the study.

We used a polynomial generalized additive model (GAM) to analyze a time-series data set, given that the independent variable was time. The covariates were CLDD reports, diarrhea-like reports, fever-like reports, and vomiting/nausea-like reports, and they were chosen as nonparametric due to the uncertain relationship with the outcome.

The GAM is a flexible regression technique that allows for nonlinear relationships between the predictor variables and the response variable. To evaluate the performance of the polynomial GAM, we calculated \(R^2\), deviance explained, generalized cross-validation (GCV), and scale estimate. The \(R^2\) values ranged from 0 to 1, with higher values indicating a better fit of the model to the data. The following formula was used to estimate the coefficient of determination:

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} \text{res}_i^2}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]

where \(SS_{res}\) is the sum of squared residuals, and \(SS_{tot}\) is the total sum of squares.

Deviance explained measures the reduction in deviance, which is a measure of the difference between the observed data and the fitted values, achieved by the model. Deviance-explained values range from 0 to 1, with higher values indicating a better fit of the model to the data. The formula is described as follows:

\[
\text{Deviance Explained} = 1 - \frac{\text{Deviance}_{null}}{\text{Deviance}_{model}}
\]

where \(\text{Deviance}_{null}\) is the deviance of a null model with no predictors, and \(\text{Deviance}_{model}\) is the deviance of the model being evaluated.

The GCV score is the average of the squared differences between the predicted values and the actual values. To estimate the GCV score, we used the following formula:

\[
\text{GCV} = \frac{\sum_{i=1}^{n} \text{res}_i^2}{n - p - 1}
\]

where \(n\) is the sample size, \(p\) is the number of predictors in the model, and \(\text{res}\) are the residual differences between the observed response variable and the predicted response variable.

A kernel estimator was used to evaluate the spatial density of the events. This technique is a nonparametric method that can estimate the probability density function (PDF) of a random variable based on observed data. In the case of CLDD reports, the PDF represents the distribution of cases across a population, which is an important factor in understanding the spread of the disease and identifying high-risk areas. The formula is defined as:

\[
\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)
\]

where \(x\) is the point at which the estimate is being made, \(x_i\) represents the \(i\)th observation in the data set, \(n\) is the number of observations in the data set, \(h\) is the bandwidth parameter, and \(K\) is the kernel function.

In this study, to safeguard participant privacy and avoid disclosing sensitive information, we used the coordinates of CBCCs. Notably, CBCCs cater to multiple villages that are usually close to the facility. The households included in this study had at least 1 child attending the CBCC. Consequently, given the dynamic social landscape within these villages, the CBCC locations served as a suitable proxy for the frequently visited vicinities by households. These measures ensured participant confidentiality while allowing for accurate spatial analysis within the study’s scope.

Results

During the study period, our data comprised 16,280 observations, achieving a weekly participation rate of 35%. Maganga TA showed the highest average of completed calls, at 144.83 (SD 10.587), while Ndindi TA showed an average of 123.66 (SD 13.176) completed calls. The participation rates were slightly higher at the beginning of the study and declined over time, with no significant drop, possibly influenced by the sensitization efforts implemented at the CBCC level. Figure 2 displays the participation rates for each week throughout the study period, while Figure 3 presents these data disaggregated by TA area.

Regarding the attack rates for CLDD, we found similar rates between Maganga TA and Ndindi TA, at 16% and 15%, respectively (Table 3). While fever-like reports could be taken as more sensitive than diarrhea-like reports and vomiting/nausea-like reports, using the proxy for CLDD might have helped to envelop the sensitivity of the signal, decreasing the potential number of false positives Figure 4.

Considering the aggregated view, CLDD showed a consistent signal over time (\(R^2=0.681404\)) and was even stronger when isolating diarrhea-like symptoms and fever-like symptoms, as
shown in Table 3. When breaking down by TA area, Maganga (Table 4) demonstrates a better performance than Ndindi (Table 5). At the beginning of the cohort, the highest activity in the time-series distribution could be explained by the initial efforts and participants’ fresh recall of the project. After 5 weeks of reports, the signal began increasing again, which might be related to the occurrence of the cholera outbreaks that hit that region during those months. The first report of cholera case in the Salima District was found on September 14, 2022, 2 days after all signals showed a spike (Figure 5). Although official reports were scarce, a hyperlocal media website confirmed the first cholera case in the Salima District [41].

**Figure 2.** Participation profile rate according to the status of the interactive voice response (IVR) call aggregated.

**Figure 3.** Participation profile rate according to the status of the interactive voice response (IVR) call disaggregated by Traditional Authority (TA) areas.
Table 3. Attack rates at the household level based on the signals reported during the study.

<table>
<thead>
<tr>
<th>Location</th>
<th>Signal</th>
<th>Diarrhea-like reports</th>
<th>Fever-like reports</th>
<th>Vomiting/nausea-like reports</th>
<th>CLDD(^a) reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maganga</td>
<td>CLDD(^a)</td>
<td>0.1398</td>
<td>0.2043</td>
<td>0.1327</td>
<td>0.1699</td>
</tr>
<tr>
<td>Ndindi</td>
<td>CLDD(^a)</td>
<td>0.1203</td>
<td>0.2059</td>
<td>0.1224</td>
<td>0.1539</td>
</tr>
</tbody>
</table>

\(^a\)CLDD: cholera-like diarrheal disease.

Figure 4. Time-series visualization of the signals, including the trend line (polynomial generalized additive model, GAM) of cholera-like diarrheal disease (CLDD), diarrhea-like reports, fever-like reports, and vomiting/nausea-like reports for Maganga Traditional Authority (TA) and Ndindi TA. The y-axis shows the percentage of reports that answered positively for the symptoms or syndrome.

Table 4. Statistical summary for the time series of signals captured by IVR\(^a\) technology, considering the aggregated data for Salima District.

<table>
<thead>
<tr>
<th>Signal</th>
<th>(R^2)</th>
<th>Deviance explained</th>
<th>GCV(^b)</th>
<th>Scale estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLDD(^c)</td>
<td>0.681404</td>
<td>0.774625</td>
<td>0.0624334</td>
<td>0.0423252</td>
</tr>
<tr>
<td>Diarrhea-like reports</td>
<td>0.831161</td>
<td>0.885469</td>
<td>0.0422496</td>
<td>0.0274658</td>
</tr>
<tr>
<td>Fever-like reports</td>
<td>0.92951</td>
<td>0.954495</td>
<td>0.0420778</td>
<td>0.0260313</td>
</tr>
<tr>
<td>Vomiting/nausea-like reports</td>
<td>0.587607</td>
<td>0.692355</td>
<td>0.081265</td>
<td>0.0580976</td>
</tr>
</tbody>
</table>

\(^a\)IVR: interactive voice response.
\(^b\)GCV: generalized cross-validation.
\(^c\)CLDD: cholera-like diarrheal disease.
Table 5. Statistical summary of GAM\(^a\) for the time series of the signals captured, considering the disaggregated data for Maganga TA\(^b\) and Ndindi TA areas.

<table>
<thead>
<tr>
<th>TA area and signal</th>
<th>(R^2)</th>
<th>Deviance explained</th>
<th>GCV(^c)</th>
<th>Scale estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maganga</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLDD(^d)</td>
<td>0.596143</td>
<td>0.69602</td>
<td>0.020795</td>
<td>0.015</td>
</tr>
<tr>
<td>Diarrhea-like reports</td>
<td>0.711076</td>
<td>0.796532</td>
<td>0.0256</td>
<td>0.017277</td>
</tr>
<tr>
<td>Fever-like reports</td>
<td>0.899293</td>
<td>0.9338</td>
<td>0.016462</td>
<td>0.0103705</td>
</tr>
<tr>
<td>Vomiting/nausea-like reports</td>
<td>0.537684</td>
<td>0.632479</td>
<td>0.0247917</td>
<td>0.0188871</td>
</tr>
<tr>
<td>Ndindi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLDD</td>
<td>0.421066</td>
<td>0.576651</td>
<td>0.0414793</td>
<td>0.0290681</td>
</tr>
<tr>
<td>Diarrhea-like reports</td>
<td>0.711076</td>
<td>0.796532</td>
<td>0.0256</td>
<td>0.017277</td>
</tr>
<tr>
<td>Fever-like reports</td>
<td>0.785284</td>
<td>0.851777</td>
<td>0.0343511</td>
<td>0.0227253</td>
</tr>
<tr>
<td>Vomiting/nausea-like reports</td>
<td>0.408624</td>
<td>0.560045</td>
<td>0.047594</td>
<td>0.0304854</td>
</tr>
</tbody>
</table>

\(a\)GAM: generalized additive model.

\(b\)TA: Traditional Authority.

\(c\)GCV: generalized cross-validation.

\(d\)CLDD: cholera-like diarrheal disease.

Figure 5. Time-series visualization of the signals, including the trend line polynomial generalized additive model, (GAM) of cholera-like diarrheal disease (CLDD), diarrhea-like reports, fever-like reports, and vomiting/nausea-like reports aggregated by Salima District. It also includes the date of the first cholera case reported in the district. The y-axis shows the percentage of reports that answered positively for the symptoms or syndrome.

In the subsequent weeks, the peak of cholera cases in the Maganga TA area was detected, where an ascending pattern in the diarrhea-like reports could be observed (Figure 6; [42]). Three weeks before this peak, CLDD reports were at their highest level when this area was considered in isolation.

Regarding the spatial distribution of the CLDD reports, even assuming a homogeneous distribution of these events over the regions, there is a notable concentration near the lake vicinity (Figure 7). The Salima District Council locally issued a “Cautionary Statement on Cholera” on October 13, 2022, aiming to provide the populace with guidance. In this statement, they reported: “Salima District has registered a total of 54 cases as of 13 October 2022 with TA Maganga along the lake shore being the most affected, with 45 cases” [43]. This observation corroborates the proxies’ ability to identify trends indicative of a possible cholera outbreak in high-risk areas.
Discussion

Principal Findings

In this study, we implemented a PS system using IVR in 2 districts of Malawi and assessed its feasibility, acceptability, and effectiveness in detecting early signals of CLDD outbreaks. During the 24-week follow-up period, the system captured 16,280 observations, and the average weekly participation rate was 35%, indicating the strategy’s feasibility and acceptability according to the literature [24,44-46].

In summary, our findings demonstrate that this method might be effective in identifying CLDD with a notable and consistent
signal being captured over time ($R^2 = 0.681404$). The signal showed a significant increase coinciding with cholera outbreaks in the region. This pattern was notably observed with the first cholera case in the Salima District, detected shortly after a spike in our CLDD data (Figure 5). Despite limited official reporting, this outbreak was confirmed through hyperlocal media sources. Furthermore, a subsequent analysis highlighted a peak in cholera cases in the Maganga TA area, preceded by a rise in CLDD reports of diarrhea-like symptoms underscoring the potential of CLDD in early outbreak detection and response facilitation. The local media released a breaking news report on October 15, 2022, highlighting the “Cautionary Statement on Cholera” that the Salima District Council issued on October 13, 2022. Signed by the Director of Health and Social Services, the statement aimed to provide guidance to the population. It mentioned a total of 54 cholera cases in Salima, emphasizing that TA Maganga accounted for 45 cases. The statement also underscored that the regions most affected were those in close proximity to the lake [43].

Traditional surveillance systems can have limitations in detecting early signals of outbreaks, particularly in resource-limited settings [7,10,14]. Therefore, PS has been proposed as an additional approach that leverages the local knowledge and resources of communities to identify early outbreak signals.

The successful implementation of PS must consider the site's geolocation, epidemiological profile, seasonality, and available telecommunication infrastructure [47]. In this study, Maganga and Ndindi encompass a vast territorial area bordered by Malawi Lake, the ninth-largest freshwater lake on Earth [48,49]. The lake plays a vital role in the region's economy, demographic density, and definition of 2 different seasons—the dry and wet seasons [48]. Thus, during the wet season between December and April, the population expects floods followed by disruptions in the WASH infrastructure, and consequently, an increase in overall diarrhea cases [16,48].

Moreover, in 2019, there were more than 5 million mobile subscriptions, with a gradual increase in mobile penetration over the years, predominantly through prepaid subscriptions [50]. The growth in mobile penetration is evident, with reports indicating approximately 12.27 million mobile subscribers at the beginning of 2022 [51]. Regarding internet coverage, the International Telecommunication Union showed that only 10% of the Malawi population was using the internet until 2019 [49]. Although internet penetration is growing, reports from 2022 showed that almost 80% of the country’s population remained offline, with internet access being particularly low in areas away from cities [52]. Bearing that in mind, models of PS conducted through mobile apps, such as Flu Near You, AlyaData, or the Guardians of Health platform, would likely not achieve significant engagement in the most remote areas of Malawi. Therefore, the choice of a self-report survey via IVR may be the most appropriate strategy for this setting [33,40,44,53,54].

PS strategies have been applied worldwide with positive results in different settings and events, successfully tracing pandemics such as Zika, H1N1, and COVID-19. PS has the potential to anticipate outbreaks and provide a quicker overview, guiding authorities to the hotspots of specific diseases [25,53-55]. Despite limitations in the field of technology and communication, PS strategies can rely on affordable systems that have been successfully implemented in underresourced areas, such as Malawi [26]. By leveraging the IVR design, it is possible to establish a less expensive, flexible, scalable, and reliable system that captures data voluntarily and provides information that is not possible to capture using traditional surveillance methods. This approach can empower communities to take an active role in anticipating disease outbreaks, even in settings where internet coverage is limited [34,56,57].

**Study Limitations**

Our study has several limitations. In terms of study design, population displacement and changes in the household configuration affected the consistency of the individuals being reported using the same household phone number. Constant phone number changes resulted in us losing contact with an entire household. Pitfalls in the adherence rate over the weeks could occur due to recipients perceiving the repeated calls with identical questions as spam, potentially leading them to avoid answering that specific phone number.

Selection bias is another significant limitation. Not all individuals may have access to a mobile phone, and those who have may differ in socioeconomic status, education level, or other factors that may affect their willingness to participate in the study. This could result in an unrepresentative sample and affect the generalizability of the results.

Another limitation is that the quality and accuracy of the data collected through IVR may be affected by factors such as poor network coverage, low battery, or other technical issues that may prevent individuals from completing the survey or lead to missing data. Additionally, individuals may not always report their symptoms accurately or truthfully, which could affect the validity of the data collected through IVR. Validating PS data with traditional sources in low-income settings is also challenging due to the scarcity of disease surveillance data. In many low-income settings, disease surveillance systems may be underresourced or underdeveloped, resulting in delays in the collection, analysis, and reporting of disease data. Consequently, it can be challenging to validate the accuracy of the data collected through PS with traditional sources. This sum of constraints can lead to a lack of confidence in the accuracy of the PS data and limit its usefulness for public health decision-making. Moreover, traditional sources of disease data, such as hospital records or laboratory test results, may not always be available or accessible in low-income settings, further complicating the validation process. Therefore, while PS has the potential to provide valuable data on disease trends and outbreaks in low-income settings, its usefulness and advantages may be limited by the lack of timely and reliable validation data from traditional sources.

**Conclusion**

IVR systems have the potential to facilitate PS in low-income and low-resource countries by enabling efficient and cost-effective data collection. In this study, weekly automated phone calls were made to a representative sample of the
population over a 6-month period, during which participants answered a consistent set of 10 questions related to cholera and malaria and associated risk factors. In low-income countries and settings with limited technological resources, PS using this method can be particularly useful in preventing CLDD outbreaks for several reasons. First, IVR allows for the timely detection of cases by rapidly identifying suspected CLDD cases within communities, providing public health officials with real-time data to respond promptly and contain potential outbreaks. Second, by using a standardized set of questions, the IVR system ensures that data collected across different participants and time points are consistent and comparable, thus improving the surveillance system’s reliability. Additionally, the use of phone-based surveys enables data collection from geographically dispersed and hard-to-reach populations, overcoming logistical barriers typically encountered in low-resource settings. Finally, by longitudinally monitoring the same set of participants over 6 months, the IVR-based PS system can capture temporal trends and identify emerging risk factors, enabling targeted and context-specific interventions to prevent and control disease outbreaks.

Acknowledgments
Generative artificial intelligence (AI) was used solely for proofreading purposes in this paper, focusing on grammar and phrase concordance since some of the authors are nonnative English speakers. It was not used to generate any content or intellectual material.

Data Availability
Data related to this study are available upon request. Interested parties can contact the corresponding author to access the data sets, which are shared in accordance with applicable privacy and ethical guidelines.

Authors’ Contributions
MV oversaw the investigation, utilized resources, and produced the initial draft of the manuscript. BL was responsible for data curation and reviewed the final draft. JP and GL handled project administration. GL also secured funding for the study and was responsible for writing review. DP contributed through critical review and editing of the final text. OLN was involved in conceptualization, investigation, data curation, supervision, methodology, and formal analysis, and also oversaw the writing, reviewing, and editing processes of the manuscript.

Conflicts of Interest
None declared.

References


42. Nkhata Bay leads with 1,002 Cholera cases. Kulinji. URL: https://kulinji.com/article/nkha-bay-leads-1002-cholera-cases [accessed 2023-05-05]


47. Smolinski MS, Crawley AW, Olsen JM, Jayaraman T, Libel M. Participatory disease surveillance for a mass gathering at FIFA World Cup 2014, Brazil. JMIR Public Health Surveill 2017 May 04;3(2):e26 [FREE Full text] [doi: 10.2196/publichealth.7313] [Medline: 28473308]


Abbreviations

- CBCC: community-based childcare center
- CFR: case-fatality ratio
- CLDD: cholera-like diarrheal disease
- COMREC: College of Medicine Research and Ethics from Malawi
- ETL: extract, transform, and load
- GAM: generalized additive model
- GCV: generalized cross-validation
- IVR: interactive voice response
- KUHES: Kamuzu University of Health Sciences
- PDF: probability density function
- PS: participatory surveillance
- TA: Traditional Authority
- WASH: water, sanitation, and hygiene
- WHO: World Health Organization

Edited by A Mavragani; submitted 01.06.23; peer-reviewed by C Kjelsø, M Gasmi, M Tarnas; comments to author 24.01.24; revised version received 16.02.24; accepted 16.05.24; published 16.07.24.

Please cite as:
Valerio MGP, Laher B, Phuka J, Lichand G, Paolotti D, Leal Neto O. Participatory Disease Surveillance for the Early Detection of Cholera-Like Diarrheal Disease Outbreaks in Rural Villages in Malawi: Prospective Cohort Study. JMIR Public Health Surveill 2024;10:e49539
URL: https://publichealth.jmir.org/2024/1/e49539
doi:10.2196/49539
PMID:39012690

©Mariana Gasparoto Pereira Valerio, Beverly Laher, John Phuka, Guilherme Lichand, Daniela Paolotti, Onicio Leal Neto. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 16.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Estimation of the Population Size of Street- and Venue-Based Female Sex Workers and Sexually Exploited Minors in Rwanda in 2022: 3-Source Capture-Recapture

Elysee Tuyishime¹,², MSc; Eric Remera³, PhD; Catherine Kayitès³, BSc; Samuel Malamba², PhD; Beata Sangwayire², MPH; Ignace Habimana Kabano¹, PhD; Horacio Ruisenor-Escudero⁴, PhD; Tom Oluoch², PhD; Angela Unna Chukwu¹,⁵, Prof Dr

¹African Center of Excellence in Data Science, University of Rwanda, Kigali, Rwanda
²Division of Global HIV and TB, Global Health Center, US Centers for Disease Control and Prevention, Kigali, Rwanda
³Institute of HIV Disease Prevention and Control, Rwanda Biomedical Centre, Rwanda Ministry of Health, Kigali, Rwanda
⁴Key Population Surveillance Team, Epidemiology and Surveillance Branch, Division of Global HIV/TB, Global Health Center, US Centers for Disease Control and Prevention, Atlanta, GA, United States
⁵Department of Statistics, University of Ibadan, Ibadan, Nigeria

Corresponding Author:
Elysee Tuyishime, MSc
Division of Global HIV and TB
Global Health Center
US Centers for Disease Control and Prevention
337H+F58
30 KG 7 Avenue
Kigali, P.O. Box 28
Rwanda
Phone: 250 788381437
Email: obx5@cdc.gov

Abstract

Background: HIV surveillance among key populations is a priority in all epidemic settings. Female sex workers (FSWs) globally as well as in Rwanda are disproportionately affected by the HIV epidemic; hence, the Rwanda HIV and AIDS National Strategic Plan (2018-2024) has adopted regular surveillance of population size estimation (PSE) of FSWs every 2-3 years.

Objective: We aimed at estimating, for the fourth time, the population size of street- and venue-based FSWs and sexually exploited minors aged ≥15 years in Rwanda.

Methods: In August 2022, the 3-source capture-recapture method was used to estimate the population size of FSWs and sexually exploited minors in Rwanda. The field work took 3 weeks to complete, with each capture occasion lasting for a week. The sample size for each capture was calculated using shinyrecap with inputs drawn from previously conducted estimation exercises. In each capture round, a stratified multistage sampling process was used, with administrative provinces as strata and FSW hotspots as the primary sampling unit. Different unique objects were distributed to FSWs in each capture round; acceptance of the unique object was marked as successful capture. Sampled FSWs for the subsequent capture occasions were asked if they had received the previously distributed unique object in order to determine recaptures. Statistical analysis was performed in R (version 4.0.5), and Bayesian Model Averaging was performed to produce the final PSE with a 95% credibility set (CS).

Results: We sampled 1766, 1848, and 1865 FSWs and sexually exploited minors in each capture round. There were 169 recaptures strictly between captures 1 and 2, 210 recaptures exclusively between captures 2 and 3, and 65 recaptures between captures 1 and 3 only. In all 3 captures, 61 FSWs were captured. The median PSE of street- and venue-based FSWs and sexually exploited minors in Rwanda was 37,647 (95% CI 31,873-43,354), corresponding to 1.1% (95% CI 0.9%-1.3%) of the total adult females in the general population. Relative to the adult females in the general population, the western and northern provinces ranked first and second with a higher concentration of FSWs, respectively. The cities of Kigali and eastern province ranked third and fourth, respectively. The southern province was identified as having a low concentration of FSWs.
Conclusions: We provide, for the first time, both the national and provincial level population size estimate of street- and venue-based FSWs in Rwanda. Compared with the previous 2 rounds of FSW PSEs at the national level, we observed differences in the street- and venue-based FSW population size in Rwanda. Our study might not have considered FSWs who do not want anyone to know they are FSWs due to several reasons, leading to a possible underestimation of the true PSE.

(JMIR Public Health Surveill 2024;10:e50743) doi:10.2196/50743

KEYWORDS
population size; female sex workers; capture-recapture; 3-source; Rwanda; HIV; surveillance; population; epidemiology; prevention; AIDS; sexually transmitted disease; STD; minor; young adult; sexually exploited minor; children

Introduction
Increased risk for HIV transmission is highly associated with social marginalization, and those individuals who are socially marginalized may not identify themselves as such when accessing services. This makes it difficult to track them in HIV program registers and impedes efforts to plan and have informed resource allocations for high impact. Those individuals are considered to be key populations for HIV and include female sex workers (FSWs), men who have sex with men, transgender women, and people who inject drugs—all of whom are at increased risk for HIV infection compared with the population at large [1,2]. In 2021, key populations and their sexual partners accounted for 70% of HIV infections globally, with 51% in sub-Saharan Africa, and the risk of HIV acquisition among FSWs is 30 times higher than that in adult females globally [2-4].

Rwanda, an East African country, is surrounded by 4 neighboring countries: Tanzania, Uganda, Democratic Republic of the Congo, and Burundi. Rwanda is divided into 5 administrative regions—4 provinces and the City of Kigali—and 30 districts as another subnational unit level. Rwanda experiences a mixed HIV epidemic, generalized in the adult population, with an adult (age 15-49 years) HIV prevalence stabilized at around 2.6% and with aspects of a concentrated epidemic among specific key populations at higher risk of HIV infection, with 45% among FSWs [5]. In Rwanda, FSWs are considered among the key populations for HIV prevention and care in the HIV and AIDS National Strategic Plan (2018-2024) [6].

Since 2010, the Rwanda Biomedical Center (RBC) has conducted 3 rounds of street- and venue-based FSW population size estimations (PSEs) nationally. The first was implemented in 2010 [7], where 3 methods were used, namely, 2-source capture-recapture (CRC), enumeration, and the multiplier method [8]. Using CRC and enumeration methods yielded the FSW PSE slightly more than 3000 [7]. Two years later, in 2012, through a household survey, RBC estimated the size of FSWs to range from 25,000 to 45,000 in Rwanda [9]. Lastly, in 2018, RBC conducted a national exercise of FSW size estimation by using the 3-source CRC (3S-CRC) method, where findings showed FSW PSE to range from 8328 to 22,806 with a median of 13,716 [10].

The recent 5th Rwanda Population and Housing Census in 2022 reported an average annual growth rate of 2.3%, where the current resident population increased beyond 13 million from 10.5 million in 2012 [11]. This rapid increase in the general population reflects the need for regular update of key population size estimates to inform HIV programming and planning in Rwanda. As affirmed by the Joint United Nations Program on HIV/AIDS or World Health Organization in the publication guidelines for second generation HIV surveillance [12], HIV surveillance among key populations is a priority in all epidemic settings. Identifying the key population groups, their locations, and their sizes helps in understanding and prioritizing the current needs for HIV prevention, diagnosis, treatment, and care services. It also helps in projecting the future needs for those services. Rwanda’s HIV and AIDS National Strategic Plan (2018-2024) has adopted routine surveillance of key populations, including regularly conducting PSEs of FSWs every 2-3 years [13]. This study aims at providing for the fourth time the population size estimate of street- and venue-based FSWs and sexually exploited minors aged 15 years and older operating in Rwanda, where the 3S-CRC method was used for the second time.

Methods
Study Population
The study population consists of biologically born females (girls or women), aged 15 years and older, who self-reported having any type of sex with men in exchange for goods, money, or services in the last 3 months and practicing sex work at street- and venue-based hotspots. Those fulfilling the above criteria and who were younger than 18 years are herein referred to as sexually exploited minors.

Study Design and Setting
This was a cross-sectional national FSW and sexually exploited minor PSE by using the 3S-CRC method [14]. The method involved visiting hotspots, where FSWs are known to congregate, on 3 separate occasions and sampling FSWs who were found at the hotspots on each occasion, calculating the degree to which FSW samples overlapped across 3 consecutive occasions. In this framework, an encountered FSW at the visited hotspot was referred to as captured, and each encounter occasion was referred to as a capture round in the CRC method context. A resampled FSW at a subsequent capture round was referred to as recaptured, and the intuition was that the degree to which FSW samples overlap across the 3 consecutive capture rounds was inversely proportional to the population size.

The objects used to tag FSWs who were presented at the hotspots were small, inexpensive, and branded with specific messages so that they would have a memorable design, and
these would only be available from the study staff who distributed them. During the first capture, a small bag branded with the “imigongo” traditional art form was offered; for the second capture, a purse branded with a flower and the key message “Rinda ubuzima” (protect your life) was offered; and during the third capture, a hair comb branded with a tree picture as a key message was offered.

**Sampling and Sample Size**

**Sampling Design**

A stratified multistage sampling design was used, with administrative provinces considered as strata and FSW hotspots as primary sampling units. Information from FSW’s hotspot mapping exercise was used as the sampling frame for this FSW PSE in 2022. Prior to this survey, the RBC conducted an FSW hotspot mapping exercise across the country from March to May 2022 to collect some key information that would inform future studies involving FSWs. Hotspot mapping consisted of teams going to the field to identify active venues and streets where FSWs congregate to find sexual clients. The FSW hotspot mapping exercise was facilitated by key informants identified by implementing partners who provide health services to FSWs to guide mapping teams. The mapping exercise identified 668 hotspots (street- and venue-based) countrywide and collected some beneficial data, including hotspot name, hotspot size, pick days, pick hours, and corresponding geocoordinates, to guide the sampling process.

The principal sampling processes were as follows: from the national list of FSW hotspots resulting from the hotspot mapping exercise, FSW hotspots were stratified by administrative provinces and the City of Kigali, and then a specific number of hotspots was selected using probability proportional to the size of the FSW populations. Hotspot sampling was performed using probability proportional to size for generating samples. In probability proportional to size sampling, the probability that a hotspot was sampled was proportional to the estimated size of FSWs observed at that hotspot during the hotspot mapping exercise. In practice, this means that hotspots with many FSWs are more likely to be sampled than hotspots with fewer FSWs. To enhance the geographical representativeness of the sample, hotspots were listed by the corresponding administrative provinces, and provinces were considered strata. To execute hotspot sampling, we listed all the hotspots in the order of the number of FSWs observed during the mapping exercise within a strata (to reflect the relative sizes of the FSW populations).

Then, we calculated the cumulative number of FSWs for each hotspot listed, determined the sampling interval, picked a random starting point, and finally selected a hotspot based off the random starting point, sampling interval, and cumulative FSW population size. This process was repeated in each capture round to minimize list dependency between capture occasions, and this resulted in the selection of 62 hotspots countrywide in each capture round.

**Sample Size Calculation**

The expected sample size for each capture round and statistical power were calculated using MS-CRC Power Analysis of the shinyrecap application [15]. Using the previously estimated size of FSWs in Rwanda of 23,495 [10], we set the application to simulate 500 CRC studies and report the amount of variability in the estimates based on the posited population and the sample size in each capture event to 2000 at an α level of .05. We found that there is a 95% chance that the CRC study’s population size estimate will be within 7.6% of the true value, corresponding to 1780 absolute accuracy. Considering the 11% nonresponse rate from the previous study, 2000 was found to be an appropriate sample size for each capture round (Multimedia Appendix 1). The number of objects distributed to each hotspot was proportional to the total number of FSWs estimated at the hotspot according to the 2022 mapping data.

To select the number of FSWs to be offered unique objects within a selected FSW hotspot, we used a systematic sampling approach for the distribution of unique objects. The unique object distribution process started with the FSW key informant conducting visual head counts of FSWs present at the hotspot and then estimating the distribution interval by dividing the head counts by the assigned hotspot unique objects. If the result of the division was 1, every FSW present at the hotspot should have received the unique object; otherwise, a random start would be randomly selected within the distribution interval following the physical standing position of FSWs in the hotspot. Table 1 shows the provincial distribution of the sampled 62 hotspots and the 2000 unique objects assigned in each capture round.

**Table 1.** Provincial level sample size replicated at each of the 3 capture rounds in Rwanda in 2022.

<table>
<thead>
<tr>
<th>Province</th>
<th>Information from hotspot mapping (n)</th>
<th>3-source capture-recapture sampling (n)</th>
<th>Average number of FSWs to be sampled per hotspot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hotspots per province (n=697)</td>
<td>Estimated total FSWs&lt;sup&gt;a&lt;/sup&gt; at hotspots during mapping exercise (n=22,471)</td>
<td>FSWs to be sampled (n=2000)</td>
</tr>
<tr>
<td>City of Kigali</td>
<td>100</td>
<td>3883</td>
<td>346</td>
</tr>
<tr>
<td>East</td>
<td>237</td>
<td>5825</td>
<td>518</td>
</tr>
<tr>
<td>North</td>
<td>74</td>
<td>3095</td>
<td>275</td>
</tr>
<tr>
<td>South</td>
<td>61</td>
<td>2858</td>
<td>255</td>
</tr>
<tr>
<td>West</td>
<td>225</td>
<td>6810</td>
<td>606</td>
</tr>
</tbody>
</table>

<sup>a</sup> FSW: female sex worker.
Data Collection

Data collection was conducted in sampled hotspots and lasted for a 3-week period from August 1 to August 21, 2022. All data collectors were trained about questionnaire administration using tablets as well as good clinical practice [16]. The 3 captures followed each other consecutively, each lasting for 1 week. Two data collectors from the 32 trained data collectors were randomly assigned to 28 districts (2 districts did not meet the minimum requirement to be included in the survey) of Rwanda in a team of 2 and shuffled on every capture round. Once they arrived in the selected hotspots, the data collection team was assigned a local FSW guide (key informant) who helped in object distribution among FSWs who were congregated in the selected hotspots. The hotspot visiting time depended on the selected pick days and hours for each venue or street, and the objects were distributed systematically by determining the interval according to the number of FSWs present at the hotspots and the number of assigned unique objects. The unique objects were distributed by FSW key informants under the supervision of data collectors, and the latter only recorded responses of whether the approached FSW accepted or refused the unique object. Those who accepted the unique object were marked as successful capture. For the successful captures, FSWs were asked if they had received a previously distributed unique object, and this information was also recorded on the tablet. If a successfully captured FSW claims to have received the previously distributed unique object, then she was asked to present it to the key informant in case she had it with her. Otherwise, a laminated card with different unique objects, including the correct distributed unique objects, was shown to her to point out the object she claimed to have received, and the data collector recorded if she successfully pointed out the correct unique object. As different unique objects were distributed in each capture round, depending on the type of object presented physically or pointed out on the laminated card, the data collector would mark the corresponding capture round from which that object was distributed on the tablet. Furthermore, a visual estimation of the age group for the approached FSWs was recorded. Data were collected at individual level encounters, that is, for each approached FSW, a record would be opened, filled, and saved in the tablet before moving forward to the next FSW. In summary, 4 major assumptions must be met for the CRC to give reliable population estimates; these were considered during data collection. These assumptions include that individual captures should be independent, the population should be closed during the data collection period, each target population member’s capture history should be correct, and the chance of getting captured should be homogeneous [17]. To minimize dependencies between captures, we repeated the sampling process of FSW hotspots in each capture round. To reduce recall bias and to ensure that the closed population assumption was met for all 3 captures, we maintained a 1-week period between consecutive capture rounds. Within sampled FSW hotspots, FSWs were sampled systematically to receive the distributed unique objects to ensure that the probability of being sampled was homogeneous. For details, please refer to the sampling and sample size section above. Finally, data collection was monitored in real time to ensure data quality and to ensure that each individual capture history was correct.

Data Management

Data were collected electronically using a tablet questionnaire programmed in Open Data Kit and transferred to the central server, with the process monitored by a qualified data manager. Electronic data files, computers, and other storage devices that contained data were password-protected, and electronic survey data files had encryption protection. Data files were transferred to the central encrypted server immediately after individual encounters. If the internet connection was not strong enough to upload data to the server, the records would be backed up on the tablet and sent later once the team reached the area with strong internet connectivity.

Data collectors were trained on the process of data collection and the use of tablets to ensure data quality. All data were anonymized, and no personal identifying information was collected. Participant-level data were line-listed and uniquely identified by a tablet-generated unique study ID. Data extracted from the central server were transferred into Microsoft Excel, maintained on a password-protected computer, and backed up on an external hard drive to ensure the security of the data, and kept in a locked, secure location. On a daily basis, the data manager who had access to the secure data would download the received data in Excel and conduct a quality check to ensure that high-quality data standards were met, including data logical flow and skip patterns. Whenever an error was identified, the data manager would immediately reach out to the concerned data collector for clarifications and rectify the error.

Statistical Analysis

Data cleaning was conducted using STATA 17 and RStudio (shinyrecap) for 3S-CRC for data analysis [18]. In preparation for the analysis, participant-level data were exported into RStudio (R package: shinyrecap) for Windows, and cleaning was performed based on preset exclusion criteria (self-reporting to be an FSW and accepting the offered unique object) and data logical flow following skip patterns. The data set was subset by province to have provincial-level FSW population size estimates. Aggregated data sets detailing counts of each CRC combination were produced for each subset. Table 2 shows how data were aggregated by overall and provincial $2^k - 1$ contingency tables for analysis preparation, where k stands for the number of capture occasions and $n_i$ represents aggregated counts, where i stands for a specific capture occasion.
Frequentist log-linear models [19], Bayesian nonparametric latent class [20], and Bayesian Model Averaging [21], which are flexible and able to accommodate various forms of heterogeneity in capture probabilities, were used to produce the final PSE with credibility sets from aggregated data sets. The median population size with 95% credibility sets and confidence intervals for 3S-CRC data were produced overall and by province. The selection of the best model to report among the 3 mentioned ones was based on whether the data presented list dependency (capture events independently drawn from the population) of captures or capture heterogeneity (individuals in the population have the same probability of being captured).

**Ethics Approval**

The survey received ethics approval from the Rwanda National Ethics Committee (IRB00001497). It was also reviewed in accordance with the US Centers for Disease Control and Prevention human research protection procedures and was determined to be not research. A waiver for consent was obtained, and no compensation was offered, as data collectors did not interact with the participants. The survey protected the anonymity of participants to avoid any stigmatization, and no personally identifiable information was collected. A referral form was available for sexually exploited minors, which included referrals to health and legal services.

**Results**

Of the 1778 FSWs approached during capture 1, 1768 (99.4%) were newly captured (ie, they were not captured elsewhere within the same week). Among those newly captured, unique object acceptance was high at 99.9% (1766/1778). For 1870 FSWs approached during capture 2, 1851 (98.9%) were newly captured within the second week of capture. Among those newly captured in capture 2, unique object acceptance was high at 99.8% (1848/1851). During capture 3, 1910 FSWs were approached, and 1867 (97.7%) were newly captured. The main reasons for unique object refusal documented were not being willing to receive the object and being willing to receive money instead of a unique object. Table 3 presents the results by capture round.

The majority of the FSWs sampled were presumed to be 25 years old, while only few sexually exploited minors aged 15-17 years were captured across all 3 capture rounds. Table 4 describes the sampled FSWs in each capture round by age and province.

### Table 2. Three-source capture-recapture aggregated data set of Rwanda in 2022.

<table>
<thead>
<tr>
<th>Capture 1</th>
<th>Capture 2</th>
<th>Capture 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>n₁</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>n₂</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>n₃</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>n₁₂</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>n₁₃</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>n₂₃</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>n₁₂₃</td>
</tr>
</tbody>
</table>

Approached female sex workers (n) 1778

Already in current capture, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Capture 1</th>
<th>Capture 2</th>
<th>Capture 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>10 (0.6)</td>
<td>19 (1.1)</td>
<td>43 (2.3)</td>
</tr>
<tr>
<td>No</td>
<td>1768 (99.4)</td>
<td>1851 (98.9)</td>
<td>1867 (97.7)</td>
</tr>
</tbody>
</table>

Unique object acceptance, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Capture 1</th>
<th>Capture 2</th>
<th>Capture 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accepted</td>
<td>1766 (99.9)</td>
<td>1848 (99.8)</td>
<td>1865 (99.9)</td>
</tr>
<tr>
<td>Refused</td>
<td>2 (0.1)</td>
<td>3 (0.2)</td>
<td>2 (0.1)</td>
</tr>
</tbody>
</table>

Reason for refusal (n)

<table>
<thead>
<tr>
<th></th>
<th>Capture 1</th>
<th>Capture 2</th>
<th>Capture 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not want or refused unique object</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Wanted money and not objects</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 3. Results of the 3-source capture-recapture by capture round during female sex worker population size estimation in Rwanda in 2022.

<table>
<thead>
<tr>
<th>Approach female sex workers (n)</th>
<th>Capture 1</th>
<th>Capture 2</th>
<th>Capture 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>10 (0.6)</td>
<td>19 (1.1)</td>
<td>43 (2.3)</td>
</tr>
<tr>
<td>No</td>
<td>1768 (99.4)</td>
<td>1851 (98.9)</td>
<td>1867 (97.7)</td>
</tr>
</tbody>
</table>

Unique object acceptance, n (%)

<table>
<thead>
<tr>
<th></th>
<th>Capture 1</th>
<th>Capture 2</th>
<th>Capture 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accepted</td>
<td>1766 (99.9)</td>
<td>1848 (99.8)</td>
<td>1865 (99.9)</td>
</tr>
<tr>
<td>Refused</td>
<td>2 (0.1)</td>
<td>3 (0.2)</td>
<td>2 (0.1)</td>
</tr>
</tbody>
</table>

Reason for refusal (n)

<table>
<thead>
<tr>
<th></th>
<th>Capture 1</th>
<th>Capture 2</th>
<th>Capture 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not want or refused unique object</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Wanted money and not objects</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
### Table 4. Sampled female sex workers by capture round, age group, and province as per the female sex worker population size estimation in Rwanda in 2022.

<table>
<thead>
<tr>
<th>Capture, age group (years)</th>
<th>City of Kigali (n)</th>
<th>Eastern province (n)</th>
<th>Northern province (n)</th>
<th>Southern province (n)</th>
<th>Western province (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capture 1 (n=1766)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-17 (n=28)</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>18-24 (n=628)</td>
<td>122</td>
<td>109</td>
<td>162</td>
<td>106</td>
<td>129</td>
</tr>
<tr>
<td>25+ (n=1110)</td>
<td>127</td>
<td>142</td>
<td>222</td>
<td>247</td>
<td>372</td>
</tr>
<tr>
<td>Capture 2 (n=1848)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-17 (n=11)</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18-24 (n=911)</td>
<td>116</td>
<td>130</td>
<td>206</td>
<td>126</td>
<td>329</td>
</tr>
<tr>
<td>25+ (n=926)</td>
<td>239</td>
<td>164</td>
<td>162</td>
<td>124</td>
<td>237</td>
</tr>
<tr>
<td>Capture 3 (n=1865)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-17 (n=35)</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>18-24 (n=851)</td>
<td>184</td>
<td>155</td>
<td>85</td>
<td>92</td>
<td>335</td>
</tr>
<tr>
<td>25+ (n=979)</td>
<td>131</td>
<td>152</td>
<td>276</td>
<td>142</td>
<td>278</td>
</tr>
</tbody>
</table>

A total of 1766 unique objects were distributed countrywide during capture 1, 1848 objects during capture 2, and 1865 objects during capture 3. In a 3-week survey implementation exercise, 62 hotspots were visited countrywide in each capture round; however, bigger hotspots were resampled in the subsequent capture rounds. Two hotspots were resampled between capture 1 and capture 2; 8 hotspots were resampled between capture 2 and capture 3; 6 hotspots were resampled between capture 1 and capture 3; and 2 hotspots were resampled in all 3 capture rounds. Figure 1 shows the maps of the individual captures, highlighting the venue and street hotspots visited. The aggregated and cleaned final 3S-CRC data set was imported into shinyrecap for analysis.

For all 3 capture rounds, 1766 FSWs, 1848 FSWs, and 1865 FSWs were sampled, of which 1408 FSWs, 1471 FSWs, and 1529 FSWs were observed strictly during capture 1, capture 2, and capture 3, respectively. There were 169 exclusive overlaps between capture 1 and capture 2, 210 exclusive overlaps between capture 2 and capture 3, and 65 recaptures between capture 1 and capture 3. Finally, 61 FSWs were recaptured in all 3 capture rounds. Figure 2 presents the Venn diagram illustrating the aggregated data of the capture history results for single, double, and triple captures to construct a structured 3S-CRC data set.
Out of the 231 FSWs recaptured between capture 1 and capture 2, 96 physically presented the unique objects received during capture 1, while out of 135 who did not have the unique objects with them, 134 were able to correctly describe and identify the received unique object on a laminated card, bringing the total number of recaptures to 230. Of the 127 FSWs recaptured between capture 1 and capture 3, 53 brought the unique objects with them, while of the 74 who did not have the unique objects with them, 73 were able to correctly describe and identify the received unique object on a laminated card. Of the 272 FSWs recaptured between capture 2 and capture 3, 111 had the received unique objects with them, while 160 who did not have the unique objects with them were able to correctly describe and identify the received unique object on a laminated card. Table 5 highlights the 2 methods used to record the recapture histories.
Table 5. The recapture identification cascade in the female sex worker population size estimation in Rwanda in 2022.

<table>
<thead>
<tr>
<th>Recapture round (n)</th>
<th>Capture 2</th>
<th>Capture 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capture 1 (C1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total recaptured from C1</td>
<td>231</td>
<td>126</td>
</tr>
<tr>
<td>Showed C1 object</td>
<td>96</td>
<td>53</td>
</tr>
<tr>
<td>Did not have unique objects with them</td>
<td>135</td>
<td>74</td>
</tr>
<tr>
<td>Correctly identified C1 object</td>
<td>134</td>
<td>73</td>
</tr>
<tr>
<td>Capture 2 (C2)</td>
<td>N/Aa</td>
<td></td>
</tr>
<tr>
<td>Total recaptured from C2</td>
<td></td>
<td>271</td>
</tr>
<tr>
<td>Showed C2 object</td>
<td>N/A</td>
<td>111</td>
</tr>
<tr>
<td>Did not have unique objects with them</td>
<td>N/A</td>
<td>161</td>
</tr>
<tr>
<td>Correctly identified C2 object</td>
<td>N/A</td>
<td>160</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

Out of the 135 FSWs who claimed to have been offered capture 1 unique object during capture 2, only 1 was unable to describe and correctly identify the object received on the laminated card. Out of the 74 FSWs in capture 3 who claimed to have been offered a unique object but who did not have the objects with them, 73 were able to describe and correctly identify the object received on the laminated card. Only 1 FSW out of 272 FSWs who claimed to have been offered capture 2 unique object during capture 3 was unable to describe and correctly identify the object received on the laminated card. The FSW population size presented in Table 6 is based on 3 models: log-linear, Bayesian Model Averaging (using noninformative prior), and Bayesian nonparametric latent class models.

Table 6. National population size estimates of female sex workers aged ≥15 years by using the 3-source capture-recapture method during the population size estimation in Rwanda in 2022.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Proportion (%) of women (95% CI)a</th>
<th>Median population size estimation (95% credible set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-linear (Mth Poisson2)b</td>
<td>1 (0.8-1.2)</td>
<td>34,370 (28,164-42,246)</td>
</tr>
<tr>
<td>Bayesian Model Averaging (noninformative prior)</td>
<td>1.1 (0.9-1.3)</td>
<td>37,647 (31,873-43,354)</td>
</tr>
<tr>
<td>Bayesian Latent Class</td>
<td>1 (0.4-1.6)</td>
<td>35,954 (14,736-55,215)</td>
</tr>
</tbody>
</table>

aDenominators are the national total number of adult females aged 15 years and older in the 5th Rwanda Population and Housing Census in 2022.
bPoisson model, which assumes that captures may have different probabilities and that individuals may be heterogeneous.

Based on the outputs and model diagnostics (Multimedia Appendices 1-4), the data were found to contain list dependence; therefore, Bayesian Model Averaging with noninformative prior was chosen, which best dealt with list dependence, as it automatically detected potential dependencies in the data. After fitting the model, the population size of street- and venue-based FSWs in Rwanda was estimated to be within a credible set ranging from 31,873 to 43,354 with a median of 37,647, corresponding to 1.1% (95% CI 0.9%-1.3%) of the general population of adult females aged 15-49 years in Rwanda (Table 7). Relative to adult females in the general population, the western and northern provinces ranked first and second with a higher concentration of FSWs, respectively. The City of Kigali and eastern province ranked third and fourth, respectively. The southern province was identified as having a lower concentration of FSWs.
In 2012, the population size of FSWs was estimated to range from 23,000 to 39,000. Later, after 6 years, in 2018, the national population size of FSWs was estimated to range from 8328 to 22,806, with a median of 13,716. These differences in the population size of FSWs might be attributed to different reasons, including but not limited to methodological or geographical coverage differences. Compared with the previous 3 rounds of FSW PSE exercises, we observed a difference in the FSW population size in our study, which also might be attributed to the reasons stated above.

There are several methods to estimate the population size of population groups without sampling frames [23,24]. Each method presents its own unique strengths and weaknesses. Our study uses the 3S-CRC method to produce FSWs and sexually exploited minors’ PSEs nationally and at the subnational level for 5 administrative provinces. Two captures are used in the classic CRC method; however, additional captures can be added to increase the number of data points from which estimates are generated, resulting in increased ability to account for the potential interaction; so, the assumption of independence between captures may be relaxed [20]. Due to its mathematically grounded and logical results, 3S-CRC is increasingly utilized in epidemiology to estimate the size of the key populations targeted by health intervention programs for certain health disorders [25-27]. The 3S-CRC approach has been utilized in numerous studies to estimate the size of specific population groups—including FSWs, men who have sex with men, and people who inject drugs—without sampling frames [28-32].

As affirmed by the Joint United Nations Program on HIV/AIDS or World Health Organization in the publication guidelines for the second generation Know Your HIV epidemic [12], HIV surveillance among key populations is a priority in all epidemic settings. Identifying the key population groups, their locations, and their sizes helps in understanding and prioritizing the current needs for HIV prevention, diagnosis, treatment, and care services. It also helps in projecting the future needs for those services.

Despite the limitations mentioned above, policy makers and planners should be able to monitor HIV pandemic control nationwide, specifically among FSWs, by using the findings from the 2022 PSE of FSWs. They should also be able to plan for other health services such as sexually transmitted infection prevention and treatment. Even though these estimates can be applied at the national and provincial levels, more work on small area estimation is needed to match PSE results with the targeted geographical coverage.

### Discussion

This study provides both national and provincial-level estimates of the population size of street- and venue-based FSWs and sexually exploited minors aged 15 years and older in Rwanda. In our study, the population size of street- and venue-based FSWs and sexually exploited minors was estimated to be within a credible set ranging from 31,873 to 43,354, with a median of 37,647, corresponding to 1.1% (95% CI 0.9%-1.3%) of the adult females aged 15 years and older in the general population. Our results indicate a significant difference in the FSW population size as compared to the 2018 population size of FSWs aged 15 years and older, which was estimated to range from 8328 to 22,806 credible sets with a median of 13,716 [10]. This difference may be attributed to several factors, including but not limited to differences in the estimation models used and the geographical coverage.

Furthermore, our study provides a provincial-level population size estimate of street- and venue-based FSWs and sexually exploited minors aged 15 years and older for the very first time. The largest population size estimate was obtained in the western province, followed by the northern and eastern provinces. The City of Kigali and southern province were found to have relatively lower estimates of the FSW population as compared to other provinces. Differences in the estimates distribution across the country may reflect long-term internal movement patterns among FSWs, from rural to more urbanizing areas as well as from smaller to larger urbanized contexts, as indicated by the Rwanda Population and Household Census 2022 [22].

The findings from the 2022 FSW PSE might not have considered high-profile FSWs and those FSWs using web-based and social media platforms to reach their clients, leading to a slight possible underestimation of the true population size. Furthermore, we acknowledge possible methodological limitations that might influence the final FSW PSE in this study. Compared to the program coverage data of the Rwanda Health Management Information System, the key strength of our study is that it is powered to provide national and provincial-level PSE for FSWs in Rwanda for the very first time.

So far, 3 rounds of FSW PSE have been conducted in Rwanda since 2010 [7,9,10]. The 2010 FSW size estimation using CRC and multiplier methods estimated the national population size of FSWs to range from 2998 to 3412 with a median of 3205. In 2012, the population size of FSWs was estimated to range from 23,000 to 39,000. Later, after 6 years, in 2018, the national population size of FSWs was estimated to range from 8328 to 22,806, with a median of 13,716. These differences in the population size of FSWs might be attributed to different reasons, including but not limited to methodological or geographical coverage differences. Compared with the previous 3 rounds of FSW PSE exercises, we observed a difference in the FSW population size in our study, which also might be attributed to the reasons stated above.

There are several methods to estimate the population size of population groups without sampling frames [23,24]. Each method presents its own unique strengths and weaknesses. Our study uses the 3S-CRC method to produce FSWs and sexually exploited minors’ PSEs nationally and at the subnational level for 5 administrative provinces. Two captures are used in the classic CRC method; however, additional captures can be added to increase the number of data points from which estimates are generated, resulting in increased ability to account for the potential interaction; so, the assumption of independence between captures may be relaxed [20]. Due to its mathematically grounded and logical results, 3S-CRC is increasingly utilized in epidemiology to estimate the size of the key populations targeted by health intervention programs for certain health disorders [25-27]. The 3S-CRC approach has been utilized in numerous studies to estimate the size of specific population groups—including FSWs, men who have sex with men, and people who inject drugs—without sampling frames [28-32].

As affirmed by the Joint United Nations Program on HIV/AIDS or World Health Organization in the publication guidelines for the second generation Know Your HIV epidemic [12], HIV surveillance among key populations is a priority in all epidemic settings. Identifying the key population groups, their locations, and their sizes helps in understanding and prioritizing the current needs for HIV prevention, diagnosis, treatment, and care services. It also helps in projecting the future needs for those services.

Despite the limitations mentioned above, policy makers and planners should be able to monitor HIV pandemic control nationwide, specifically among FSWs, by using the findings from the 2022 PSE of FSWs. They should also be able to plan for other health services such as sexually transmitted infection prevention and treatment. Even though these estimates can be applied at the national and provincial levels, more work on small area estimation is needed to match PSE results with the targeted geographical coverage.

### Table 7. Female sex worker provincial population size estimates produced using Bayesian Model Averaging with noninformative prior population size estimation in Rwanda in 2022.

<table>
<thead>
<tr>
<th>Province</th>
<th>Proportion (%) of women 15-49 years who were female sex workers (95% CI)a</th>
<th>Median population size estimation (95% credible set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>City of Kigali</td>
<td>0.8 (0.5-1)</td>
<td>3974 (2815-5197)</td>
</tr>
<tr>
<td>Eastern province</td>
<td>0.6 (0.3-1)</td>
<td>5022 (2535-8601)</td>
</tr>
<tr>
<td>Northern province</td>
<td>1.1 (0.7-1.6)</td>
<td>5993 (3710-8876)</td>
</tr>
<tr>
<td>Southern province</td>
<td>0.5 (0.2-0.9)</td>
<td>3884 (1548-6727)</td>
</tr>
<tr>
<td>Western province</td>
<td>1.2 (0.9-1.6)</td>
<td>8983 (6536-11,791)</td>
</tr>
</tbody>
</table>

aDenominators are the provincial total number of adult females aged 15 years and older in the 5th Rwanda Population and Housing Census in 2022.
HIV treatment and prevention efforts at lower subnational levels. Further, we admit that some FSW groups can still be difficult to estimate, including but not limited to those reaching their clients on web-based platforms; this could be a topic for future research.

**Acknowledgments**

The first author would like to thank all the study investigators and coauthors of this manuscript for their valuable contributions to making this project a success. The population size estimation of female sex workers was implemented by the Rwanda Biomedical Center and was supported by the President’s Emergency Plan for AIDS Relief through the Centers for Disease Control and Prevention under the terms of GH002203. The findings and conclusions in this paper are those of the author(s) and do not necessarily represent the official position of the funding agencies.

**Data Availability**

The data that inform the findings in this manuscript are client-level data fully owned by the Rwanda Ministry of Health and the Rwanda Biomedical Center. The data cannot be shared publicly to protect client privacy and confidentiality, considering that female sex workers is a sensitive key population group. However, researchers who meet set criteria can access the data upon request and approval by the Rwanda Biomedical Center (data access requests should be sent to info@rbc.gov.rw).

**Authors’ Contributions**

All authors made significant contributions to the study’s design, conduct, or data analysis and interpretation, as well as writing or editing portions of the work and approving the final version for clearance and publication.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Multiple capture-recapture power analysis.

[PNG File, 76 KB - publichealth_v10i1e50743_app1.png ]

**Multimedia Appendix 2**

Log-linear analysis for producing the final population size estimate with credibility sets from aggregated data sets in Rwanda in 2022.

[PNG File, 149 KB - publichealth_v10i1e50743_app2.png ]

**Multimedia Appendix 3**

Bayesian Latent Class for producing the final population size estimate with credibility sets from aggregated data sets in Rwanda in 2022.

[PNG File, 87 KB - publichealth_v10i1e50743_app3.png ]

**Multimedia Appendix 4**

Bayesian Model Averaging for the final population size estimate in Rwanda in 2022.

[PNG File, 55 KB - publichealth_v10i1e50743_app4.png ]

**References**


9. Estimating the size of populations through a household survey (ESPHS) - Rwanda 2011. Rwanda Biomedical Center/Institute of HIV/AIDS. URL: medbox.org/pdf/5e148832db60a2044c2d4b27 [accessed 2024-02-27]


18. shinyrecap. fellstat. 2023 Nov 05. URL: fellstat.github.io/shinyrecap/ [accessed 2023-11-25]


Abbreviations

3S-CRC: 3-source capture-recapture
CRC: capture-recapture
FSW: female sex worker
PSE: population size estimation
RBC: Rwanda Biomedical Center

©Elysee Tuyishime, Eric Remera, Catherine Kayitesi, Samuel Malamba, Beata Sangwayire, Ignace Habimana Kabano, Horacio Ruisenor-Escudero, Tom Oluoch, Angela Unna Chukwu. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 15.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Triangulating Truth and Reaching Consensus on Population Size, Prevalence, and More: Modeling Study

Ian E Fellows¹*, PhD; Carl Corcoran¹*, PhD; Anne F McIntyre¹, MPH, PhD

Division of Global HIV & TB, Global Health Center, Centers for Disease Control and Prevention, Atlanta, GA, United States

*these authors contributed equally

Corresponding Author:
Carl Corcoran, PhD
Division of Global HIV & TB
Global Health Center
Centers for Disease Control and Prevention
1600 Clifton Rd NE
Atlanta, GA, 30333
United States
Phone: 1 (404) 713 3545
Email: ruu2@cdc.gov

Abstract

Background: Population size, prevalence, and incidence are essential metrics that influence public health programming and policy. However, stakeholders are frequently tasked with setting performance targets, reporting global indicators, and designing policies based on multiple (often incongruous) estimates of these variables, and they often do so in the absence of a formal, transparent framework for reaching a consensus estimate.

Objective: This study aims to describe a model to synthesize multiple study estimates while incorporating stakeholder knowledge, introduce an R Shiny app to implement the model, and demonstrate the model and app using real data.

Methods: In this study, we developed a Bayesian hierarchical model to synthesize multiple study estimates that allow the user to incorporate the quality of each estimate as a confidence score. The model was implemented as a user-friendly R Shiny app aimed at practitioners of population size estimation. The underlying Bayesian model was programmed in Stan for efficient sampling and computation.

Results: The app was demonstrated using biobehavioral survey-based population size estimates (and accompanying confidence scores) of female sex workers and men who have sex with men from 3 survey locations in a country in sub-Saharan Africa. The consensus results incorporating confidence scores are compared with the case where they are absent, and the results with confidence scores are shown to perform better according to an app-supplied metric for unaccounted-for variation.

Conclusions: The utility of the triangulator model, including the incorporation of confidence scores, as a user-friendly app is demonstrated using a use case example. Our results offer empirical evidence of the model’s effectiveness in producing an accurate consensus estimate and emphasize the significant impact that the accessible model and app offer for public health. It offers a solution to the long-standing problem of synthesizing multiple estimates, potentially leading to more informed and evidence-based decision-making processes. The Triangulator has broad utility and flexibility to be adapted and used in various other contexts and regions to address similar challenges.

(JMIR Public Health Surveill 2024;10:e48738) doi:10.2196/48738

KEYWORDS

HIV; epidemiology; population size estimation; key populations; Bayesian models; consensus estimation; statistical tool; prevalence; Bayesian model; population; estimate; consensus; population size

Introduction

One of the more daunting tasks for policy makers in the public health arena is the synthesis of data and information into actionable insights. Population size, prevalence, and incidence are crucial metrics that play a significant role in shaping public health programs and policies. However, stakeholders are often faced with multiple estimates of these quantities, originating from different sources and of differing quality. Often, estimates
of the same quantity will be incompatible, with confidence bounds completely disjoint from one another.

Stakeholders face the task of weighing their knowledge about the population along with the various estimates to triangulate to a single number that represents the most likely value. Typically, this has been done without a formal framework, where decisions about how the final number was arrived are shrouded by a nontransparent process. How much stock was each estimate given by the stakeholders? Did any of the stakeholders express strong prior beliefs about what the true number was that could affect the result? Did political considerations color the findings, and if so, how? Without transparency, these questions are difficult to answer.

The goal of this study is to present a statistical tool that can be used to guide the triangulation process. Prior beliefs, study quality, and estimate uncertainty are all components of the process of generating a consensus estimate. However, because these components are put together in a rigorous statistical framework, all of them are made completely transparent and inspectable by third parties.

A particular focus is paid to population size estimation (PSE). PSE is one of the fundamental estimates required for policy decisions in HIV treatment and prevention. Understanding the population size is essential for determining the appropriate scale of the public health response. Without this knowledge, it becomes challenging to tailor interventions effectively. This is especially true for key populations (KP), which are less visible groups disproportionately at risk for HIV, such as female sex workers (FSW), men who have sex with men (MSM), and people who inject drugs [1]. Globally, KP and their sexual partners account for more than half of the new HIV infections as of 2019 [2]. Reliable estimates of KP sizes are needed to inform the planning and implementation of prevention and treatment programs as well as to assess the outcomes of these control measures. In particular, evaluating progress toward the Joint United Nations Programme on HIV/AIDS 95-95-95 goals (95% of people living with HIV know their status, 95% of those who know their status are receiving antiretroviral therapy [ART], and 95% of those receiving ART are virally suppressed) [3] requires accurate population size estimates.

Although these estimates have significant implications, there is currently no gold standard method for PSE [4]. In the absence of such a standard, various techniques have been developed, each with varying degrees of rigor [5]. On the one hand, estimates can be based on nonempirical data such as the opinion or experience of subject matter experts, such as the wisdom of crowds [6,7] or the Delphi method [8,9]. On the other hand, empirical data–driven biobehavioral survey (BBS)–based techniques including service or unique object multipliers [10] and recruitment information captured in surveys using respondent-driven sampling [11] can provide more reliable population size estimates [12] but can still produce incongruous results [13]. Moreover, CIs only capture part of the uncertainty we have in an estimate. Mistakes in implementation, violation of method assumptions, and incorrect statistical model application can all add additional bias and uncertainty that are not captured by the CI. A major question is how to synthesize multiple estimates of varying quality and certainty, given the practical need of policy makers and scientists for a single best estimate of population size. Although a range of estimates with uncertainty bounds derived from multiple statistical methods may be a more comprehensive snapshot, multiple estimates are often challenging to interpret if they are disparate and difficult to apply to problem-solving or performance evaluation.

To address this issue, a consensus best estimate is often arrived at somewhat arbitrarily by round table discussions of subject area experts; however, the absence of empirical evidence in such processes makes them susceptible to bias or statistical errors.

Recently, more rigorous statistical methods have emerged for finding a consensus-based estimate based on approaches in the field of meta-analysis. Meta-analysis is a branch of statistics concerned with synthesizing the results of multiple studies that aim to estimate the same quantity, often an effect size, in medical or epidemiological studies [14]. Traditional meta-analysis approaches have been based in frequentist statistics, although Bayesian techniques are becoming more common. Table 1 shows a component of the spectrum of frequentist models in meta-analysis for combining the effects of studies. In the fixed effects model, it is assumed that each study estimates the same effect, and only within-study variation is incorporated. In the random effects model, each study is assumed to estimate a different effect drawn from a population distribution and thus incorporates between-study variation. However, both models take the results of each study at face value. With the goal of incorporating bias adjustment into more traditional models, the quality effects model [15] has recently been introduced, and it has been shown to have fewer limitations than competing models [16]. In the quality effects model, each study considered is appraised on its methodological quality and given a quality score, which is then used to adjust the study variance within the model. Although other meta-analysis methods have Bayesian PSE analogs, such as the Anchored Multiplier, which draws on both fixed and random effects models [17,18], the quality effects model has yet to be adapted for use triangulating public health quantities.

In this study, we present a Bayesian hierarchical model to synthesize multiple study estimates that allow the user to incorporate the quality of each estimate as a study confidence score. We developed a mathematical framework for the model as well as a metric to assess the variation between estimates that have been captured by the model. We also introduced an implementation of the model as a user-friendly R Shiny app aimed at practitioners of PSE. To demonstrate both the model and the app, we provide several examples of their use in combining KP size estimates in several cities in a country in sub-Saharan Africa, showing the potential of our model to provide a valid best estimate in a practical context.
Methods
Overview
We began the construction of our model by considering a Bayesian hierarchical model that is fundamental in meta-analysis [19]. The primary goal of such a meta-analysis is to estimate the mean of the distribution of effect sizes, synthesizing the estimates of individual studies. Our goal was similar: to estimate a population quantity of interest \( \theta \) which can be a population size, population proportion, incidence, among others. We aimed to synthesize several estimates \( y_j \) of this quantity from different studies, where the variance \( \sigma_j^2 \) of each estimate is known.

Despite the assumption that the variance \( \sigma_j^2 \) is known (ie, has been computed during the processing of the \( j \)th study), it may not truly reflect the uncertainty in the estimate of \( y_j \). The variance \( \sigma_j^2 \) only accounts for the sampling uncertainty and not any additional uncertainty that there may be about potential bias in the study design. For example, in PSE, certain size estimation techniques result in incongruous estimates and nonoverlapping CIs [13], indicating that there is additional nonsampling error not accounted for by the CIs and \( \sigma_j^2 \). Nonsampling bias can be introduced for a number of reasons, including study design assumption violation, a mismatch between the study population and the population of interest, or an out-of-date sample. Our model provides a simple way for the user to account for their assessment of the degree of this additional uncertainty in the study. We allow the model user to adjust this uncertainty based on their confidence level \( c_j \in (0,1) \) in the study, where \( c_j = 1 \) represents full confidence in the study’s estimate and uncertainty. The SD is scaled by the transformation \( \sigma_j \to \sigma_j / c_j \) to adjust for this additional uncertainty at the discretion of the user. This allows practitioners and experts, whose expertise may carry information about each estimate not accounted for in the given CIs, to manually incorporate the potential for study bias.

The synthesis itself is based on a Bayesian hierarchical model that is often used in meta-analysis. At the first level of the hierarchy, we assume that each \( y_j \) is a point estimate of the quantity \( y_j \) estimated by the \( j \)th study:

\[
y_j 
\]

A normal model is chosen because many point estimates tend to be asymptotically normal according to the central limit theorem. That the \( y_j \) are estimates of \( v_j \), and not \( \theta \) directly, is an important distinction; individual studies may have bias, and that bias may not be accounted for by the user-specified study confidences. At the second level of the hierarchy, we assume that each study’s quantity of interest \( v_j \) itself is centered around the true population quantity of interest \( \theta \) with uncertainty characterized by the between-study variance \( \tau^2 \). At this level, we assume that the \( v_j \) are distributed normally as

\[
\]

Although the \( v_j \) provide a bridge between the data \( y_j \) and the target quantity \( \theta \), the between-study variance \( \tau^2 \) provides an important source of uncertainty beyond the adjusted study-level variances (\( \sigma_j^2 / c_j^2 \)). When \( \tau^2 \) is small, the \( v_j \) are similar, and most of the uncertainty in the model comes from the data level; when \( \tau^2 \) is large, most of the uncertainty comes from differing \( v_j \). However, from a practical perspective, the \( v_j \) represent a nuisance parameter that can significantly slow down the computational model. Eliminating this parameter by marginalizing it out yields the reduced model (refer to Multimedia Appendix 1 [18] for details):

\[
\]

In collapsing the hierarchical model to a single level, we can quickly and efficiently infer the mean of the distribution of the quantity of interest, given our data \( y_j \).

Prior Distribution for \( \theta \) and \( \tau^2 \)
As this is a Bayesian model, we must consider the distribution of the hyperparameters \( \theta \) and \( \tau^2 \). In general, we assume a normal prior for \( \theta \) with mean \( \mu_0 \) and variance \( \sigma_0^2 \):

\[
\]

The specification of this prior distribution for \( \theta \) is an important part of the inference process, as experts and stakeholders will often have informed beliefs about reasonable values for the population quantity.

Setting a prior distribution for \( \tau^2 \) is more complex, as there are multiple candidate distributions, each with advantages and
disadvantages; refer to the study by Williams et al [20] for discussion of the choice of prior on $\tau$ and why it may be desirable to place the prior on $\tau$ instead of $\tau^2$. We choose a half-Cauchy distribution for our prior on $\tau$, which is a Cauchy distribution truncated so that only positive values have a nonzero probability density. The distribution is heavy tailed, which is important because the model must allow for the possibility of large between-study variance. That said, we wish to trust that our stakeholders have accounted for the additional study variability in the confidence score step and thus wish to construct an informative prior that puts most of its mass at a very small $\tau^2$ but which allows for it to be large if the data are incompatible with a small $\tau^2$. This has the effect of flexibly controlling the model complexity, with the $v_j$ values only being appreciably different from $\theta$ if the data require it.

We control this distribution with a single-scale parameter $\gamma$, and thus, we choose that parameter to accumulate most of the probability mass for $\tau^2$ at 0. We use an empirical Bayes–flavored choice, which is scaled to be a fraction of the sample SD of the estimates $y_j$:

$$s^2 = \text{Var}(y_j).$$

As a heuristic justification for the choice of 0.1 as the value of the multiplier on the sample SD, we note that the quantile function for a half-Cauchy distribution with scale parameter $\gamma$ is $\gamma$, and thus, the 95th percentile is given by $Q(0.95) = 12.7\gamma$. As such, with $\gamma = 0.1$, a little more than 95% of the probability mass of $\tau^2$ is below the sample SD $s$ which we have found is often sufficient to create a high burden for values of $\tau^2$ away from 0.

### Transforming the Data and Parameter Space

The hierarchical model assumed a normal distribution for the estimates. This is generally justifiable, as most estimates follow the central limit theorem and are therefore approximately normal if the sample size of the study is large enough. However, depending on the quantity, the central limit theorem may be applied to smaller sample sizes when transformed. Regardless, the Delta method provides some assurance that transformed point estimates are asymptotically normal if the untransformed estimate is also asymptotically normal.

For instance, a log transform is typically used in PSE, where log-linear models are a prime example. Models of proportions often involve a logistic transformation, as seen in techniques such as logistic regression. With our model, the user may apply a transformation to the data, run the model, and then apply an inverse transformation to obtain inferences regarding the quantity of interest.

Working in a transformed space also transforms the prior, which can be desirable in and of itself. The log transform of PSE causes the prior to be log-normal. This may be a more reasonable prior distribution for this quantity because it is always >0. Similarly, for the logistic transform, the prior becomes a logistic normal distribution whose values are bounded from 0 to 1.

Transformations may induce infinities in some edge cases (eg, a point estimate of zero or one with a logistic transformation). In these cases, either the estimate may be dropped, or the analyst can choose not to use a transformation.

### Relationship With Frequentist Meta-Analysis Models

The frequentist methods in Table 1 describe increasingly complex models for estimating the combination. The fixed effect model simply calculates a weighted average of the estimates, weighted by the inverse variance of each estimate. Our model is the Bayesian analog of this fixed effect model when all study confidences are set to 100% and $\tau \rightarrow 0$. For $\tau >0$ and the study confidences set to 100%, our model is the Bayesian analog of the frequentist random effects model, where the $\tau$ in the random effects model is approximately analogous to our $\tau$. By positing a heavy-tailed prior for $\tau$ highly concentrated at 0, our model leans toward a fixed effect model when the data are compatible with it and a random effects model when the data are not compatible with it.

The quality effects model introduces a researcher-defined quality metric to the random effects model in much the same way that our study confidence does. Thus, when study confidence was included, the features of the quality effects model were incorporated.

### Explained Variance and Unaccounted-For Variation

Although we use the reduced model for computational reasons, the full multilevel model contains information about the sources of uncertainty that may have a significant impact on the confidence a user has in the resulting estimate of $\theta$. To quantify the sources of uncertainty, we computed a metric called explained variance, or $R^2$, which describes each level of the full model [21]. In the full multilevel model, $R^2$ calculated at the study (data) level is defined as

$$R^2 = \frac{\text{var}(\text{Estimate}) - \text{var}(\text{Estimate} - \text{Null})}{\text{var}(\text{Estimate})}.$$  

where $\text{var}(\cdot)$ is the posterior mean and $\text{var}(\cdot)$ is the sample variance operator. From this definition, an observation can be made: when the $v_j$ more closely approximate the $y_j$, $R^2$ is closer to 1, and when the $v_j$ more closely approximate $\theta$, $R^2$ is closer to 0. In the case where there is considerable variation in the estimates $y_j$, $R^2$ reflects the proportion of estimate variability attributable to unaccounted-for study bias. When the variation in the estimates is largely owing to unaccounted-for study bias, the $v_j$ will approximate the $y_j$, leading to high $R^2$. When study bias is accounted for by increasing the uncertainty of the estimates, the $v_j$ will more closely approximate $\theta$, leading to low $R^2$. To reflect this relationship, we use the term unaccounted-for variation to refer to $R^2$.

A complication with the abovementioned formula is that it relies on the parameters $v_j$ that have been eliminated in the reduced model. However, we can compute $R^2$ at the study level using the data and posterior draws of $\theta$ and $\tau^2$ that are specific to the model at hand. We find that
A full derivation of the preceding formulas can be found in Multimedia Appendix 1 [18].

Example Using BBS-Based PSE Methods

Overview

The population size estimates used to demonstrate the Bayesian hierarchical model using the Shiny app were all part of a 2019 BBS conducted among FSW in 3 cities (location A, location B, and location C) and MSM in 2 cities (location B and location C) of a country in sub-Saharan Africa. Priors and corresponding plausibility bounds were based on nonempirical stakeholder consensus during the 2014 BBS. In each survey city and for each key population (FSW and MSM), 2 methods of PSE were implemented: multipliers and successive sampling.

Multipliers

As part of a BBS, unique object, event, and service multipliers provide data sources to facilitate PSE. Multipliers have 2 overlapping components, both of which include members of a target population such as FSW or MSM. For each of the 3 multipliers, the first component takes place shortly before the BBS implementation. For unique object multipliers, small, inexpensive, and memorable unique objects (gifts) are distributed to target populations by teams of key population peers within the survey catchment area. For event multipliers, a party or concert is hosted for target population members. For service multipliers, client lists from HIV service providers are cleaned and deduplicated. The counts are tallied of individuals who received a gift, attended an event such as a survey kick-off party or concert, or accessed HIV services from a particular provider during defined dates. The second component is the weighted proportion of BBS respondents who reported receiving a gift, attending an event, or accessing HIV services. A population size estimate results when the counts are divided by the weighted proportion. These are reported with the corresponding 95% CIs.

For FSW in all 3 locations and MSM in location C, the survey multipliers included a unique object distributed in the survey catchment areas, a special event for the target population, 2 providers of testing services, and 1 outreach provider. The MSM survey at location B included an additional fourth outreach service multiplier.

Successive Sampling

As part of a BBS using respondent-driven sampling for recruitment, participants were asked a series of questions to determine their personal network size that informs the survey weighting. One assumption is that individuals with larger social networks are more likely to be sampled and more socially visible and, thus, recruited into surveys earlier than those with smaller networks [22]. The successive sampling PSE (SS-PSE) method has been described elsewhere [23] and applies this assumption to self-reported personal network size and the order of recruitment to estimate the population size. The SS-PSE used in this study uses imputed visibility to adjust for errors in network size reporting [24]. The SS-PSE was generated for FSW and MSM in all survey locations.

Confidence Scores

The purpose of the confidence score is to quantify, on a scale of 0 to 100, the quality or reliability of each individual PSE generated by various methods. It aims to provide a measure of how much trust or confidence one should place in each PSE. Confidence scoring requires knowledge of the methods and quality of PSE. Elements to consider for each PSE include whether the definitions of the populations are aligned; when (ie, Were the various studies conducted concurrently as part of the same BBS? Within 1 to 2 years of each other? Several years apart?) and how each of the PSE methods were implemented and analyzed; and how to apply that information to each PSE independent of, not relative to, other PSE. We have more confidence in empirical methods implemented concurrently in the same catchment area among populations with the same or similar characteristics, such as age range, sex, residence, and behaviors, with high-quality, complete data and minimal errors. Therefore, we assigned a high score that reflects the confidence we have in the PSE derived from an activity meeting these criteria. In contrast, if we have data from similar but not the same populations, conducted several years apart and sampling different catchment areas, with some missing data, we have less confidence in those PSE and assign a low score. Adjusting confidence scores or dropping low-quality PSE to produce a specific, desired outcome is strongly discouraged.

All PSE methods are subject to errors during implementation. Unique object, service, and event multipliers are commonly included with BBS. The first of the multipliers (or, if sampled twice before a survey, 3-source capture-recapture) involves offering unique objects to key population members in hot spots, typically carried out by teams of their peers (eg, FSW teams distributing unique objects to fellow FSW in venues where they congregate, often referred to as hot spots). The selection of which hot spots to visit (geographic coverage), who is responsible for distributing the unique objects (ideally, peers within the key population to increase acceptability), and the distribution method used (eg, hasty, clustered distribution vs a more systematic approach to peers in the hot spot) can vary significantly, potentially introducing bias into the data collection process. Broad geographic coverage implemented by peers using a systematic or random approach to distribute enough (eg, twice the survey sample size) unique objects would earn a high score. Next, health service client lists may be outdated (ie, include former or expired clients), contain duplicates, or exclusively cover a portion of the population (eg, an ART client registry representing only individuals who are HIV positive and actively receiving treatment). These challenges would result in a low score. Events may draw only a specific subgroup of the target population or fail to attract enough participants, which may compromise the reliability of the PSE. This was particularly apparent when the PSE produced by the event multiplier was smaller than the survey sample size. This scenario would result in a nonplausible PSE and a very low score.

In respondent-driven sampling surveys, participants were tasked with providing information regarding their personal network
size. This can present a significant challenge for respondents to estimate, as it is a function of the clarity of survey questions and a respondent’s ability to quantify the size of their network quickly and accurately. The response may not reflect the actual size. Even after adjusting the self-reported network size for social visibility, the resulting data could potentially impact the subsequent PSE during the successive sampling process. Confidence scores were independently applied to each PSE method and reflected the quality of implementation based on the elements described in this section as well as the plausibility of the estimates.

Ethical Considerations

The BBS survey was reviewed and approved by the ministry of health in the sub-Saharan African country; the implementing partner; and the Centers for Disease Control and Prevention Global Health Center in Atlanta, United States. The data collection staff completed training on human subjects research and signed a confidentiality agreement before commencing their survey duties. All participants provided written informed consent.

Results

Confidence Scores

We reviewed all PSE methods for FSW and MSM at each survey location and assigned a confidence score (Table 2). The confidence score was based on our knowledge of the rigor of the methods used for PSE and the quality of implementation of those methods. The scores were assigned to each method independent of other methods implemented in the survey; thus, a confidence score for one method was neither relative nor comparative to another method. Confidence scores were selected from the interval (0,100), with 100 representing full confidence. These scores were then divided by 100 to yield the confidence score $c_j \in (0,1)$, as described in the Methods section.

Table 2. Survey-based population size estimates by location and key population.

<table>
<thead>
<tr>
<th>Estimate method</th>
<th>Confidence score</th>
<th>Population size estimate (95% CI)</th>
<th>FSW$^a$ in location A</th>
<th>FSW in location B</th>
<th>MSM$^b$ in location B</th>
<th>FSW in location C</th>
<th>MSM in location C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>0</td>
<td>800 (300-2000)</td>
<td>3000 (1800-3400)</td>
<td>2416 (850-4000)</td>
<td>900 (825-1500)</td>
<td>610 (475-685)</td>
<td></td>
</tr>
<tr>
<td>Unique object</td>
<td>80</td>
<td>679 (525-1024)</td>
<td>382 (228-949)</td>
<td>221 (174-904)</td>
<td>690 (483-1191)</td>
<td>379 (280-681)</td>
<td></td>
</tr>
<tr>
<td>Event</td>
<td>5</td>
<td>162 (126-249)</td>
<td>144 (69-354)</td>
<td>96 (81-205)</td>
<td>205 (146-378)</td>
<td>122 (87-276)</td>
<td></td>
</tr>
<tr>
<td>Service 1</td>
<td>60</td>
<td>849 (630-1369)</td>
<td>1984 (1033-7367)</td>
<td>515 (384-4398)</td>
<td>1540 (1181-2260)</td>
<td>2322 (1158-6985)</td>
<td></td>
</tr>
<tr>
<td>Service 2</td>
<td>60</td>
<td>2766 (1995-4622)</td>
<td>4459 (3082-10,219)</td>
<td>1917 (3493-1218)</td>
<td>2937 (2127-4613)</td>
<td>1917 (1394-4093)</td>
<td></td>
</tr>
<tr>
<td>Service 3</td>
<td>60</td>
<td>668 (500-1082)</td>
<td>5449 (1769-30,997)</td>
<td>778 (660-5256)</td>
<td>2004 (1138-5916)</td>
<td>156 (97-494)</td>
<td></td>
</tr>
<tr>
<td>Service 4</td>
<td>60</td>
<td>N/A$^d$</td>
<td>5900 (2674-45,597)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>SS-PSE$^e,f$</td>
<td>70</td>
<td>674 (318-2426)</td>
<td>2196 (1651-2382)</td>
<td>2205 (382-10,409)</td>
<td>1057 (576-3369)</td>
<td>664 (405-1614)</td>
<td></td>
</tr>
<tr>
<td>Consensus: scaled$^g$</td>
<td>—</td>
<td>890 (614-1342)</td>
<td>2744 (2046-3768)</td>
<td>1412 (727-3111)</td>
<td>1038 (781-1333)</td>
<td>612 (517-730)</td>
<td></td>
</tr>
<tr>
<td>Consensus: unscaled$^e$</td>
<td>—</td>
<td>729 (403-1335)</td>
<td>2821 (2080-3856)</td>
<td>1638 (812-3416)</td>
<td>922 (701-1205)</td>
<td>606 (508-724)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$FSW: female sex workers.
$^b$MSM: men who have sex with men.
$^c$Not applicable (this refers to the fact that confidence scores are only applicable to the estimates).
$^d$N/A: not applicable (this refers to the fact that an estimate of type Service 4 was not collected).
$^e$95% CI refers to the credible interval.
$^f$SS-PSE: successive sampling population size estimation.
$^g$The implementation of PSE methods was similar across all survey locations for FSW and MSM, so our confidence scores were applied consistently by method. This was appropriate because our FSW and MSM populations had the same population definitions across survey sites and the methods were concurrently and consistently implemented. Had any survey site or population struggled with implementation, that site and the PSE method would have been assigned a lower score. At all survey sites, enough unique objects were broadly distributed across the survey catchment area, and the PSE from unique object distribution in FSW and MSM hot spots was plausible; thus, we assigned a high confidence score of 80. Providers of HIV testing and outreach services reviewed and used similar methods to clean their client lists; however, the services did not represent the entire FSW or MSM population in any of the cities, only those who pursued testing or engaged in outreach services. The service providers were unable to guarantee that the lists contained only current clients (ie, no former or expired clients) and were fully deduplicated; thus, they were assigned confidence scores of 60. The service providers varied by key population and city but were provided with the same instructions on how to clean and deduplicate the client lists, so the confidence scores...
were applied consistently for FSW and MSM services across all cities. Finally, all event multipliers indicated low overall participation, with only a small proportion of survey respondents reporting attendance. The events were similar in content and attendance for FSW and MSM across the survey cities. The resulting PSE for FSW and MSM in each survey city was smaller than the survey sample sizes. Despite being a low-quality, nonsensical PSE, the results were not dropped from the model; they were assigned a low confidence score of 5. The distributions of self-reported personal network sizes for FSW and MSM in each survey city were plausible with a few outliers; therefore, we adjusted for social visibility to reduce the impact of those outliers on the PSE. The SS-PSE using imputed visibility for FSW and MSM was consistent with previous study results and program data, so SS-PSE was assigned a high confidence score of 70.

**Using the Shiny App**

The confidence-scaled consensus estimation model is deployed in an easy-to-use Shiny app that requires only a few inputs from the user. There are 3 pages on the app: Enter Estimates, Define Prior Beliefs, and Synthesis.

As an example, we consider estimating the size of the population of FSW in location A of our example country in sub-Saharan Africa. The 6 PSEs and the prior are all specified in Table 2. The results of the model after running the app are also included in Table 2.

The prior estimate for the population is 800 FSW, with plausible bounds of (300, 2000). This prior was arrived at by consensus among stakeholders and experts based on the 2014 BBS-based population size estimates using nonempirical methods such as the wisdom of crowds and literature reviews. The relatively wide bounds cover most of the new PSEs, which are all similar except for the Event and Service 2. The low Event estimate is mitigated by its low confidence, and as such, it will have less of an effect during synthesis. However, the high Service 2 estimate has relatively high confidence, so we would expect this estimate to pull the posterior mean higher.

The estimates are entered in the Enter Estimates tab of the app (Figure 1A) along with the upper and lower bounds of the CI and the confidence in the study (ranked out of 100). The other choice that the user must make on this tab is whether to apply a transformation to the data, with the options being no transformation, log transformation, and logit transformation. The log transformation is recommended when working with population size estimates, whereas the logit transformation is recommended for percentage or proportion data.

**Figure 1.** The (A) Enter Estimate tab and the (B) Define Prior Beliefs tab of the Shiny app. The Enter Estimates tab requires the user to input any number of estimates, upper and lower bounds of the 95% CIs of those estimates, and confidence scores (out of 100) as well as input the estimate type to determine the most appropriate transformation of the input data. The Define Prior Beliefs tab requires users to input the median and 75th percentile for the prior distribution of the quantity being estimated and optionally maximum and minimum values that the quantity estimated can take on.

The prior beliefs are then entered in the Define Prior Beliefs tab (Figure 1B). The median of the prior should match the prior estimate of the population size. The 75th percentile, which is used to determine the spread of the prior, is also required. This is meant to be a practical and flexible way to characterize the uncertainty in the prior. If the user has reliable information about the prior uncertainty, such as the 95% CI, the 75th percentile can be computed and entered. If the user does not have reliable information about the uncertainty, they may try several options for the 75th percentile, and the app will resample to create the prior distribution shown in Figure 1B until the user believes the distribution shown reasonably reflects their knowledge of the population size.

The lower and upper bounds on the prior are not required for the model to run but can speed up the computation and provide a more reliable synthesis of the estimates by ruling out impossible prior values. In this example, the lower bound is set to 1 because there must be at least 1 FSW to make the population size worth estimating. The upper bound is set to 3000, which is somewhat arbitrary but makes improbably large population sizes unlikely.

Once prior information has been entered, the model can be run on the Synthesis tab. One additional input is the \( \tau \) multiplier, which helps to set the scale parameter for the prior on \( \tau \). We have found that this default works well in most cases, although it is possible that it may be appropriate to change it in some circumstances. For instance, one might wish to decrease the
multiplier (eg, to 0.01 or 0.001) if there is good reason to believe that all PSEs are of precisely the same population, such as if the studies were done at the exact same time with the exact same catchment area. In this case, barring errors in the implementation of the studies, the between-study variance would be expected to be close to 0, and thus, the prior on \( \tau \) can be adjusted to make high values of \( \tau \) unlikely. In contrast, if less is known about the implementation of the studies and there is reason to believe that there might be some true between-study variance, the multiplier could be increased (eg, to 1) to allow \( \tau \) to take on larger values or to make the prior less informative.

The results show both the prior and the posterior distributions of the population size \( \theta \) as well as descriptive statistics for the posterior distribution of \( \theta \) (Figure 2). Notably, the posterior median was higher and the uncertainty was lower when compared with the prior. The observation that the posterior variance is smaller than the prior variance is generally true: even if the PSEs are given very low confidence (and thus very high uncertainty), the posterior will look very similar to the prior. Refer to the Discussion section for further observations on this point. This tab also reports the unaccounted-for variation at the study level (\( R^2 \)) as the “percent of estimate variability attributable to unaccounted-for study bias.” In this example, \( R^2 = 11\% \), indicating that the confidence scaling in the model accounts for most of the nonsampling error that explains the differences among the study estimates. This is compared with the unscaled case (Table 2), where \( R^2 = 92\% \), indicating the incongruity of the estimates. A forest plot of the prior, estimates, and consensus results (for both confidence scaled and unscaled) is also shown in Figure 2, with both the reported CIs for the estimates (solid) as well as the scaled CIs (dotted).

Figure 2. The Synthesis tab of the Shiny app, which includes the consensus estimate and a forest plot of the prior, the estimates, and the consensus estimate as well as the unaccounted-for variation. For the estimates in the forest plot (black), the solid lines represent the unscaled 95% CI, whereas the dotted lines represent the confidence-scaled 95% CI.

### Other Cases

To further demonstrate the range of the model, we considered 4 other populations to estimate in our example country in sub-Saharan Africa: FSW and MSM in location B and FSW and MSM in location C (Figures 3 and 4). For each population, we consider an unscaled consensus estimate (each estimate’s confidence level is set to \( c_j = 1 \)) as well as a confidence-scaled estimate, where the confidence levels for each estimate are selected by expert opinion. The priors, estimates, and consensus estimates for these 4 populations are shown in Table 2. These 4 estimates show that the effects of confidence scaling are not limited to always increasing or decreasing the consensus estimate. Using confidence scaling over the unscaled version of the model, the point estimate for the population size decreased by 3% for FSW in location B, decreased by 14% for MSM in location B, increased by 13% for FSW in location C, and increased by 1% for MSM in location B. Furthermore, the magnitude of difference in the estimates for FSW in location C and MSM in location B shows that confidence scaling can have a major effect on the estimates.

The unaccounted-for variation for both the confidence-scaled and unscaled estimates for each population also shows the positive effects of having expert-selected confidence levels for the estimates. For the unscaled estimates, the proportion of estimate variation owing to unaccounted-for study bias is 0.89 for FSW in location B, 0.84 for MSM in location B, 0.90 for FSW in location C, and 0.91 for MSM in location C. For the confidence-scaled estimates, the proportion of estimate variability attributable to unaccounted-for study bias is 0.011 for FSW in location B, 0.02 for MSM in location B, 0.16 for FSW in location C, and 0.23 for MSM in location C. These results indicate that for the confidence-scaled models, there is
less unexplained variation in the $y_j$ owing to between-estimate variation when compared with the unscaled models.

Figure 3. Forest plots of prior distribution of the true population size $\theta$, population size estimates, and consensus estimates for (A) female sex workers (FSW) in location B, (B) men who have sex with men (MSM) in location B, (C) FSW in location C, and (D) MSM in location C. Consensus estimates include unscaled confidence (ie, all study confidence levels set at 100) and confidence scaling (with confidence scores given in Table 2).

Figure 4. Prior and posterior distributions of the true population size $\theta$ for (A) female sex workers (FSW) in location B, (B) men who have sex with men (MSM) in location B, (C) FSW in location C, and (D) MSM in location C. The prior is shown in red, the posterior with unscaled confidence (ie, all study confidence levels set at 100) is shown in green, and the posterior with confidence scaling is shown in blue.
Discussion

Principal Findings

In this study, we presented a Bayesian hierarchical model for triangulating estimates of public health interest, with our examples focusing on PSE. It is a major challenge for stakeholders to come up with a single population size estimate for a given population informed by multiple, potentially contradictory, estimates. Approaches used in the past could be overly simplistic, such as the use of a simple median, or opaque, with stakeholders picking a number after a discussion with no record of their full reasoning.

The model we have proposed assists and formalizes the triangulation process. It makes the decision for a consensus estimate from stakeholders transparent and explicit, while allowing them to retain control over how credible they find each individual estimate. Furthermore, the tool we developed to implement this model has a built-in assessment of the consensus process in the form of unaccounted-for variation; thus, the Shiny app can provide instant feedback on how successfully the individual estimates were combined in consensus. For users working with population size estimates, we recommend rounding results when presenting them to stakeholders to avoid any confusion stemming from noninteger counts of people. For all estimates, we strongly encourage transparency by presenting confidence scores with a rationale for each.

We presented a use case that represents a common challenge: the need to synthesize multiple estimates, often derived from methods of varying quality. In our example, we had the benefit of all empirical PSE and knowing the limitations and errors made during survey implementation. This provided a valuable context for confidence scoring, which affects the consensus estimate. When this information on implementation is not available, the next best option may be to base confidence scores on the strength of the method used to produce each estimate. For example, we presented the PSE derived from events, and all were smaller than the survey sample size. We did not need to know about the implementation errors to know these were nonsensical results that merited very low confidence scores. The PSE were all produced from the same survey; therefore, there were no differences in the age of our PSE to impact confidence scores (eg, we would have low confidence in older PSE and high confidence in the current PSE). As described earlier, our priors were several years old and based on stakeholder consensus. We were not present for the discussions years before our survey and do not know whether this consensus effort was influenced by a few with strong voices or more representative of the entire group of experts. Analysis of our data presented in Figures 3 and 4 suggests that the more disparate our individual estimates were, the more influential the prior was in the consensus estimate. Therefore, we encourage users to gather as much information as possible on the quality of the prior and how it was derived to provide context for the consensus estimate.

Strengths and Limitations

Other Bayesian approaches such as the Anchored Multiplier model [17,18] have been proposed for PSE. Our model differs from the Anchored Multiplier (and its Variance-Adjusted counterpart) in 4 ways. First, our model can directly handle any estimated quantity, including both population proportions and absolute sizes from any PSE technique, whereas the Anchored Multiplier is designed for estimates of population proportions and requires the selection of denominators to handle absolute sizes. Second, the Anchored Multiplier uses a binomial distribution as the sampling distribution and a beta distribution as the prior distribution, whereas our model uses normal distributions for both the sampling and prior distributions, with transformations tailoring the model to the estimate type. The effect of this difference in model structure is explored in detail in Multimedia Appendix 1 [18]; the differing structures give the models strengths in different cases, and as such, they represent complementary rather than competing approaches. Third, in the variance-adjusted case, the Anchored Multiplier uses a frequentist estimator of $\tau^2$ to extend the CIs of the estimates, whereas in our model, $\tau$ is a parameter that is fit during the sampling process. This distinction is important, as the DerSimonian-Laird estimator [25] used by the Anchored Multiplier has the potential to underestimate the true value of tau squared [26] and produces confidence (or credible) intervals that are overly narrow, particularly when the number of estimates is small or the true between-study variance is large [27], both of which are frequently encountered when dealing with PSE. A wholly Bayesian approach, such as the one offered by the Triangulator, may better account for between-study variance [27]. Fourth, and most critically, our model provides users with a clear and transparent platform to input their confidence for each estimate, which is displayed alongside the estimates data and consensus results on the final tab of the Shiny app. This is in contrast with the Anchored Multiplier and other techniques, where any adjustments to the uncertainty surrounding the estimates must be made before entering the data, which may lead to confusion among observers regarding the source of the data.

Results from the Triangulator are subject to several user limitations. First, if a user lacks knowledge of the quality of the estimates to be synthesized by the Triangulator, the confidence scores may not reflect the actual quality. High scores given to poor-quality data in the presence of other confidence-scaled estimates may result in a biased consensus estimate, as can low scores given to high-quality data. Next, users may try to adjust the confidence scores and model parameters after running the models to fit a specific, desired consensus estimate. Whether this affects the Triangulator more than similar models is debatable; the Triangulator, the data, confidence scores, and prior information are all transparently presented in the app, whereas with other methods, the inputs may be obfuscated. We provided tutorials for the proper use of the Triangulator to avoid spurious results by making such adjustments.

Conclusions

As a user-friendly app, the Triangulator has broad utility for statisticians, epidemiologists, and other public health officials—anyone seeking to combine multiple estimates of population size, incidence, prevalence, or other quantities into a single consensus estimate. It offers a solution to the
long-standing problem of synthesizing multiple estimates, potentially leading to more informed and evidence-based decision-making processes. Single-point estimates are widely used in decision-making, resource allocation, and policy development. Ministries of health rely on them to meet global health reporting requirements. Humanitarian organizations and multilateral government donors use single-point estimates to establish targets and assess their performance against those targets. The Triangulator leverages the user’s knowledge of the quality of the original estimates as well as prior knowledge of the quantity being estimated. Although technical parameters can be adjusted to meet the needs of certain users, for most use cases, no additional information is required outside of the estimates to be combined (and their uncertainty), confidence scores for each estimate, and parameters to set the prior on the quantity of interest. The functionality increases accessibility to public health teams that may not have statistical support but need to synthesize multiple estimates into a single estimate with uncertainty bounds. The Triangulator has the flexibility to be adapted and used in various other contexts and regions to address similar challenges.

Software, including the Shiny web app source code, is freely available and can be accessed on the internet [28]. The Shiny app is hosted at Epi apps for convenient use [29].

Acknowledgments
The authors would like to thank Dr. Ray Shiraishi (Division of Global Health & TB, Centers for Disease Control) for his comments and contributions to the model design and analysis in this study. This project has been supported by the President’s Emergency Plan for AIDS Relief through the Centers for Disease Control and Prevention. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the funding agencies.

Data Availability
All data generated or analyzed during this study are included in this published article.

Authors’ Contributions
IEF conceived of the model, CC and IEF designed the model and performed technical analysis, AFM, as a survey coinvestigator, conducted the original population size estimation analysis and assigned confidence scores included in the model and CC wrote and revised the manuscript with input from IEF and AFM. All authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Derivations of the model equations, a sensitivity analysis of the posterior estimate of tau, and a comparison of the model presented in this manuscript to an existing model in the literature.

References
survey and stakeholder consensus. JMIR Public Health Surveill 2018 Aug 07;4(3):e10188 [FREE full text] [doi: 10.2196/10188] [Medline: 30087089]


Leveraging Routinely Collected Program Data to Inform Extrapolated Size Estimates for Key Populations in Namibia: Small Area Estimation Study

Talia Loeb¹, MHS; Kalai Willis¹, MSPH; Frans Velishavo², BBA; Daniel Lee³, MPH; Amrita Rao¹, PhD; Stefan Baral¹, MD; Katherine Rucinski⁴, PhD

¹Data for Implementation (Data.FI), Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
²IntraHealth Namibia, Windhoek, Namibia
³United States Agency for International Development Dominican Republic, Santo Domingo, Dominican Republic
⁴Data for Implementation (Data.FI), Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Talia Loeb, MHS
Data for Implementation (Data.FI)
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street
Baltimore, MD, 21205
United States
Phone: 1 410 955 3227
Email: tloeb2@jh.edu

Abstract

Background: Estimating the size of key populations, including female sex workers (FSW) and men who have sex with men (MSM), can inform planning and resource allocation for HIV programs at local and national levels. In geographic areas where direct population size estimates (PSEs) for key populations have not been collected, small area estimation (SAE) can help fill in gaps using supplemental data sources known as auxiliary data. However, routinely collected program data have not historically been used as auxiliary data to generate subnational estimates for key populations, including in Namibia.

Objective: To systematically generate regional size estimates for FSW and MSM in Namibia, we used a consensus-informed estimation approach with local stakeholders that included the integration of routinely collected HIV program data provided by key populations’ HIV service providers.

Methods: We used quarterly program data reported by key population implementing partners, including counts of the number of individuals accessing HIV services over time, to weight existing PSEs collected through bio-behavioral surveys using a Bayesian triangulation approach. SAEs were generated through simple imputation, stratified imputation, and multivariable Poisson regression models. We selected final estimates using an iterative qualitative ranking process with local key population implementing partners.

Results: Extrapolated national estimates for FSW ranged from 4777 to 13,148 across Namibia, comprising 1.5% to 3.6% of female individuals aged between 15 and 49 years. For MSM, estimates ranged from 4611 to 10,171, comprising 0.7% to 1.5% of male individuals aged between 15 and 49 years. After the inclusion of program data as priors, the estimated proportion of FSW derived from simple imputation increased from 1.9% to 2.8%, and the proportion of MSM decreased from 1.5% to 0.75%. When stratified imputation was implemented using HIV prevalence to inform strata, the inclusion of program data increased the proportion of FSW from 2.6% to 4.0% in regions with high prevalence and decreased the proportion from 1.4% to 1.2% in regions with low prevalence. When population density was used to inform strata, the inclusion of program data also increased the proportion of FSW in high-density regions (from 1.1% to 3.4%) and decreased the proportion of MSM in all regions.

Conclusions: Using SAE approaches, we combined epidemiologic and program data to generate subnational size estimates for key populations in Namibia. Overall, estimates were highly sensitive to the inclusion of program data. Program data represent a supplemental source of information that can be used to align PSEs with real-world HIV programs, particularly in regions where population-based data collection methods are challenging to implement. Future work is needed to determine how best to include...
and validate program data in target settings and in key population size estimation studies, ultimately bridging research with practice to support a more comprehensive HIV response.

*(JMIR Public Health Surveill 2024;10:e48963) doi:10.2196/48963*

**KEYWORDS**

female sex workers; HIV; key populations; men who have sex with men; Namibia; population size estimation; small area estimation

**Introduction**

There have been significant declines in HIV incidence in countries across southern Africa over the last 10 years, including in Namibia [1]. This progress largely reflects the expanded availability of comprehensive HIV prevention and treatment, including the rollout of pre-exposure prophylaxis and increased access to antiretrovirals for people living with HIV. In Namibia, approximately 95% of those living with HIV are aware of their status, 95% of those aware of their status are on treatment, and 92% of those on treatment are virally suppressed [2]. Despite these investments and progress toward the Joint United Nations Programme on HIV/AIDS (UNAIDS) 95-95-95 targets, key populations in Namibia, such as female sex workers (FSW) and gay, bisexual, and other men who have sex with men (MSM), continue to experience a disproportionate burden of HIV compared to other adults of reproductive age [3], with an estimated prevalence of 20.9% and 8.4%, respectively. Viral suppression among individuals living with HIV ranges between 31.2% and 75.5% among FSW and between 55.8% and 76.1% among MSM, with significant variation across geographic regions [2].

Key populations in Namibia face individual, network, and structural barriers to HIV prevention and treatment that increase population-level risks for HIV acquisition and transmission. Sex work is criminalized in Namibia [4], limiting the extent to which FSW are able to safely access sexual and reproductive services without the fear of stigma or penalization by law enforcement. Power imbalances within sexual relationships, high rates of sexual and physical violence, and a lack of social protections for sex workers can compromise sexual autonomy and increase the risk of HIV [5,6]. For MSM, laws that criminalize same-sex relationships between men and other intersectional stigmas have been found to increase depression and impact stress-response behaviors in Namibia, such as substance use, increasing the risk of acquiring HIV [7]. Further, enacted stigmas in clinical settings have been found to limit access to comprehensive health services for MSM, including prevention and treatment for HIV and other sexually transmitted infections [2,8-10].

Differrentiated HIV services that are responsive to the needs of key populations can improve the overall quality of care and increase the uptake of HIV testing and treatment, but data are needed to guide the rollout and scale-up of these programs in Namibia [11]. Studies to determine population size for FSW and MSM in Namibia have historically been conducted in mostly urban settings, where community-led HIV programs are well-established. However, due to the high cost of incorporating less densely populated areas into empiric data collection efforts, population size estimates (PSEs) for key populations are lacking for the majority of the country.

Determining the number or size of key populations in Namibia may help identify existing gaps in programming and guide the prioritization of interventions aimed at addressing the transmission, prevention, and treatment of HIV [12]. Methodological approaches that use existing PSEs along with auxiliary data sources to determine the size of key populations are well-established [13-15]. However, these small area estimation (SAE) approaches typically exclude the use of program data as key inputs. Additionally, key population PSEs are often generated using multiple methods, and thus numerous and competing estimates may be produced for the same district or region [16,17]. While HIV programs frequently provide services in the same regions where size estimates are available and thus collect substantial data on the number and type of key populations accessing services, program data have not been used to systematically inform the triangulation of these size estimates [18]. By comparison, sentinel surveillance data and national HIV service delivery program data are regularly combined to inform estimates for the “general population,” but this approach is not currently used for key population-related estimation.

In this study, we used a consensus-informed approach with local stakeholders alongside routinely collected HIV program data provided by key population implementing partners to systematically generate subnational size estimates and demonstrate the use of the inclusion of program data for FSW and MSM in Namibia.

**Methods**

**Direct Estimates and Sources of Auxiliary Data**

Direct size estimates were generated using empiric data collection methods between 2012 and 2014 through Integrated Bio-Behavioral Surveillance Studies (IBBSS) in Zambesi, Ohangwena, Erongo, and Khomas for FSW [19] and in Karas, Oshana, Erongo, and Khomas for MSM (Table S1 in Multimedia Appendix 1) [20]. In the IBBSS, direct estimates were derived using multiple methods of size estimation, including mapping, key informant interviews, unique object identifiers, wisdom of the crowds, literature review, and stakeholder consensus [16,19]. These methods have been described previously [16,19]. Respondent-driven sampling was used to recruit all participants. Briefly, a set of 6-9 seeds from 4 study sites were selected, and each seed used coupons to refer others into the study. Recruitment continued until the desired sample size was reached [19,20]. Eligible FSW were aged 18 years or older and had received monetary payment for sex in the previous 6 months. Eligible MSM were aged 18 years or older and reported...
engaging in sex with other men in the previous 6 months. Both FSW and MSM were required to have resided in the study region for the last 6 months.

Auxiliary data to inform extrapolated estimates were identified in consultation with implementing partners and key stakeholders of key population programs in Namibia and were accessed through the Data.FI Consortium, a 5-year collaboration funded by the President’s Emergency Fund for AIDS Relief (PEPFAR) through the United States Agency for International Development (USAID) [21]. Specific data types included census data and geographically disaggregated population-level data such as regional literacy rates and HIV prevalence estimates. Auxiliary data sources comprised the Namibia 2011 Census [22], Namibia Population Projections calculated by the Namibia Statistics Agency [23], and the Namibia Population-Based HIV Impact Assessment (NAMPHIA) [24]. Census data and population projections were publicly available, while NAMPHIA data were obtained by request from the Centers for Disease Control and Prevention. Program data sources included indicators collected through PEPFAR’s key populations Data for Accountability, Transparency, and Impact Monitoring (DATIM) efforts [25] as well as indicators reported by the partner-led Key Populations—Strengthening Technical Assistance and Response for sustainable HIV prevention and treatment (KP-STAR) program [26].

Population denominators for direct estimates were extracted from Namibia’s Census Population Projections for the years in which data were collected in order to transform size estimates into standardized proportions. For FSW, this included regional estimates of cisgender women aged between 18 and 49 years; for MSM, this comprised adult cisgender men aged between 18 and 49 years. Age bands for each population were selected to align denominators with the age composition of the direct size estimates (numerators) collected through the initial IBBSS respondent-driven sampling survey.

Auxiliary variables identified for potential inclusion in SAE approaches included population density, employment rates, literacy proportions, and HIV prevalence. Population density, literacy, and employment rates were abstracted from the publicly available Namibia 2011 Census, while HIV prevalence estimates came from NAMPHIA (Tables S2 and S3 in Multimedia Appendix 1 provide comparisons of these characteristics across regions, stratified by sex). Variables were selected for inclusion in analyses based on their availability across on- and off-sample regions and through previously identified associations with key population size in the literature [13,14,27]. Population density and HIV prevalence were ultimately selected, given the variation in these data across regions where direct estimates were available. Program data used to inform priors included counts of the number of individuals who accessed services, reported through both DATIM and KP-STAR.

**Use of Program Data to Inform Imputations**

Each method of direct size estimation produces a single estimate. Given the variation between available direct estimates within regions, we used a consensus estimation approach and a publicly available tool (ie, the consensus estimator tool) to generate a single estimate for each region where direct estimates were collected [28]. This allowed for the weighting of the IBBSS-collected direct estimates both by precision and by quality of study implementation. The methods of direct size estimation were ranked by level of confidence, with successive sampling population size estimation (SS-PSE) having the highest confidence, followed by unique object multiplier, mapping, stakeholder consensus, literature review, key informant interview, and lastly, wisdom of the crowds. This initial ranking scheme was based on the rigor and quality of methods, guided by the World Health Organization’s synthesis of these direct size estimation methods and investigator experience implementing these approaches [16,29-31].

Prior beliefs for the distribution of the derived size estimates were defined based on program data. Quarterly data between fiscal year (FY) 2020 quarter 3 and FY 2021 quarter 2 from KP-STAR and between FY 2016 quarter 4 and FY 2021 quarter 1 from DATIM were used. Program data included the number of FSW or MSM that were reached by programs during each quarter. A 2-tailed, 2-sample t test was performed to identify any differences in the underlying distribution between the KP-STAR and PEPFAR data. The data from PEPFAR and KP-STAR were pooled due to similarity in the underlying distribution, and the minimum, maximum, and median numbers were calculated, in addition to the SD for the pooled numbers from 2016 quarter 4 and 2021 quarter 2. These 4 values, along with the assumed log-normal distribution for population sizes, served as the prior beliefs for the consensus estimator tool in order to derive a single direct estimate for each region. Of note, the data availability limited the accessibility of consistently reported direct estimates across regions, such as missing direct estimates collected through SS-PSE for MSM in Karas and Oshana. The USAID team implemented routine data quality assessments of data from DATIM and KP-STAR. This involved a random sample review of data captured in client files in comparison to the implementing partner’s client-level electronic database. A similar comparison was routinely collected at the aggregate level, comparing site-level monthly reports to the quarterly data.

Sensitivity analyses were completed to determine the extent to which adjusting the numerical level of confidence but retaining the rank order affected the weighting scheme and the resulting combined direct estimates. Multiple permutations and values for the confidence levels were assigned to assess the sensitivity of the rankings across the resulting triangulated estimates. The resulting size estimates generated from the consensus estimator were similar across the varying confidence levels as long as rank order was preserved.

**SAE Using Imputation and Regression**

We conducted simple imputation, stratified imputation, and regression modeling to determine the regional proportion of female individuals aged between 18 and 49 years and the proportion of male individuals aged between 18 and 49 years that were FSW and cisgender MSM, respectively. The derived direct estimates from the consensus estimator tool were used in the simple imputation and stratified imputation approaches, whereas all the direct estimates were used in the regression approach to maximize the number of input data points.
For the imputation approaches, the consensus-generated direct estimates were divided by the total cisgender female population aged between 18 and 49 years for FSW and the total cisgender male population aged between 18 and 49 years for MSM to derive proportions. Using simple imputation, the mean of these proportions was applied to each of the regions where direct size estimates were not available. In the stratified imputation, HIV prevalence estimates reported by NAMPHIA were used as the stratification variable. Regions were categorized by whether their HIV prevalence fell above or below the median HIV prevalence of all the regions, as reported by NAMPHIA (17.85% among women and 7.45% among men of reproductive age). The mean of the proportions for each respective category was applied to each of the regions where direct size estimates were not available. In sensitivity analyses, simple and stratified imputations were also conducted using all the available direct size estimates to evaluate how the incorporation of program data as prior beliefs in the consensus estimator tool changed the results.

In a supplemental analysis, multivariable Poisson regression models were fit using the full set of available direct estimates to generate size estimate proportions for FSW and MSM. To ensure a maximum number of available inputs were available to fit the regression models, the consensus estimator tool was not used. Generalized estimating equations accounted for clustering by region, and candidate predictors included HIV prevalence, population density, literacy rate, employment, and annual projected population growth, all overall and stratified by sex. Model selection was determined using Akaike information criterion (AIC) and Bayesian information criterion (BIC) values, with the lowest values indicating the best fit.

The resulting estimated proportions from all approaches were applied to the projected 2021 population of female individuals for FSW and male individuals for MSM aged between 15 and 49 years, resulting in the final PSEs.

Software

Imputations were conducted using Microsoft Excel [32], multivariable Poisson regression was performed using Stata (version 16) [33], and maps were constructed using ArcGIS software [34,35].

Consensus Building With Local Stakeholders

The process of generating the extrapolated size estimates was an iterative process, requiring collaboration and discussion between researchers and government, programs, and community stakeholders in order to refine and select a final set of estimates. Meetings with implementing partners occurred on multiple occasions, with discussion around how these size estimates could be used programmatically to inform service delivery targets. A brief synopsis of these meetings has been provided in Multimedia Appendix 1.

Ethical Considerations

This study comprised a secondary analysis of deidentified surveillance data and routinely collected programmatic data; patient-level data were not shared with the study team nor included in analyses. The Johns Hopkins Bloomberg School of Public Health Institutional Review Board designated this work as not being a human subjects research study (IRB00007442).

Results

Direct PSEs

Direct PSEs collected through the IBBSS varied across direct estimation methods and by region for FSW (Table 1). The largest number of FSW was generally seen in Khomas, ranging from 100 (95% CI 50-1700) using key informant interviews to 5240 (95% CI 3373-11,706) using the unique object multiplier method. Khomas also produced the largest PSEs for MSM, ranging from 300 (95% CI 100-1600) using key informant interviews to 2416 (95% CI 850-4000) using stakeholder consensus. After incorporating program data as priors in the consensus estimator tool, the resulting generated PSEs in Khomas were 1480 (95% CI 1099-1799) for FSW and 511 (95% CI 143-958) for MSM. Direct PSEs for each region, including the resulting weighted estimates from the consensus estimator tool, are presented in Table 1.
<table>
<thead>
<tr>
<th>Direct estimation method</th>
<th>Female sex worker regions, n (95% CI)</th>
<th>Men who have sex with men regions, n (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Khomas</td>
<td>Karas</td>
</tr>
<tr>
<td></td>
<td>Erongo</td>
<td>Oshana</td>
</tr>
<tr>
<td></td>
<td>Erongo</td>
<td>Khomas</td>
</tr>
<tr>
<td>Mapping</td>
<td>246 (142-426)</td>
<td>282 (141-423)</td>
</tr>
<tr>
<td>Key informant interview</td>
<td>300 (50-4300)</td>
<td>1132 (200-2948)</td>
</tr>
<tr>
<td>Unique object multiplier</td>
<td>5299 (3500-8575)</td>
<td>1714 (2379-5632)</td>
</tr>
<tr>
<td>Wisdom of the crowds</td>
<td>300 (100-1000)</td>
<td>100 (50-1000)</td>
</tr>
<tr>
<td>Literature review</td>
<td>84 (47-251)</td>
<td>84 (24-138)</td>
</tr>
<tr>
<td>Stakeholder consensus</td>
<td>800 (380-2000)</td>
<td>500 (300-650)</td>
</tr>
<tr>
<td>SS-PSE</td>
<td>674 (318-2426)</td>
<td>2196 (1651-2382)</td>
</tr>
<tr>
<td>Consensus estimator tool</td>
<td>208 (88-420)</td>
<td>1480 (1099-1779)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Female sex worker regions, n (95% CI)</th>
<th>Khomas</th>
<th>Erongo</th>
<th>Khomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zambezi</td>
<td>158</td>
<td>322</td>
<td>528</td>
</tr>
<tr>
<td>Erongo</td>
<td>300</td>
<td>100</td>
<td>528</td>
</tr>
<tr>
<td>Oshana</td>
<td>100</td>
<td>330</td>
<td>1132</td>
</tr>
<tr>
<td>Karas</td>
<td>5240</td>
<td>1714</td>
<td>1714</td>
</tr>
<tr>
<td>Unique object multiplier</td>
<td>400</td>
<td>70</td>
<td>150</td>
</tr>
<tr>
<td>Wisdom of the crowds</td>
<td>85</td>
<td>241</td>
<td>84</td>
</tr>
<tr>
<td>Literature review</td>
<td>47-251</td>
<td>134-723</td>
<td>1582</td>
</tr>
<tr>
<td>Stakeholder consensus</td>
<td>900</td>
<td>900</td>
<td>3000</td>
</tr>
<tr>
<td>SS-PSE</td>
<td>900</td>
<td>1057</td>
<td>2196</td>
</tr>
<tr>
<td>Consensus estimator tool</td>
<td>302</td>
<td>387</td>
<td>1480</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Men who have sex with men regions, n (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karas</td>
</tr>
<tr>
<td>Oshana</td>
</tr>
<tr>
<td>Erongo</td>
</tr>
<tr>
<td>Khomas</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Karas</th>
<th>Oshana</th>
<th>Erongo</th>
<th>Khomas</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>488</td>
<td>460</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>100</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>250-5184</td>
<td>(70-300)</td>
<td>(100-1600)</td>
<td></td>
</tr>
<tr>
<td>1249-1822</td>
<td>(2379-5632)</td>
<td>(2013-5808)</td>
<td></td>
</tr>
<tr>
<td>1699-3240</td>
<td>(122-701)</td>
<td>(345-1983)</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>40</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>70</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>40-250</td>
<td>(no bounds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>427</td>
<td>1207</td>
<td>427</td>
<td></td>
</tr>
<tr>
<td>2416</td>
<td></td>
<td>2416</td>
<td></td>
</tr>
<tr>
<td>900</td>
<td>2416</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>610</td>
<td>2416</td>
<td>610</td>
<td></td>
</tr>
<tr>
<td>475-658</td>
<td>(850-4000)</td>
<td>(850-4000)</td>
<td></td>
</tr>
<tr>
<td>670</td>
<td>2416</td>
<td>670</td>
<td></td>
</tr>
<tr>
<td>410-1610</td>
<td>(850-4000)</td>
<td>(850-4000)</td>
<td></td>
</tr>
<tr>
<td>511</td>
<td></td>
<td>511</td>
<td></td>
</tr>
</tbody>
</table>

**Indirect Estimates**

Extrapolated national estimates for FSW ranged from 4777 (stratified imputation by HIV prevalence) to 13,148 (stratified imputation by population density, preprogrammatic data). For MSM, estimates ranged from 4611 (simple imputation and stratified imputation by HIV prevalence) to 10,171 (stratified imputation by population density). Regional estimates from the imputation approaches are reported in Tables 2 and 3 for FSW and Tables 4 and 5 for MSM. Regional estimates from multivariable Poisson regression models are included in Table S4 in Multimedia Appendix 1.

When program data were incorporated as priors in the imputation models, national extrapolated estimates ranged from 4777 (stratified imputation by HIV prevalence) to 9288 (simple imputation) for FSW. For MSM, these estimates ranged from 2236 (stratified imputation by population density) to 2372 (simple imputation). Regional estimates resulting from the inclusion of programmatic priors are presented by population for each approach in Tables 2-5 and are presented visually in Figures 1 and 2.

---

[a] Not available.
### Table 2. Absolute estimate differences in female sex workers (FSW) extrapolated estimates for regions in Namibia before and after the integration of program data. Results from 3 different methods of imputation are presented: simple imputation, stratified imputation (by HIV prevalence), and stratified imputation (by population density).

<table>
<thead>
<tr>
<th>Region</th>
<th>Simple imputation</th>
<th>Stratified imputation, HIV prevalence</th>
<th>Stratified imputation, population density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preprogram data proportion</td>
<td>Postprogram data proportion</td>
<td>Preprogram data proportion</td>
</tr>
<tr>
<td>Zambezi a</td>
<td>175</td>
<td>256</td>
<td>231</td>
</tr>
<tr>
<td>Ohangwena a</td>
<td>129</td>
<td>189</td>
<td>171</td>
</tr>
<tr>
<td>Erongo a</td>
<td>985</td>
<td>1441</td>
<td>695</td>
</tr>
<tr>
<td>Khomas a</td>
<td>2751</td>
<td>4022</td>
<td>1941</td>
</tr>
<tr>
<td>Hardap</td>
<td>284</td>
<td>415</td>
<td>200</td>
</tr>
<tr>
<td>Karas</td>
<td>278</td>
<td>406</td>
<td>196</td>
</tr>
<tr>
<td>Kavango</td>
<td>368</td>
<td>538</td>
<td>260</td>
</tr>
<tr>
<td>Kunene</td>
<td>140</td>
<td>204</td>
<td>99</td>
</tr>
<tr>
<td>Omaheke</td>
<td>103</td>
<td>151</td>
<td>73</td>
</tr>
<tr>
<td>Omusati</td>
<td>74</td>
<td>107</td>
<td>97</td>
</tr>
<tr>
<td>Oshana</td>
<td>512</td>
<td>749</td>
<td>361</td>
</tr>
<tr>
<td>Oshikoto</td>
<td>125</td>
<td>183</td>
<td>165</td>
</tr>
<tr>
<td>Otjozondjupa</td>
<td>428</td>
<td>626</td>
<td>302</td>
</tr>
</tbody>
</table>

*Indicates regions in which direct estimates were available.

### Table 3. Proportion estimate differences in female sex workers (FSW) extrapolated estimates for regions in Namibia before and after the integration of program data. Results from 3 different methods of imputation are presented: simple imputation, stratified imputation (by HIV prevalence), and stratified imputation (by population density).

<table>
<thead>
<tr>
<th>Stratified imputation</th>
<th>Simple imputation</th>
<th>Stratified imputation, HIV prevalence</th>
<th>Stratified imputation, population density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprogram data proportion</td>
<td>Postprogram data proportion</td>
<td>Preprogram data proportion</td>
<td>Postprogram data proportion</td>
</tr>
<tr>
<td>Greater than median</td>
<td>0.019</td>
<td>0.028</td>
<td>0.026</td>
</tr>
<tr>
<td>Less than or equal to median</td>
<td>0.019</td>
<td>0.028</td>
<td>0.014</td>
</tr>
</tbody>
</table>
Table 4. Absolute estimate differences in men who have sex with men (MSM) extrapolated estimates for regions in Namibia before and after the integration of program data. Results from 3 different methods of imputation are presented: simple imputation, stratified imputation (by HIV prevalence), and stratified imputation (by population density).

<table>
<thead>
<tr>
<th>Region</th>
<th>Simple imputation</th>
<th>Stratified imputation, HIV prevalence</th>
<th>Stratified imputation, population density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preprogram data</td>
<td>Postprogram data</td>
<td>Preprogram data</td>
</tr>
<tr>
<td></td>
<td>estimate</td>
<td>estimate</td>
<td>estimate</td>
</tr>
<tr>
<td>Zambezi</td>
<td>114</td>
<td>59</td>
<td>165</td>
</tr>
<tr>
<td>Ohangwena</td>
<td>81</td>
<td>42</td>
<td>117</td>
</tr>
<tr>
<td>Erongo(^a)</td>
<td>854</td>
<td>439</td>
<td>741</td>
</tr>
<tr>
<td>Khomas(^a)</td>
<td>1991</td>
<td>1024</td>
<td>1728</td>
</tr>
<tr>
<td>Hardap</td>
<td>215</td>
<td>110</td>
<td>186</td>
</tr>
<tr>
<td>Karas(^b)</td>
<td>196</td>
<td>100</td>
<td>282</td>
</tr>
<tr>
<td>Kavango</td>
<td>215</td>
<td>111</td>
<td>310</td>
</tr>
<tr>
<td>Kunene</td>
<td>96</td>
<td>49</td>
<td>84</td>
</tr>
<tr>
<td>Omaheke</td>
<td>78</td>
<td>40</td>
<td>67</td>
</tr>
<tr>
<td>Omusati</td>
<td>43</td>
<td>22</td>
<td>62</td>
</tr>
<tr>
<td>Oshana(^a)</td>
<td>313</td>
<td>161</td>
<td>452</td>
</tr>
<tr>
<td>Oshikoto</td>
<td>96</td>
<td>49</td>
<td>139</td>
</tr>
<tr>
<td>Otjozondjupa</td>
<td>313</td>
<td>161</td>
<td>272</td>
</tr>
</tbody>
</table>

\(^a\)Indicates regions in which direct estimates were available.

Table 5. Proportion estimate differences in men who have sex with men (MSM) extrapolated estimates for regions in Namibia before and after the integration of program data. Results from 3 different methods of imputation are presented: simple imputation, stratified imputation (by HIV prevalence), and stratified imputation (by population density).

<table>
<thead>
<tr>
<th>Stratified imputation</th>
<th>Simple imputation</th>
<th>Stratified imputation, HIV prevalence</th>
<th>Stratified imputation, population density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preprogram data</td>
<td>Postprogram data</td>
<td>Preprogram data</td>
</tr>
<tr>
<td></td>
<td>proportion</td>
<td>proportion</td>
<td>proportion</td>
</tr>
<tr>
<td>Greater than median</td>
<td>0.015</td>
<td>0.0075</td>
<td>0.021</td>
</tr>
<tr>
<td>Less than or equal to median</td>
<td>0.015</td>
<td>0.0075</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Figure 1. Map representing 2021 extrapolated estimates of the number of female sex workers (FSWs) across regions in Namibia. The darker gradations represent a higher number of FSWs in the region. The number of FSWs was calculated through 3 methods of imputation: simple imputation, stratified imputation (by HIV prevalence), and stratified imputation (by population density). Direct estimates used for the imputation were derived from a Bayesian consensus estimation approach incorporating routinely collected program data as priors.
Change in Imputed Estimates Following Inclusion of Program Data

After including program data as priors in the simple imputation model for FSW, the estimated proportion of female individuals aged between 15 and 49 years who were FSW increased from 1.9% to 2.8% (Table 3). Among male individuals aged between 15 and 49 years, the proportion of MSM decreased from 1.5% to 0.75% (Table 5).

In models where HIV prevalence informed stratified imputations, the estimated proportion of women who were FSW increased from 2.6% to 4.0% in regions where HIV prevalence was greater than the median; in regions with less than median HIV prevalence, the estimated proportion decreased from 1.4% to 1.2%. Among MSM, the proportion decreased from 2.1% to 0.8% in regions where the HIV prevalence was greater than the median and decreased from 1.3% to 0.7% in regions where the HIV prevalence was less than the median.

In models where population density informed strata, the estimated proportion of women who were FSW increased from 1.1% to 3.4% in regions with a greater than median population density and decreased from 2.8% to 1.0% in regions with less than median population density. Among MSM, the proportion decreased from 1.4% to 0.45% in regions where the population density was greater than the median and decreased from 1.5% to 1.0% in regions where the population density was less than the median.

Discussion

In this study of key population PSEs in Namibia, we calculated national and regional size estimates for FSW and MSM by incorporating routinely collected program data into established methodological approaches for SAE. These estimates were further refined by engaging implementing partners in consensus-building meetings throughout the estimation process. An iterative qualitative ranking process was used to ground truth estimates with partners. The resulting indirect size estimates varied across SAE approaches, and final estimates were calibrated to population-level data, stakeholder knowledge, and data routinely collected through HIV prevention and treatment programs in Namibia.

The inclusion of program data as previous knowledge increased the estimated proportion of reproductive-age women that were sex workers in regions with high HIV prevalence, suggesting that network-based surveys may have underestimated the true size of the sex worker population in Namibia [36]. Methods to collect direct PSEs vary in rigor and quality, but the quality of implementation and reporting of these estimates can be highly variable irrespective of method, making estimates difficult to compare [30]. Existing direct estimates for FSW from the IBBSS studies included in this analysis were also highly uneven across PSE approaches, necessitating a systematic approach to combine estimates into an interpretable and usable estimate for each region [18,36,37]. This approach may be relevant to similar settings in which direct size estimates are derived from multiple epidemiologic studies and necessitate a formalized approach for consolidation.

Historically, key population program data have not been included in size estimation exercises, given concerns around external validity. While these data are imperfect, in the case of SAEs, the number of sex workers accessing services through established key population programs can potentially serve as a lower bound for these estimates in a given region. This approach demonstrates the potential of including program data in SAE efforts to better align PSEs with real-world data in countries where surveillance efforts may be limited [14,18].

Indirect estimates for MSM were also highly sensitive to the inclusion of programmatic priors, producing smaller than expected population proportions for all regions. Despite efforts in Namibia to reach MSM with targeted HIV prevention and treatment services, including investments in peer navigators and community health workers [38], MSM remain largely marginalized from existing HIV programs relative to FSW. Provision of services to MSM remains a challenge amid Namibia’s legal framework, and as a result, the majority of MSM living with HIV are thought to be unaware of their HIV status. This is particularly true for younger MSM, who are significantly less likely to access HIV testing services relative to older MSM despite a high concentration of risk during
adolescence and early adulthood [39]. While age-oriented and culturally relevant HIV testing programs may increase the uptake of HIV services for young MSM in Namibia, in the absence of larger structural interventions that reduce stigmas and discriminatory policies, these innovations may result in only modest progress in reducing the burden of HIV among MSM [40]. As size estimation is often used for allocating funding and setting targets, methods that harmonize direct PSEs with real-world data in the context of these structural barriers are needed to develop realistic targets for epidemic control. Inclusion of program data in indirect size estimation methods such as SAE could potentially allow programs to implement more informed targets for the delivery of HIV prevention and treatment services, which is ultimately one of the key purposes of population size estimation studies.

These analyses were also grounded in a systematic consensus-building approach, bridging a gap in the coordination and integration of size estimation approaches between researchers and implementing partners [41]. Meetings with key stakeholders were highly influential in the development of indirect estimates, as the limited number of data inputs challenged the precise selection of a universal best-fitting model for each population. Results for up to 3 models were presented to stakeholders to ground-truth potential estimates and serve as an additional appraisal tool. Historically, consensus-building approaches have been used to guide the triangulation of direct PSEs resulting from multiple data collection methods [42-44]. Collaborative efforts that extend beyond the initial data collection processes are seemingly less common. While ideal estimates may vary for different stakeholders, particularly when estimates are used for program target setting, this approach may provide a way forward for ensuring that size estimates are ultimately used to guide practice, programming, and resource allocation.

There are several notable strengths in this study. First, we demonstrated the potential of program data as an underused data source to understand the size of key populations in Namibia, thus opening the possibility of routinizing this approach for size estimation activities in other settings. Second, we systematically engaged with external partners to understand the cultural and on-the-ground implementation context for SAE, which provided a robust understanding of the results and interpretation. Lastly, there was a wide availability of various program and auxiliary data sources spanning multiple regions and years. This enabled the use of longitudinal data to generate prior beliefs regarding the distribution of FSW and MSM in this specific context and to use characteristics specific to each region in order to predict size estimates, which has not been done previously in Namibia.

This study also had several limitations. Despite widely available auxiliary data, there was a lack of available direct size estimates for key populations in Namibia, and those that were available were limited to more urban regions and were from the mid-2010s. As these estimates informed the primary inputs for our imputation approaches and regression modeling, it is possible that extrapolated size estimates were upwardly biased in more rural settings or in communities where outmigration is common [45]. Furthermore, the direct estimates included in analyses may have been sensitive to other unmeasured characteristics, which could have subsequently impacted the size and direction of the resulting indirect estimates. Similarly, we were only able to use programmatic data from regions where size estimates had also been collected to inform our models, as these data informed triangulation of the available direct estimates. While these data may have improved the validity of extrapolated size estimates in regions where direct estimates were available, it is possible that their inclusion introduced additional biases in regions where these data were not available [15]. Moreover, there was limited regional heterogeneity in the availability of direct estimates and program data, resulting in a lack of nuanced indirect estimates in more rural areas.

In this study of indirect PSEs in Namibia, data routinely collected by key population implementing partners were incorporated as prior information into imputation models to produce updated regional and national FSW and MSM estimates. The findings provide a potential roadmap for the systematic integration of program data in generating HIV estimates for key populations and align size estimation efforts with best practices by UNAIDS and others that rely heavily on both sentinel surveillance and national service delivery data to inform general population HIV projections. Further, as PEPFAR and other international donors move toward a model of sustainable HIV programming in sub-Saharan Africa, integrating routinely collected data into SAE approaches can support a more precise calculation of costs and allow for an expanded evidence base for supporting targeted programmatic scale up. Finally, as funding and resources for large epidemiologic studies become increasingly scarce, harnessing underused data sources, such as routinely collected programmatic data, will be critical for improving the validity of PSEs and meeting the needs of key populations in Namibia and elsewhere.

Acknowledgments

The authors would like to thank all participants who generously donated their time to the data collection efforts that informed the present analysis. Additionally, the authors would like to acknowledge the important role of community groups and implementing partners in meeting the needs of those most marginalized in the HIV response. SB and AR were supported by the National Institute of Allergy and Infectious Diseases (R01AI170249). KR was supported by the National Institute of Mental Health (K01MH129226).

This manuscript was produced for review by the US President’s Emergency Plan for AIDS Relief through the United States Agency for International Development. It was prepared by Data for Implementation (Data.FI), a cooperative agreement funded by the US Agency for International Development (agreement 7200AA19CA0004), beginning April 15, 2019. Data for Implementation is implemented by Palladium in partnership with John Snow Research & Training Institute, Johns Hopkins University Department of Epidemiology, Right to Care, Cooper/Smith, DT Global, Jembi Health Systems, and Pendulum, and...
supported by expert local resource partners. The information provided in this study is not official US government information and does not necessarily reflect the views or positions of the US President’s Emergency Plan for AIDS Relief, the US Agency for International Development, or the US Government.

Data Availability
The data sets analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental materials include description of consensus building, example determination of confidence values, multivariable regression-derived populations size estimates by region, and regional characteristics stratified by sex.

References
33. Stata statistical software. StataCorp. College Station, TX: StataCorp LLC; 2021. URL: https://www.stata.com/ [accessed 2024-02-15]


Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
DATIM: Data for Accountability, Transparency, and Impact Monitoring
FSW: female sex workers
FY: fiscal year
IBBSS: Integrated Bio-Behavioral Surveillance Studies
KP-STAR: Key Populations—Strengthening Technical Assistance and Response for sustainable HIV prevention and treatment
MSM: men who have sex with men
NAMPHIA: Namibia Population-Based HIV Impact Assessment
PEPFAR: President’s Emergency Fund for AIDS Relief
PSE: population size estimate
SAE: small area estimation
SS-PSE: successive sampling population size estimation
UNAIDS: Joint United Nations Programme on HIV/AIDS
USAID: United States Agency for International Development

©Talia Loeb, Kalai Willis, Frans Velishavo, Daniel Lee, Amrita Rao, Stefan Baral, Katherine Rucinski. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 04.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and
Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Estimated Number of Injection-Involved Overdose Deaths in US States From 2000 to 2020: Secondary Analysis of Surveillance Data

Eric William Hall¹, MPH, PhD; Patrick Sean Sullivan², PhD, DVM; Heather Bradley², PhD

¹OHSU-PSU School of Public Health, Oregon Health and Science University, Portland, OR, United States
²Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Eric William Hall, MPH, PhD
OHSU-PSU School of Public Health
Oregon Health and Science University
1810 SW 5th Avenue
Suite 510
Portland, OR, 97201
United States
Phone: 1 503 494 4966
Email: halleri@ohsu.edu

Abstract

Background: In the United States, both drug overdose mortality and injection-involved drug overdose mortality have increased nationally over the past 25 years. Despite documented geographic differences in overdose mortality and substances implicated in overdose mortality trends, injection-involved overdose mortality has not been summarized at a subnational level.

Objective: We aimed to estimate the annual number of injection-involved overdose deaths in each US state from 2000 to 2020.

Methods: We conducted a stratified analysis that used data from drug treatment admissions (Treatment Episodes Data Set–Admissions; TEDS-A) and the National Vital Statistics System (NVSS) to estimate state-specific percentages of reported drug overdose deaths that were injection-involved from 2000 to 2020. TEDS-A collects data on the route of administration and the type of substance used upon treatment admission. We used these data to calculate the percentage of reported injections for each drug type by demographic group (race or ethnicity, sex, and age group), year, and state. Additionally, using NVSS mortality data, the annual number of overdose deaths involving selected drug types was identified by the following specific multiple-cause-of-death codes: heroin or synthetic opioids other than methadone (T40.1, T40.4), natural or semisynthetic opioids and methadone (T40.2, T40.3), cocaine (T40.5), psychostimulants with abuse potential (T43.6), sedatives (T42.3, T42.4), and others (T36-T59.0). We used the probabilities of injection with the annual number of overdose deaths, by year, primary substance, and demographic groups to estimate the number of overdose deaths that were injection-involved.

Results: In 2020, there were 91,071 overdose deaths among adults recorded in the United States, and 93.1% (84,753/91,071) occurred in the 46 jurisdictions that reported data to TEDS-A. Slightly less than half (38,253/84,753, 45.1%; 95% CI 41.1%-49.8%) of those overdose deaths were estimated to be injection-involved, translating to 38,253 (95% CI 34,839-42,181) injection-involved overdose deaths in 2020. There was large variation among states in the estimated injection-involved overdose death rate (median 14.72, range 5.45-31.77 per 100,000 people). The national injection-involved overdose death rate increased by 323% (95% CI 255%-391%) from 2010 (3.78, 95% CI 3.33-4.31) to 2020 (15.97, 95% CI 14.55-17.61). States in which the estimated injection-involved overdose death rate increased faster than the national average were disproportionately concentrated in the Northeast region.

Conclusions: Although overdose mortality and injection-involved overdose mortality have increased dramatically across the country, these trends have been more pronounced in some regions. A better understanding of state-level trends in injection-involved mortality can inform the prioritization of public health strategies that aim to reduce overdose mortality and prevent downstream consequences of injection drug use.
KEYWORDS
death rate; death; drug abuse; drugs; injection drug use; injection; mortality; National Vital Statistics System; overdose death rate; overdose; state; substance abuse; Treatment Episode Dataset-Admission; treatment

Introduction

Drug overdose mortality has increased dramatically in the United States since 1999, with over 106,000 overdose deaths occurring in 2021 alone [1,2]. Although this trend is often largely attributed to an increase in opioid-involved overdose deaths [3,4], data from toxicology reports indicate that the specific drugs and drug classes that contribute to the burden of overdose mortality have changed over time [5-8]. In addition, primary routes of drug administration have evolved, and recent work suggests much of the recent increase in overdose mortality is likely attributable to increasing injection-involved overdose deaths. From 2007 to 2018, the estimated proportion of all overdose deaths that were injection-involved more than doubled from 18.5% to 42.2% [9]. Increases in injection-involved mortality may indicate an increase in injection behavior, which highlights additional public health concerns. Compared to other routes of substance use (eg, swallowing pills or smoking), injection drug use has the highest risk of overdose and the acquisition of bloodborne infections such as HIV, viral hepatitis, and skin infections [9-13].

National trends in injection-involved overdose mortality likely mask important geographic differences at subnational levels. Previous analysis of mortality data has identified regional differences in the substances implicated in overdose mortality trends [14-16]. Additionally, analyses of infectious disease surveillance data reveal geographic differences in both the rates of acute hepatitis B virus (HBV) and hepatitis C virus (HCV) infections [17] and the trends in HCV-related death rates among persons aged 40 years or younger [18], which are all likely related to the prevalence of injection drug use [19,20]. An understanding of subnational trends in injection-involved overdose mortality has implications both for local resource allocation and the types of interventions needed to prevent overdose mortality, as well as for public health strategies to address additional adverse health outcomes related to injection behavior. For example, the availability of naloxone [21-23] and the implementation of supervised injection sites [24] have both been shown to prevent overdoses, specifically among people who inject drugs. Syringe services programs can prevent infections through the provision of sterile needles and other injection equipment [25-27]. In this study, we aimed to estimate the annual number of injection-involved overdose deaths and the injection death rate within each US state from 2000 to 2020.

Methods

Overview

Our approach used and extended a previously published method for estimating the burden of injection-involved mortality [9]. Briefly, we conducted a stratified analysis that used data from drug treatment admissions to estimate the percentage of persons injecting by reported drug type within demographic strata. Using this estimated percentage along with counts of overdose deaths from the National Vital Statistics System (NVSS), we estimated the number of persons that died from an injection-involved drug overdose death within each state and year.

Data Sources

Drug treatment admissions data were from the Treatment Episode Data Set–Admissions (TEDS-A; from 2000 to 2020) [28]. TEDS-A is a national data set that includes client-level data routinely collected from all publicly funded, and some privately funded, facilities that are licensed to provide substance abuse treatment in the United States. The national data set compiles data from individual state systems into a standardized format. Each record in TEDS-A represents an admission to a treatment program and includes demographic information, along with data on the substances used and the usual route of administration for each substance.

Data on the annual number of overdose deaths were from the NVSS detailed multiple cause of death mortality microdata files, obtained through a data request from the National Center for Health Statistics (NCHS) [29]. The NVSS microdata include a line-listing of all deaths that occurred in the United States from 2000 to 2020. Mortality data include the International Classification of Diseases, Tenth Revision (ICD-10) codes for multiple underlying causes of deaths. To align with previous definitions, overdose deaths were classified using the following ICD-10 codes: X40-X44, X60-X64, X85, and Y10-Y14 [1,30]. Additionally, overdose deaths involving selected drug types were identified by the following specific multiple-cause-of-death codes: heroin or synthetic opioids other than methadone (T40.1, T40.4), natural or semisynthetic opioids and methadone (T40.2, T40.3), cocaine (T40.5), psychostimulants with abuse potential (T43.6), sedatives (T42.3, T42.4), and others (T36-T59.0). We used National Center for Health Statistics bridged-race annual population estimates [31] for denominators in all population death rates.

Analysis

All analyses were limited to adults aged 18 years or older within 50 US states and the District of Columbia. Using TEDS-A data from 2000 to 2020, we identified all treatment admissions that included the following drugs: heroin or synthetic opioids (excluding methadone), stimulants (including methamphetamine, other amphetamines, and other stimulants), natural or semisynthetic opioids and methadone, cocaine (including crack), and sedatives (including benzodiazepines, other tranquilizers, barbiturates, and other sedatives). For each year and within each state, TEDS-A treatment admissions data were categorized into 16 strata defined by all combinations of sex (male and female), age group (18-39 years and ≥40 years), and race or ethnicity (Hispanic, non-Hispanic Black, non-Hispanic White, and non-Hispanic other). For each of the 5 drug types, within the analytic strata, we estimated the percentage of injection by calculating the percentage of admissions with data on the route of administration that reported injection as the usual route of administration. The number of treatment admissions and
percentage of reported injections for each state, year, and drug type are reported in Table S1 in Multimedia Appendix 1.

Overdose deaths from NVSS that included toxicology data were categorized into the same 5 mutually exclusive drug categories. Deaths that indicated multiple drugs were categorized based on the drug that had the highest overall percentage of persons reporting injection, as estimated from the TEDS-A treatment data [9]. This approach was used because we were interested in the percentage of any drug having been injected rather than in classifying deaths as attributable to a single drug type. Overdose deaths that did not have a specific T-code listed (ie, only listed T50.9) were distributed to the 6 categories (the 5 defined drug use categories described above and any other T-codes) based on the nonmissing distribution within each year and demographic strata [32,33]. The percentage of all overdose deaths that were categorized into each drug category for each state and year is reported in Table S2 in Multimedia Appendix 2 [1,9,29,30,32,33].

To estimate the number of injection-involved overdose deaths, we multiplied the drug-specific probabilities of injection by the respective number of reported overdose deaths within each drug type, state, year, and demographic strata. We then collapsed the results to report the estimated number of injection-involved overdose deaths for each state and year. To estimate population-level rates, we summed the bridged-race population estimates as denominators. The CIs were estimated by calculating the Clopper-Pearson confidence limits for all estimates of drug-specific probabilities of injection. If the annual number of treatment admissions that reported any of the 5 drug types was less than 50 within any state, we suppressed the results for that state and year. Additionally, if more than 15% of treatment admissions that reported a drug of interest were missing, we suppressed the results for that state and year.

To demonstrate how changes in injection-involved overdose mortality differed by state and over time, we present state-level results from 3 years: 2000, 2010, and 2020. We also calculated the relative percentage change in the national injection-involved overdose death rate from 2010 to 2020 with CIs for multiyear percentage change [34]. If a state had a suppressed rate for either 2010 or 2020, we suppressed the estimated trend. We then compared the state-level trends to the national trend during the same time period to identify states that had more or less pronounced trends.

**Ethical Considerations**

This study did not receive a review or a formal determination from an institutional review board because this analysis is not human-participant research, according to Regulation 45 CFR 46 [35].

**Results**

In 2020, 90% (46/51) of jurisdictions reported treatment data and were included in the national TEDS-A data set (states not included were Idaho, Maryland, New Mexico, Oregon, and Washington). Overall, there were 91,071 overdose deaths among adults recorded in the United States in 2020. Of those, 93.1% (84,753/91,071) occurred in the 46 jurisdictions that reported data to TEDS-A. Slightly less than half (38,253/84,753, 45.1%, 95% CI 41.1%-49.8%) of those were estimated to be injection-involved, translating to 38,253 (95% CI 34,839-42,181) injection-involved overdose deaths (Table 1). In 2020, the national injection overdose death rate was 15.97 (95% CI 14.55-17.61) per 100,000 people.
Table 1. Estimated number of injection-involved overdose deaths among adults by US state, 2020. State-level results were suppressed if the annual number of treatment admissions that reported any of the 5 drug types was <50 or if ≥15% of treatment admissions that reported a drug of interest were missing data on route of administration.

<table>
<thead>
<tr>
<th>State</th>
<th>Population, n</th>
<th>Overdose deaths, n</th>
<th>Injection-involved overdose deaths</th>
<th>Overdose deaths, n (95% CI)</th>
<th>% of overdose deaths (95% CI)</th>
<th>Rate per 100,000 (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Deaths, n (95% CI)</td>
<td>456 (407-513)</td>
<td>44.6 (39.7-50.1)</td>
<td>11.89 (10.61-13.37)</td>
</tr>
<tr>
<td>Alabama</td>
<td>3,834,249</td>
<td>1023</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaska</td>
<td>552,427</td>
<td>158</td>
<td>83 (68-100)</td>
<td>52.6 (42.8-63.3)</td>
<td>15.06 (12.25-18.11)</td>
<td></td>
</tr>
<tr>
<td>Arizona</td>
<td>5,774,978</td>
<td>2483</td>
<td>779 (539-1249)</td>
<td>31.4 (21.7-50.3)</td>
<td>13.49 (9.33-21.63)</td>
<td></td>
</tr>
<tr>
<td>Arkansas</td>
<td>2,330,808</td>
<td>540</td>
<td>248 (192-299)</td>
<td>46 (35.5-55.4)</td>
<td>10.66 (8.23-12.85)</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>30,576,844</td>
<td>8778</td>
<td>3475 (3343-3620)</td>
<td>39.6 (38.1-41.2)</td>
<td>11.37 (10.93-11.84)</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>2,838,054</td>
<td>1364</td>
<td>543 (510-578)</td>
<td>39.8 (37.4-42.4)</td>
<td>19.13 (17.99-20.36)</td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>782,153</td>
<td>444</td>
<td>175 (159-192)</td>
<td>39.3 (35.8-43.3)</td>
<td>22.33 (20.31-24.59)</td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>583,228</td>
<td>421</td>
<td>Suppressed</td>
<td></td>
<td>Suppressed</td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>17,482,580</td>
<td>7202</td>
<td>3767 (3475-4073)</td>
<td>52.3 (48.3-56.6)</td>
<td>21.54 (19.88-23.30)</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>8,210,067</td>
<td>1902</td>
<td>773 (683-868)</td>
<td>40.7 (35.9-45.7)</td>
<td>9.42 (8.32-10.58)</td>
<td></td>
</tr>
<tr>
<td>Hawaii</td>
<td>1,111,188</td>
<td>273</td>
<td>Suppressed</td>
<td></td>
<td>Suppressed</td>
<td></td>
</tr>
<tr>
<td>Idaho</td>
<td>1,375,870</td>
<td>284</td>
<td>_a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illinois</td>
<td>9,809,562</td>
<td>3532</td>
<td>1063 (971-1169)</td>
<td>30.1 (27.5-33.1)</td>
<td>10.84 (9.90-11.92)</td>
<td></td>
</tr>
<tr>
<td>Indiana</td>
<td>5,188,514</td>
<td>2305</td>
<td>1250 (1161-1344)</td>
<td>54.2 (50.4-58.3)</td>
<td>24.10 (22.38-25.90)</td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>2,438,002</td>
<td>428</td>
<td>171 (150-194)</td>
<td>40 (35-45.3)</td>
<td>7.01 (6.14-7.95)</td>
<td></td>
</tr>
<tr>
<td>Kansas</td>
<td>2,217,059</td>
<td>474</td>
<td>222 (178-265)</td>
<td>46.8 (37.5-55.8)</td>
<td>10.01 (8.03-11.94)</td>
<td></td>
</tr>
<tr>
<td>Kentucky</td>
<td>3,475,334</td>
<td>2075</td>
<td>1104 (1025-1185)</td>
<td>53.2 (49.4-57.1)</td>
<td>31.77 (29.50-34.09)</td>
<td></td>
</tr>
<tr>
<td>Louisiana</td>
<td>3,564,038</td>
<td>1889</td>
<td>847 (718-986)</td>
<td>44.8 (38-52.2)</td>
<td>23.76 (20.16-27.67)</td>
<td></td>
</tr>
<tr>
<td>Maine</td>
<td>1,101,973</td>
<td>495</td>
<td>229 (202-257)</td>
<td>46.3 (40.8-52)</td>
<td>20.80 (18.34-23.34)</td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td>4,721,883</td>
<td>2757</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts</td>
<td>5,552,051</td>
<td>2298</td>
<td>1375 (1335-1416)</td>
<td>59.8 (58.1-61.6)</td>
<td>24.77 (24.05-25.51)</td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>7,839,742</td>
<td>2742</td>
<td>1295 (1237-1355)</td>
<td>47.2 (45.1-49.4)</td>
<td>16.52 (15.78-17.29)</td>
<td></td>
</tr>
<tr>
<td>Minnesota</td>
<td>4,356,123</td>
<td>1035</td>
<td>413 (379-450)</td>
<td>39.9 (36.6-43.5)</td>
<td>9.47 (8.69-10.33)</td>
<td></td>
</tr>
<tr>
<td>Mississippi</td>
<td>2,273,653</td>
<td>582</td>
<td>256 (202-318)</td>
<td>44.1 (34.7-54.6)</td>
<td>11.28 (8.89-13.99)</td>
<td></td>
</tr>
<tr>
<td>Missouri</td>
<td>4,780,119</td>
<td>1859</td>
<td>921 (856-987)</td>
<td>49.5 (46.1-53.1)</td>
<td>19.26 (17.92-20.64)</td>
<td></td>
</tr>
<tr>
<td>Montana</td>
<td>850,894</td>
<td>160</td>
<td>Suppressed</td>
<td></td>
<td>Suppressed</td>
<td></td>
</tr>
<tr>
<td>Nebraska</td>
<td>1,462,537</td>
<td>209</td>
<td>94 (69-115)</td>
<td>45 (33.2-55.2)</td>
<td>6.44 (4.74-7.88)</td>
<td></td>
</tr>
<tr>
<td>Nevada</td>
<td>2,440,679</td>
<td>817</td>
<td>292 (185-440)</td>
<td>35.7 (22.6-53.8)</td>
<td>11.95 (7.58-18.01)</td>
<td></td>
</tr>
<tr>
<td>New Hampshire</td>
<td>1,113,141</td>
<td>391</td>
<td>Suppressed</td>
<td></td>
<td>Suppressed</td>
<td></td>
</tr>
<tr>
<td>New Jersey</td>
<td>6,947,836</td>
<td>2827</td>
<td>1175 (1130-1223)</td>
<td>41.6 (40-43.3)</td>
<td>16.91 (16.26-17.60)</td>
<td></td>
</tr>
<tr>
<td>New Mexico</td>
<td>1,633,828</td>
<td>778</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>15,348,422</td>
<td>4950</td>
<td>2082 (2025-2142)</td>
<td>42.1 (40.9-43.3)</td>
<td>13.57 (13.20-13.95)</td>
<td></td>
</tr>
<tr>
<td>North Carolina</td>
<td>8,294,423</td>
<td>3129</td>
<td>1529 (1411-1654)</td>
<td>48.9 (45.1-52.9)</td>
<td>18.44 (17.01-19.94)</td>
<td></td>
</tr>
<tr>
<td>North Dakota</td>
<td>583,680</td>
<td>110</td>
<td>Suppressed</td>
<td></td>
<td>Suppressed</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>9,124,576</td>
<td>5179</td>
<td>2688 (2485-2903)</td>
<td>51.9 (48-56)</td>
<td>29.46 (27.23-31.81)</td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>3,027,263</td>
<td>750</td>
<td>289 (245-336)</td>
<td>38.5 (32.7-44.7)</td>
<td>9.55 (8.09-11.09)</td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>3,380,729</td>
<td>794</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>Population, n</td>
<td>Overdose deaths, n</td>
<td>Injection-involved overdose deaths</td>
<td>Rate per 100,000 (95% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>--------------------</td>
<td>------------------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Deaths, n (95% CI) % of overdose deaths (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>10,162,497</td>
<td>5141</td>
<td>2877 (2684-3077) 56 (52.2-59.8)</td>
<td>28.31 (26.41-30.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>855,276</td>
<td>395</td>
<td>182 (162-203) 46 (41-51.4)</td>
<td>21.24 (18.95-23.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>4,100,115</td>
<td>1730</td>
<td>487 (392-607) 28.2 (22.7-35.1)</td>
<td>11.89 (9.57-14.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>674,238</td>
<td>81</td>
<td>37 (26-49) 45.4 (32-59.9)</td>
<td>5.45 (3.84-7.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>5,373,433</td>
<td>3021</td>
<td>1587 (1446-1734) 52.5 (47.9-57.4)</td>
<td>29.53 (26.91-32.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texas</td>
<td>21,925,627</td>
<td>4116</td>
<td>1677 (1548-1823) 40.7 (37.6-44.3)</td>
<td>7.65 (7.06-8.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utah</td>
<td>2,320,603</td>
<td>619</td>
<td>217 (193-250) 35.1 (31.1-40.4)</td>
<td>9.36 (8.30-10.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vermont</td>
<td>510,181</td>
<td>189</td>
<td>93 (82-104) 49.4 (43.6-55.2)</td>
<td>18.29 (16.14-20.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>6,724,143</td>
<td>2226</td>
<td>990 (888-1101) 44.5 (39.9-49.5)</td>
<td>14.72 (13.21-16.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>6,027,818</td>
<td>1705</td>
<td>—                    —</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td>1,428,520</td>
<td>1328</td>
<td>Suppressed           Suppressed</td>
<td>Suppressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>4,574,131</td>
<td>1523</td>
<td>771 (694-857) 50.6 (45.6-56.3)</td>
<td>16.86 (15.18-18.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyoming</td>
<td>449,237</td>
<td>99</td>
<td>Suppressed           Suppressed</td>
<td>Suppressed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>National(^a)</td>
<td>256,662,010</td>
<td>91,071</td>
<td>38,253 (34,839-42,181) 45.1 (41.1-49.8)</td>
<td>15.97 (14.55-17.61)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Not available. These states did not report data to Treatment Episodes Data Set–Admissions.

\(^b\)National rate does not include populations from states that did not report data to Treatment Episodes Data Set–Admissions (Idaho, Maryland, New Mexico, Oregon, and Washington).

In addition to the 5 states that did not report TEDS-A data, we suppressed data from 7 states, resulting in 39 (77%) states with estimated results for 2020. In 2020, the estimated number of injection-involved overdose deaths by state ranged from 37 (95% CI 26–49) in South Dakota to 3767 (95% CI 3475–4073) in Florida (Table 2). There was large variation in the estimated injection-involved overdose death rate by state (median 14.72, range 5.45–31.77 per 100,000 people). Of the 10 states with the highest estimated injection overdose death rate, 5 were located in the South (Kentucky, Tennessee, Louisiana, Delaware, and Florida), 3 were in the Northeast (Pennsylvania, Massachusetts, and Rhode Island), and 2 (Ohio and Indiana) were in the Midwest Census regions. Of the 8 states in the Northeast that have results, 7 (86%) had an estimated injection overdose death rate higher than the national rate (Figure 1). All 6 states with results in the West had an estimated injection overdose death rate lower than the national rate. Over half (19682/38253, 51.5%) of the total number of estimated injection-involved overdoses occurred in 8 states (Florida, California, Pennsylvania, Ohio, New York, Texas, Tennessee, and North Carolina).
Table 2. Estimated number of injection-involved overdose deaths among adults, by US state and year, 2000-2020. State-level results were suppressed if the annual number of treatment admissions that reported any of the 5 drug types was <50 or if ≥15% of treatment admissions that reported a drug of interest were missing data on route of administration.

<table>
<thead>
<tr>
<th>State</th>
<th>2000 Overdose deaths, n</th>
<th>2020 Overdose deaths, n</th>
<th>Injection-involved deaths</th>
<th>% of overdose deaths (95% CI)</th>
<th>% of overdose deaths (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deaths, n (95% CI)</td>
<td>Deaths, n (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alabama</td>
<td>192</td>
<td>1023</td>
<td>31 (22-41)</td>
<td>16.1 (11.4-21.2)</td>
<td>44.6 (39.7-50.1)</td>
</tr>
<tr>
<td>Alaska</td>
<td>47</td>
<td>158</td>
<td>Suppressed</td>
<td>83 (68-100)</td>
<td>52.6 (42.8-63.3)</td>
</tr>
<tr>
<td>Arizona</td>
<td>517</td>
<td>2483</td>
<td>Suppressed</td>
<td>779 (539-1249)</td>
<td>31.4 (21.7-50.3)</td>
</tr>
<tr>
<td>Arkansas</td>
<td>135</td>
<td>540</td>
<td>36 (30-42)</td>
<td>27 (22.6-31.1)</td>
<td>46 (35.5-55.4)</td>
</tr>
<tr>
<td>California</td>
<td>1921</td>
<td>8778</td>
<td>565 (541-598)</td>
<td>29.4 (28.2-31.1)</td>
<td>39.6 (38.1-41.2)</td>
</tr>
<tr>
<td>Colorado</td>
<td>346</td>
<td>1461</td>
<td>97 (85-112)</td>
<td>28 (24.7-32.5)</td>
<td>37.9 (35.1-41)</td>
</tr>
<tr>
<td>Connecticut</td>
<td>317</td>
<td>1364</td>
<td>__a</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Delaware</td>
<td>54</td>
<td>444</td>
<td>10 (8-16)</td>
<td>19.1 (14-28.8)</td>
<td>39.3 (35.8-43.3)</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>76</td>
<td>421</td>
<td>Suppressed</td>
<td>Suppressed</td>
<td>Suppressed</td>
</tr>
<tr>
<td>Florida</td>
<td>1145</td>
<td>7202</td>
<td>329 (305-356)</td>
<td>28.7 (26.7-31.1)</td>
<td>52.3 (48.3-56.6)</td>
</tr>
<tr>
<td>Georgia</td>
<td>355</td>
<td>1902</td>
<td>25 (16-42)</td>
<td>7.2 (4.6-11.9)</td>
<td>40.7 (35.9-45.7)</td>
</tr>
<tr>
<td>Hawaii</td>
<td>63</td>
<td>273</td>
<td>8 (5-17)</td>
<td>12.4 (7.6-27.5)</td>
<td>39.8 (37.4-42.4)</td>
</tr>
<tr>
<td>Idaho</td>
<td>63</td>
<td>284</td>
<td>7 (5-15)</td>
<td>11.3 (8.2-23.3)</td>
<td>—</td>
</tr>
<tr>
<td>Illinois</td>
<td>856</td>
<td>3532</td>
<td>64 (51-83)</td>
<td>7.5 (5.9-9.7)</td>
<td>30.1 (27.5-33.1)</td>
</tr>
<tr>
<td>Indiana</td>
<td>214</td>
<td>2305</td>
<td>43 (38-48)</td>
<td>20.3 (17.9-22.6)</td>
<td>54.2 (50.4-58.3)</td>
</tr>
<tr>
<td>Iowa</td>
<td>71</td>
<td>428</td>
<td>11 (8-15)</td>
<td>16.2 (12-21.4)</td>
<td>40 (35-45.3)</td>
</tr>
<tr>
<td>Kansas</td>
<td>103</td>
<td>474</td>
<td>23 (17-27)</td>
<td>21.9 (16.7-26.7)</td>
<td>46.8 (37.5-55.8)</td>
</tr>
<tr>
<td>Kentucky</td>
<td>236</td>
<td>2075</td>
<td>Suppressed</td>
<td>Suppressed</td>
<td>Suppressed</td>
</tr>
<tr>
<td>Louisiana</td>
<td>246</td>
<td>1889</td>
<td>28 (21-37)</td>
<td>11.6 (8.6-15.2)</td>
<td>44.8 (38-52.2)</td>
</tr>
<tr>
<td>Maine</td>
<td>59</td>
<td>495</td>
<td>13 (11-17)</td>
<td>22.5 (18.8-29.5)</td>
<td>46.3 (40.8-52)</td>
</tr>
<tr>
<td>Maryland</td>
<td>614</td>
<td>2757</td>
<td>41 (37-47)</td>
<td>6.7 (6-7.6)</td>
<td>—</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>461</td>
<td>2298</td>
<td>44 (41-48)</td>
<td>9.6 (8.9-10.5)</td>
<td>59.8 (58.1-61.6)</td>
</tr>
<tr>
<td>Michigan</td>
<td>554</td>
<td>2742</td>
<td>164 (153-177)</td>
<td>29.6 (27.3-31.9)</td>
<td>47.2 (45.1-49.4)</td>
</tr>
<tr>
<td>Minnesota</td>
<td>128</td>
<td>1035</td>
<td>17 (13-23)</td>
<td>13.5 (10.4-17.7)</td>
<td>39.9 (36.6-43.5)</td>
</tr>
<tr>
<td>Mississippi</td>
<td>110</td>
<td>582</td>
<td>18 (12-25)</td>
<td>16.7 (10.7-22.6)</td>
<td>44.1 (34.7-54.6)</td>
</tr>
<tr>
<td>Missouri</td>
<td>299</td>
<td>1859</td>
<td>88 (76-102)</td>
<td>29.4 (25.5-34.3)</td>
<td>49.5 (46.1-53.1)</td>
</tr>
<tr>
<td>Montana</td>
<td>41</td>
<td>160</td>
<td>7 (4-11)</td>
<td>17.5 (10.6-26.2)</td>
<td>Suppressed</td>
</tr>
<tr>
<td>Nebraska</td>
<td>44</td>
<td>209</td>
<td>Suppressed</td>
<td>94 (69-115)</td>
<td>45 (33.2-55.2)</td>
</tr>
<tr>
<td>Nevada</td>
<td>273</td>
<td>817</td>
<td>74 (59-112)</td>
<td>26.9 (21.8-40.9)</td>
<td>35.7 (22.6-53.8)</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>45</td>
<td>391</td>
<td>4 (3-9)</td>
<td>10 (6.6-19.2)</td>
<td>Suppressed</td>
</tr>
<tr>
<td>New Jersey</td>
<td>639</td>
<td>2827</td>
<td>155 (145-169)</td>
<td>24.3 (22.7-26.5)</td>
<td>41.6 (40-43.3)</td>
</tr>
<tr>
<td>New Mexico</td>
<td>263</td>
<td>778</td>
<td>75 (55-112)</td>
<td>28.4 (21-42.6)</td>
<td>—</td>
</tr>
<tr>
<td>New York</td>
<td>762</td>
<td>4950</td>
<td>55 (50-61)</td>
<td>7.2 (6.6-8)</td>
<td>42.1 (40.9-43.3)</td>
</tr>
<tr>
<td>North Carolina</td>
<td>485</td>
<td>3129</td>
<td>135 (114-158)</td>
<td>27.9 (23.5-32.6)</td>
<td>48.9 (45.1-52.9)</td>
</tr>
<tr>
<td>North Dakota</td>
<td>14</td>
<td>110</td>
<td>Suppressed</td>
<td>Suppressed</td>
<td>Suppressed</td>
</tr>
<tr>
<td>Ohio</td>
<td>555</td>
<td>5179</td>
<td>137 (123-157)</td>
<td>24.7 (22.1-28.3)</td>
<td>51.9 (48-56)</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>233</td>
<td>750</td>
<td>63 (47-80)</td>
<td>27.1 (20.4-34.4)</td>
<td>38.5 (32.7-44.7)</td>
</tr>
<tr>
<td>State</td>
<td>2000 Overdose deaths, n</td>
<td>Injection-involved deaths</td>
<td>% of overdose deaths (95% CI)</td>
<td>2020 Overdose deaths, n</td>
<td>Injection-involved deaths</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------</td>
<td>---------------------------</td>
<td>-------------------------------</td>
<td>-------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Oregon</td>
<td>206</td>
<td>56 (51-62)</td>
<td>27.2 (25-30)</td>
<td>794</td>
<td>—</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>1134</td>
<td>306 (290-325)</td>
<td>27 (25.6-28.7)</td>
<td>5141</td>
<td>2877 (2684-3077)</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>73</td>
<td>6 (5-9)</td>
<td>8.4 (6.8-12.5)</td>
<td>395</td>
<td>182 (162-203)</td>
</tr>
<tr>
<td>South Carolina</td>
<td>252</td>
<td>37 (26-52)</td>
<td>14.8 (10.5-20.7)</td>
<td>1730</td>
<td>487 (392-607)</td>
</tr>
<tr>
<td>South Dakota</td>
<td>19</td>
<td>Suppressed</td>
<td>—</td>
<td>81</td>
<td>37 (26-49)</td>
</tr>
<tr>
<td>Tennessee</td>
<td>388</td>
<td>48 (34-70)</td>
<td>12.5 (8.7-18)</td>
<td>3021</td>
<td>1587 (1446-1734)</td>
</tr>
<tr>
<td>Texas</td>
<td>1006</td>
<td>290 (263-325)</td>
<td>28.9 (26.1-32.3)</td>
<td>4116</td>
<td>1677 (1548-1823)</td>
</tr>
<tr>
<td>Utah</td>
<td>210</td>
<td>65 (58-77)</td>
<td>31.2 (27.7-36.6)</td>
<td>619</td>
<td>217 (193-250)</td>
</tr>
<tr>
<td>Vermont</td>
<td>36</td>
<td>9 (7-14)</td>
<td>24.2 (19.1-39.7)</td>
<td>189</td>
<td>93 (82-104)</td>
</tr>
<tr>
<td>Virginia</td>
<td>407</td>
<td>59 (44-81)</td>
<td>14.5 (10.9-19.9)</td>
<td>2226</td>
<td>990 (888-1101)</td>
</tr>
<tr>
<td>Washington</td>
<td>551</td>
<td>131 (118-149)</td>
<td>23.7 (21.5-27.1)</td>
<td>1705</td>
<td>—</td>
</tr>
<tr>
<td>West Virginia</td>
<td>112</td>
<td>—</td>
<td>—</td>
<td>1328</td>
<td>Suppressed</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>244</td>
<td>47 (32-71)</td>
<td>19.3 (13.2-29.1)</td>
<td>1523</td>
<td>771 (694-857)</td>
</tr>
<tr>
<td>Wyoming</td>
<td>22</td>
<td>Suppressed</td>
<td>—</td>
<td>99</td>
<td>Suppressed</td>
</tr>
</tbody>
</table>

aNot available. These states did not report data to Treatment Episodes Data Set–Admissions.
Figure 1. Estimated injection-involved overdose death rate among adults, by state and region, United States, 2020. The vertical line represents the national estimated injection overdose death rate for 2020. States without an estimate either did not report data to TEDS-A or were suppressed. State-level results were suppressed if the annual number of treatment admissions that reported any of the 5 drug types was <50 or if ≥15% of treatment admissions that reported a drug of interest were missing data on route of administration.

The number of overdose deaths among adults increased more than 5-fold from 2000 (n=17,196) to 2020 (n=91,071; Table 2). Among jurisdictions that reported TEDS-A data, the estimated number of injection-involved overdose deaths increased more than 10-fold from 2000 (50/51 states reported TEDS-A data; n=3549, 95% CI 3103-4248) to 2020 (46/51 states reported TEDS-A data; n=38,253, 95% CI 34,839-42,181). Over the same time period, the national injection overdose death rate increased from 1.72 (95% CI 1.51-2.05) to 15.97 (95% CI 14.55-17.61). Much of that increase occurred from 2010 to 2020, in which the injection overdose death rate increased by 323% (95% CI 255%-391%). We estimated a state-level relative percentage change in the injection overdose death rate from 2010 to 2020 for 38 states. States that experienced the largest relative increase in injection overdose death rates were disproportionately concentrated in the Northeast (Figure 2). The injection overdose death rate increased more than 6-fold in Massachusetts (733% relative increase), Maine (678% relative increase), New York (610% relative increase), Virginia (600% relative increase), New Jersey (573% relative increase), Louisiana (545% relative increase), Florida (523% relative increase), and Connecticut (507% relative increase). There were 6 states (Utah, Texas, Arkansas, Missouri, Nebraska, and Iowa) that experienced a more gradual increase in the injection overdose death rate than the national increase. The estimated number of injection-involved overdose deaths and injection overdose death rate for each state and year are reported in Table S3 in Multimedia Appendix 3. Additionally, Figure S1 in Multimedia Appendix 4 visualizes the relationship between state-level overdose death rates and the percentage of overdose deaths that were estimated to be injection-involved for each year.
Figure 2. Estimated injection-involved overdose death rate among adults, by state and year, United States, from 2000 to 2020. The state-level results were suppressed if the annual number of treatment admissions that reported any of the 5 drug types was <50 or if ≥15% of treatment admissions that reported a drug of interest were missing data on route of administration.

Discussion

This work provides additional evidence that drug overdose mortality and other adverse outcomes related to injection drug use continue to be growing public health challenges across the United States. We previously demonstrated that much of the increase in overdose mortality since 2000 [1] has been attributable to injection-involved overdose deaths [9], and this analysis indicates that trend continued through 2020. We estimated that over half (34,704/67,986, 51%) of the increase in overdose deaths between 2000 and 2020 were injection-involved deaths. Importantly, these results provide additional granularity that reveals geographic differences in the magnitude of that trend. Although the injection-involved overdose death rate has increased dramatically in all states, the increase in injection overdose mortality is most pronounced in the Northeast and Appalachia regions, further emphasizing the need to prioritize harm reduction strategies in these areas. These geographic differences may be indicative of differences in the drug supply [36], an ongoing shift from prescription opioid mortality to injection-involved mortality [37], or geographic differences in access to overdose prevention interventions.

The estimated injection-involved overdose death rate for any jurisdiction is driven by both trends in overdose deaths (the total number and distribution of substances involved) and trends in reported injection behavior by substance type. Nationally, the likelihood of injecting heroin or synthetic opioids steadily increased before peaking in 2014 at 69.4% and declining to 58.3% in 2020 (Table S1 in Multimedia Appendix 1). However, during the same time period, the number of overdose deaths (46,764 in 2014 and 91,071 in 2020) and the proportion of overdose deaths attributable to heroin or synthetic opioids (39.6% in 2014 and 70.4% in 2020; Table S2 in Multimedia Appendix 2) continually increased, resulting in an increase in the overall injection-involved overdose death rate. As seen in Figure 2, some states have experienced a slight decrease in the estimated injection-involved overdose death rate in recent years due to slightly different temporal trends. For example, in Massachusetts, the injection-involved overdose death rate peaked in 2016. Similar to the national trend, the likelihood of injecting heroin or synthetic opioids has declined in recent years (79.6% in 2016 and 68.8% in 2020). This decrease outweighed the corresponding gradual increase in the percent of overdose deaths attributable to heroin or synthetic opioids (80.0% in 2016 and 85.8% in 2020), resulting in a decrease in the injection-involved overdose death rate. Similarly, recent trends in stimulant use and, subsequently, deaths attributed to stimulant use also influence these estimates. Generally, areas with a higher proportion of deaths attributable to stimulant use are estimated to have relatively fewer injection-involved overdose deaths.
based on a lower probability of injecting stimulants compared to heroin or synthetic opioids. In Tables S1 and S2 in Multimedia Appendices 1 and 2, we present both the probability of injecting by substance type and the percentage of deaths attributable to each substance as components of trends in injection-involved overdose in each state. Disaggregating the drivers of state-level trends in injection-involved overdose death rates can help inform the composition of harm reduction strategies.

Understanding changes in injection-involved overdose mortality provides additional value for designing overdose prevention and intervention strategies that cannot be gained from overall trends in overdose mortality alone. For example, states that are experiencing a particularly sharp increase in injection-involved overdose mortality may prioritize strategies that focus on injection harm reduction, such as naloxone access or supervised injection sites. Although the Centers for Disease Control and Prevention (CDC) has started to implement efforts to expand overdose surveillance and collect better data on the evidence of injection through the development of the State Unintentional Drug Overdose Report System (SUDORS) [38], the system does not yet have complete data on the route of administration and is too new to investigate temporal trends in injection-drug use.

Beyond the implications for overdose mortality prevention, these results can inform efforts to reduce infectious disease transmission due to injection drug use. From a public health surveillance perspective, injection drug use behavior, which is associated with an increased risk of transmission of several viruses and bacteria, is extremely difficult to measure and monitor [39]. These results can help state health departments and researchers better understand the burden of injection drug use in their jurisdiction. Recently, our team developed an approach that built upon this methodology to estimate that there were 3.7 million “people who injected drugs” in the United States in 2018 [40]. Following that framework, researchers could use these results, along with their own local data, to estimate the population size of people who injected drugs at the state level. There has been an increase in federal and state-led efforts to eliminate HCV [41] and HIV [42] infections, both of which disproportionately occur among people who injected drugs in the United States. The results from this analysis can be interpreted alongside data on the prevalence of HCV [17,43] or HIV [44] to highlight areas in most need of infectious disease prevention and treatment efforts. The comparison across states can inform the allocation of federal funding to scale up infectious disease prevention interventions [25,26,45] (eg, syringe service programs, substance use treatment centers, and hepatitis B vaccination clinics) or screening and treatment resources (eg, testing for acute viral hepatitis or HIV infections and providing linkage to care) [46,47].

As previously described in the development of this method [9], we note that we assumed the likelihood of injection within demographic strata and drug type is the same among living persons that enter treatment and persons who died of a drug overdose. This is a conservative approach because if the injection of any of these substances is more lethal than other routes of administration, these results would be an underestimate of the true burden of injection-involved mortality. Additionally, the selection of TEDS-A data to estimate the likelihood of injection comes with strengths and limitations. TEDS-A is the most comprehensive and complete single data source on substance abuse treatment admissions across the country. TEDS-A data collection has been an ongoing systematic activity for more than 30 years, which enables us to look at changes in reported behaviors over time. However, there may be differences in the data collection systems or the use of public funding for substance use treatment within each state [48]. As a result, the degree to which TEDS-A captures all treatment admissions may differ between states. However, this methodology does not depend on the complete enumeration of substance use treatment admissions but rather on the probability of reporting injection among those that are captured by TEDS-A data systems. In 2020, 99.2% of all admissions for a drug of interest reported data on the route of administration. The state-level response rate ranged from 44.6% in Wyoming to >99% in many states. After Wyoming, the second-lowest response rate was 95.2% in North Dakota.

Additionally, the analytic decision to categorize overdose deaths with multiple substances recorded for the drug with the highest likelihood of injection could result in an overestimate if polydrug use is less likely to involve injection than just using one of the drugs independently. In general, this assumption becomes more important, and the use of these estimates may be more limited in time periods or areas in which there is a high proportion of polysubstance use. A future analysis could validate these estimates with an analysis of reported injections from SUDORS [38] as those data become available. Finally, we assumed overdose deaths without T-codes were missing at random within strata and followed the same distribution as overdose deaths with recorded T-codes. However, the potential for this assumption to bias results has been reduced in recent years, as the completeness of these data has increased substantially over time [49].

Measuring trends in injection behavior and understanding how those trends impact the ongoing overdose mortality epidemic remains a huge public health challenge. These estimates provide longitudinal data points that reveal additional understanding of geographic heterogeneity in recent trends of injection-involved overdose mortality. As state and local public health departments continue to implement programs aimed at reducing overdose mortality and preventing infectious disease transmission, innovative and timely data sources can help inform the development of strategies best suited for their particular setting.

Acknowledgments

This study was supported by funding provided by Emory University AIDSVu (SOW #16).
Data Availability

The data sets analyzed during this study are available in the National Center for Health Statistics (NCHS) Restricted-Use Vital Statistics Data repository [29] or the Treatment Episodes Data Set–Admissions (TEDS-A) web-based repository [50]. All data generated during this study are included in this published article and its supplementary information files.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Total number of treatment admissions among adults that reported injection as a route of administration, by drug type, state and year, United States, 2000-2020.

[DOCX File, 302 KB - publichealth_v10i1e49527_app1.docx]

Multimedia Appendix 2

Percent of overdose deaths among adults attributed to each drug type, by state and year, United States, 2000-2020.

[DOCX File, 184 KB - publichealth_v10i1e49527_app2.docx]

Multimedia Appendix 3

Estimated number of injection-involved overdose deaths among adults, by state and year, United States, 2000-2020.

[DOCX File, 152 KB - publichealth_v10i1e49527_app3.docx]

Multimedia Appendix 4

Overdose death rate and the estimated percent of overdose deaths that were injection-involved among adults, by state and year, United States, 2000-2020. Note: Points represent the point estimate for the estimated percent of overdose deaths that were injection-involved.

[PNG File, 209 KB - publichealth_v10i1e49527_app4.png]

References


34. Percent changes. United States Census Bureau. 2015. URL: https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/percchg.pdf [accessed 2022-12-16]

Abbreviations

CDC: Centers for Disease Control and Prevention
HBV: hepatitis B virus
HCV: hepatitis C virus
ICD-10: International Classification of Diseases, Tenth Revision
NCHS: National Center for Health Statistics
NVSST: National Vital Statistics System
SUDORS: State Unintentional Drug Overdose Report System
TEDS-A: Treatment Episodes Data Set—Admissions
©Eric William Hall, Patrick Sean Sullivan, Heather Bradley. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 05.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Out-of-Hospital Cardiac Arrest Before and During the COVID-19 Pandemic in Hong Kong: Registry-Based Study From 2017 to 2023

Richard Huan Xu, PhD; Ruiqi Sun, MPhil; Siu-Ngor Fu, Prof Dr
Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, Hong Kong Polytechnic University, Kowloon, China (Hong Kong)

Corresponding Author:
Richard Huan Xu, PhD
Department of Rehabilitation Sciences
Faculty of Health and Social Sciences
Hong Kong Polytechnic University
Hung Hom
Kowloon
China (Hong Kong)
Phone: 852 27664199
Email: richard.xu@polyu.edu.hk

Abstract

Background: The COVID-19 pandemic has exerted a significant toll on individual health and the efficacy of health care systems. However, the influence of COVID-19 on the frequency and outcomes of out-of-hospital cardiac arrest (OHCA) within the Chinese population, both before and throughout the entire pandemic period, remains to be clarified.

Objective: This study aimed to fill the gaps by investigating the prevalence and outcomes of OHCA in Hong Kong (HK) both before and during the whole pandemic period.

Methods: This is a retrospective regional registry study. The researchers matched OHCA data with COVID-19–confirmed case records between December 2017 and May 2023. The data included information on response times, location of OHCA, witness presence, initial rhythm, bystander cardiopulmonary resuscitation (CPR), use of public-access defibrillation, resuscitation in the accident and emergency department, and survival to admission. Descriptive analyses were conducted, and statistical tests such as analysis of variance and $\chi^2$ were used to examine differences between variables. The incidence of OHCA and survival rates were calculated, and logistic regression analysis was performed to assess associations. The prevalence of OHCA and COVID-19 during the peak of the pandemic was also described.

Results: A total of 43,882 cases of OHCA were reported in HK and included in our analysis. Around 13,946 cases were recorded during the prepandemic period (2017-2019), and the remaining 29,936 cases were reported during the pandemic period (2020-2023). During the pandemic period, the proportion of female patients increased to 44.1% (13,215/29,936), and the average age increased slightly to 76.5 (SD 18.5) years. The majority of OHCAs (n=18,143, 61.1% cases) occurred at home. A witness was present in 45.9% (n=10,723) of the cases, and bystander CPR was initiated in 44.6% (n=13,318) of the cases. There was a significant increase in OHCA incidence, with a corresponding decrease in survival rates compared to the prepandemic period. The location of OHCA shifted, with a decrease in incidents in public places and a potential increase in incidents at home. We found that CPR (odds ratio 1.48, 95% CI 1.17-1.86) and public-access defibrillation (odds ratio 1.16, 95% CI 1.05-1.28) were significantly associated with a high survival to admission rate during the pandemic period. There was a correlation between the development of OHCA and the prevalence of COVID-19 in HK.

Conclusions: The COVID-19 pandemic has had a significant impact on OHCA in HK, resulting in increased incidence and decreased survival rates. The findings highlight the importance of addressing the indirect effects of the pandemic, such as increased stress levels and strain on health care systems, on OHCA outcomes. Strategies should be developed to improve OHCA prevention, emergency response systems, and health care services during public health emergencies to mitigate the impact on population health.

(JMIR Public Health Surveill 2024;10:e56054) doi:10.2196/56054
Introduction

Out-of-hospital cardiac arrest (OHCA) refers to the sudden cessation of cardiac activity outside a hospital setting. It is a critical medical emergency that poses significant challenges to individuals’ health and health care systems. The escalating prevalence of OHCA in recent years has become a pressing concern. OHCA can affect individuals across all age groups and backgrounds, making it a universal health issue. According to data from 2020, the incidence of OHCA ranges from 30.0 to 97.1 per 100,000 population, with survival to hospital discharge varying from 3.1% to 20.4% worldwide [1]. This highlights the urgent need for effective preventive measures and improved emergency response systems.

The COVID-19 pandemic was associated with an increased incidence of several diseases. The stress on health care systems, disruptions to health care services, and changes in lifestyle and behavior due to lockdown measures contributed to these effects. Some diseases that showed an increased incidence included mental health disorders, cardiovascular diseases, and respiratory diseases. The COVID-19 pandemic negatively affected all components of systems of care related to OHCA, disrupting the chain of survival. A recent review demonstrated an increase of approximately 120% in the incidence of OHCA and a 65% decrease in OHCA-related survival to hospital discharge during the pandemic, compared to the prepandemic period [2].

In addition, previous studies have reported significant changes in the incidence and epidemiological characteristics of OHCA and the prognosis of patients with OHCA during the pandemic compared to the pre–COVID-19 period. These reports have indicated a decrease in bystander cardiopulmonary resuscitation (CPR) and public-access defibrillator (PAD) use. In contrast, the emergency medical service (EMS) reaction time, frequency of arrests at home, and use of supraglottic airways have increased. Treatment-induced return of spontaneous circulation, survival to admission (STA), survival to discharge, 30-day survival, and favorable neurological outcomes all showed decreased incidence during the pandemic compared with the prepandemic period [2-5]. Despite the end of the COVID-19 pandemic, comprehensive analyses of the entire period are scarce. This lack of evaluation of the effects of COVID-19 on OHCA during each wave of the outbreak may skew the results.

While many studies have explored the effect of COVID-19 on OHCA during the pandemic, 2 research gaps need to be filled. First, a significant limitation is that no studies have comprehensively monitored changes in the prevalence of OHCA in the population during the prepandemic period and during the entire duration of the COVID-19 pandemic. The lack of continuous monitoring data hinders our ability to fully comprehend the impact of the progression of the COVID-19 pandemic on OHCA. Second, there is a notable deficiency in data concerning the epidemiology of OHCA and the corresponding survival rates within the Hong Kong (HK) population. HK, a highly developed region of China, exhibits a lower survival rate for patients with OHCA compared to other industrialized countries [1]. A previous study reported a local survival rate of 15.3% upon admission and a discharge rate of 2.3% for patients with OHCA [6]. The lack of comprehensive data on OHCA in HK during the pandemic hampers our ability to fully understand the scope of the problem and develop effective strategies to improve survival rates for patients with OHCA. It is crucial to conduct extensive research to gather detailed epidemiological data on OHCA in the HK population. This would provide valuable insights into the impact of COVID-19 on OHCA and inform the development of targeted interventions to improve patient outcomes [1,6]. Therefore, we aimed to conduct a large-scale regional registry study to systematically explore the incidence and outcomes of OHCA during the prepandemic and pandemic periods in HK.

Methods

Study Design and Participants

We established our database by matching OHCA data, based on electronic records provided by the Hong Kong Fire Service Department (HKFSD), with confirmed COVID-19 case records from open-source data [7] provided by the Hong Kong Centre for Health Protection (CHP). The HKFSD is the official sector that provides publicly funded paramedic emergency ambulance services to more than 7.4 million HK residents, operating from 41 ambulance depots. It maintains the electronic records of victims of OHCA who call for ambulance services. These electronic records include call triage and dispatch information, EMS response times, demographics, first aid provided, and short-term clinical outcomes. From January 2020, the CHP began to report the official number of confirmed cases of infection with SARS-CoV-2 variants in HK. Information on the number of cases and daily deaths due to COVID-19 in HK at the peak of the pandemic was obtained from the Hong Kong University information hub [8].

Data and Procedures

The HKFSD provided general data on OHCA cases through its internal record system. The research team collaborated with a superintendent from the HKFSD. This superintendent, who is in charge of quality assurance and ambulance services training, was appointed by the chief director of the HKFSD to provide the necessary data. The data elicitation method followed the method used in a previous study aimed at exploring the prevalence of OHCA in HK [6]. The senior author of that paper (YC Siu) is a member of our research team. A list of variables of interest developed based on internal discussion and a literature review was provided to the superintendent for internal discussion. The research team, superintendent, and senior medical officers from the HKFSD held 3 rounds of meetings to revise and confirm the data analysis protocol and the variable list. This study included patients of all ages. Victims of OHCA who were directly transferred to the public mortuary from the scene by EMS personnel and patients who did not use a ground ambulance were excluded from this study.
We collected data from December 1, 2017, to May 31, 2023. These data included the date of the incident, the response time (from the time of the call to the arrival of an ambulance or EMS personnel), the location of the OHCA (home, en route to hospital, nursing home, public place, or street), and the presence of an OHCA witness (yes or no). Clinical data consisted of the initial rhythm (pulseless electrical activity, ventricular fibrillation or ventricular tachycardia, asystole, or other), bystander CPR (yes or no), the number of times a PAD was used, resuscitation in the accident and emergency (A&E) department (yes or no), and STA (yes or no). Sociodemographic data included sex, age, and area of residence.

Although cases of COVID-19 were first reported in late December 2019 in Wuhan, the HK CHP started recording such cases in HK in January 2020. Therefore, we defined the prepandemic period as December 1, 2017, to December 31, 2019, and the pandemic period as January 1, 2020, to May 31, 2023, despite the World Health Organization announcing the end of the pandemic on May 5, 2023. We compared patient characteristics, the setting of OHCA occurrences, and OHCA outcomes between the prepandemic and pandemic periods.

Statistical Analysis
Descriptive analyses were primarily used to present our findings. Continuous data are presented as means with SDs, while categorical data are presented as frequencies (n) and proportions (%). The differences between continuous variables were examined using analysis of variance, the Mann-Whitney test, or the Kruskal-Wallis test, as appropriate. Differences between categorical variables were examined using the χ² test.

The total incidence of OHCA (‰) in HK and the incidence stratified by 3 districts (Hong Kong Island, Kowloon, and the New Territories) are presented as the number of OHCA cases divided by the mid-year population and survival to A&E department admission (%) from 2018 to 2022, respectively. Figures were created to visually represent the changing incidence of OHCA across different years and its relationship with the outbreak and progression of the COVID-19 pandemic. Additionally, we present the change in the prevalence of OHCA at different key cutoff points (eg, different waves of an outbreak) of the COVID-19 pandemic in HK.

Multivariate logistic regression analysis was used to assess the association between survival and A&E department admission, as well as the use of CPR and PAD during the prepandemic and pandemic periods. The analysis was adjusted for factors including sex, age, the cause of OHCA, the initial rhythm, the response time, whether the OHCA was witnessed, the location of the OHCA, and the area of residence of the patient. Adjusted odds ratios (aORs) and their 95% CIs were calculated. Furthermore, we collected additional data and described the prevalence of OHCA and COVID-19 during the peak of the pandemic (11 weeks from February 7, 2022, to April 24, 2022). Differences with P values less than .05 were considered statistically significant. All statistical analyses were performed using Stata (version 16; StataCorp LLC).

Ethical Considerations
The Institutional Review Board of Hong Kong Polytechnic University approved this study protocol (HSEARS 20211004004). The requirement for informed consent was waived.

Results
Participants’ Background Characteristics
From December 1, 2017, to May 31, 2023, HK reported 43,882 cases of OHCA (Table 1). Of these, 13,946 cases were recorded during the prepandemic period (2017-2019), and the remaining 29,936 cases were reported during the pandemic period (2020-2023). During the prepandemic period, 56.6% (7896/13,946) of the patients were men, with an average age of 75.81 (SD 17.12) years. The majority of OHCAs (n=7566 cases, 54.3%) occurred at home. During the pandemic period, the proportion of female patients increased to 44.1% (13,215/29,936), and the average age increased slightly to 76.5 (SD 18.5) years. The majority of OHCAs (n=18,143, 61.1% cases) occurred at home. A witness was present in 45.9% (n=10,723) of the cases, and CPR was initiated in 44.6% (n=13,318) of the cases. The average response time increased to 13.6 (SD 13.3) minutes, and only 8.3% (n=2464 cases) of the patients were alive when admitted to an A&E department, which is half the rate observed in the prepandemic period.
Table 1. Background characteristics of individuals and the outcome of OHCA<sup>a</sup>.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Prepandemic (December 2017 to December 2019)</th>
<th>Pandemic (January 2020 to May 2023)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n/N (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7896/13,946 (56.6)</td>
<td>16,721/29,936 (55.9)</td>
<td>.13</td>
</tr>
<tr>
<td>Female</td>
<td>6050/13,946 (43.4)</td>
<td>13,215/29,936 (44.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Region, n/N (%)</strong></td>
<td></td>
<td></td>
<td>.24</td>
</tr>
<tr>
<td>Hong Kong Island</td>
<td>2679/13,946 (19.2)</td>
<td>5568/29,985 (18.6)</td>
<td></td>
</tr>
<tr>
<td>Kowloon</td>
<td>4992/13,946 (35.8)</td>
<td>10,747/29,985 (35.8)</td>
<td></td>
</tr>
<tr>
<td>New Territories and Outlying Islands</td>
<td>6275/13,946 (45)</td>
<td>13,670/29,985 (45.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Location of OHCA, n/N (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Street</td>
<td>622/13,944 (4.5)</td>
<td>1077/29,715 (3.6)</td>
<td></td>
</tr>
<tr>
<td>Home</td>
<td>7566/13,944 (54.3)</td>
<td>18,143/29,715 (61.1)</td>
<td></td>
</tr>
<tr>
<td>En route to a hospital</td>
<td>692/13,944 (5)</td>
<td>1366/29,715 (4.6)</td>
<td></td>
</tr>
<tr>
<td>Home for aged</td>
<td>3439/13,944 (24.7)</td>
<td>6475/29,715 (21.8)</td>
<td></td>
</tr>
<tr>
<td>Public places</td>
<td>1625/13,944 (11.7)</td>
<td>2654/29,715 (8.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Witnessed arrest, n/N (%)</strong></td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>No</td>
<td>9165/13,946 (65.7)</td>
<td>19,179/29,902 (64.1)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4781/13,946 (34.3)</td>
<td>10,723/29,902 (35.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Initial electrocardiogram rhythm, n/N (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Asystole</td>
<td>11,098/13,942 (79.6)</td>
<td>24,327/29,875 (81.4)</td>
<td></td>
</tr>
<tr>
<td>Pulseless electrical activity</td>
<td>1880/13,942 (13.5)</td>
<td>3730/29,875 (12.5)</td>
<td></td>
</tr>
<tr>
<td>VF/VT&lt;sup&gt;b&lt;/sup&gt;</td>
<td>937/13,942 (6.7)</td>
<td>1712/29,875 (5.7)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>27/13,942 (0.2)</td>
<td>106/29,875 (0.4)</td>
<td></td>
</tr>
<tr>
<td><strong>Apparent cause of arrest, n/N (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Trauma</td>
<td>659/13,946 (4.7)</td>
<td>1128/29,972 (3.8)</td>
<td></td>
</tr>
<tr>
<td>Nontrauma</td>
<td>13,287/13,946 (95.3)</td>
<td>28,844/29,972 (96.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Resuscitation done in A&amp;E&lt;sup&gt;c&lt;/sup&gt;, n/N (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>6780/13,946 (48.6)</td>
<td>4264/8068 (52.9)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7166/13,946 (51.4)</td>
<td>3804/8068 (47.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Bystander CPR&lt;sup&gt;d&lt;/sup&gt;, n/N (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>9308/13,945 (66.7)</td>
<td>16,534/29,852 (55.4)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4637/13,945 (33.3)</td>
<td>13,318/29,852 (44.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Bystander PAD&lt;sup&gt;e&lt;/sup&gt;, n/N (%)</strong></td>
<td></td>
<td></td>
<td>.37</td>
</tr>
<tr>
<td>No</td>
<td>13,665/13,946 (98)</td>
<td>29,423/29,985 (98.1)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>281/13,946 (2)</td>
<td>562/29,985 (1.9)</td>
<td></td>
</tr>
<tr>
<td><strong>STA&lt;sup&gt;f&lt;/sup&gt;, n/N (%)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>11,478/13,946 (82.3)</td>
<td>27,309/29,773 (91.7)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2468/13,946 (17.7)</td>
<td>2464/29,773 (8.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Response time (min), mean (SD)</strong></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>10.6 (7.8)</td>
<td>13.6 (13.3)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>75.8 (17.1)</td>
<td>76.5 (18.5)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>OHCA: out-of-hospital cardiac arrest.
<sup>b</sup>VF/VT: ventricular fibrillation or ventricular tachycardia.
Prevalence of OHCA During the COVID-19 Pandemic

Figure 1 presents a comparative analysis of the prevalence of OHCA during this study’s period. The prevalence of OHCA was lower in 2018-2019 than in 2020-2023. During the pandemic, each wave of the COVID-19 outbreak in HK aligned with a significant increase in the prevalence of OHCA. Notably, the peak in OHCA incidents coincided with the onset of the Chinese New Year holiday in 2022, a surge that lasted for approximately 11 weeks. This period marked the highest recorded number of COVID-19 cases throughout the pandemic. We also found a significant increase in the prevalence of OHCA following the discontinuation of zero-tolerance measures in mainland China. However, the removal of mandatory mask-wearing regulations by the HK government in March 2023 did not result in a significant increase in the prevalence of OHCA.


OHCA Incidence Stratified by Year and Region

Table 2 demonstrates the increased incidence of OHCA from 2018 (6438/7,451,500, 0.85‰) to 2022 (10,618/7,345,000, 1.45‰). However, the STA rate significantly decreased from 16.9% (1090/6438) to 5.6% (584/10,424) during this period. When stratified by region, the incidence of OHCA was higher in Kowloon and Hong Kong Island than in the New Territories. During the prepandemic period, the STA rate was approximately 4% higher in the New Territories than in the other 2 districts. However, during the pandemic period, the gap in STA rates narrowed, with the STA rate in the New Territories becoming slightly lower than the STA rates in Kowloon in 2021 (245/3407, 7.2%, vs 212/2805, 7.6%) and 2022 (277/4837, 5.7%, vs 215/3636, 5.9%). Figure 2 provides additional information on the incidence of OHCA in all 18 districts. Generally, the incidence of OHCA was lower in districts in the north (New Territories) than in the other districts. Kwun Tong District (Kowloon, mid-part of HK) reported the highest incidence of OHCA during both the prepandemic and pandemic periods. The OCHA incidents, STA rate, and use of CPR among different demographic groups in the 3 regions over this study’s period are detailed in Multimedia Appendix 1.
Table 2. OHCA\textsuperscript{a}, incidence, STA\textsuperscript{b}, and CPR\textsuperscript{c} ratio in HK\textsuperscript{d} and stratified by regions.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Year</th>
<th>KW\textsuperscript{e}</th>
<th>HKI\textsuperscript{f}</th>
<th>NT\textsuperscript{g}</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHCA incidence, n (%)/KW</td>
<td>2018</td>
<td>2338 (1.03)</td>
<td>1232 (0.98)</td>
<td>2868 (0.73)</td>
<td>6438 (0.86)</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>2398 (1.04)</td>
<td>1314 (1.05)</td>
<td>3119 (0.79)</td>
<td>6831 (0.91)</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>2922 (1.27)</td>
<td>1504 (1.23)</td>
<td>3642 (0.92)</td>
<td>8068 (1.08)</td>
</tr>
<tr>
<td></td>
<td>2021</td>
<td>2818 (1.26)</td>
<td>1468 (1.23)</td>
<td>3410 (0.86)</td>
<td>7696 (1.04)</td>
</tr>
<tr>
<td></td>
<td>2022</td>
<td>3695 (1.68)</td>
<td>1978 (1.71)</td>
<td>4945 (1.24)</td>
<td>10,618 (1.45)</td>
</tr>
<tr>
<td>STA, n (%)</td>
<td>2018</td>
<td>342 (14.6)</td>
<td>196 (15.9)</td>
<td>552 (19.2)</td>
<td>1090 (16.9)</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>369 (15.4)</td>
<td>225 (17.1)</td>
<td>667 (21.4)</td>
<td>1261 (18.5)</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>348 (11.9)</td>
<td>181 (12)</td>
<td>570 (15.7)</td>
<td>1099 (13.6)</td>
</tr>
<tr>
<td></td>
<td>2021</td>
<td>212 (7.6)</td>
<td>84 (5.7)</td>
<td>245 (7.2)</td>
<td>541 (7)</td>
</tr>
<tr>
<td></td>
<td>2022</td>
<td>215 (5.9)</td>
<td>92 (4.7)</td>
<td>277 (5.7)</td>
<td>584 (5.6)</td>
</tr>
<tr>
<td></td>
<td>2023</td>
<td>99 (7.5)</td>
<td>31 (5)</td>
<td>110 (6.6)</td>
<td>240 (6.7)</td>
</tr>
<tr>
<td>CPR, n (%)</td>
<td>2018</td>
<td>674 (28.8)</td>
<td>338 (27.4)</td>
<td>851 (29.7)</td>
<td>1863 (28.9)</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>872 (36.4)</td>
<td>532 (40.5)</td>
<td>1215 (39)</td>
<td>2619 (38.3)</td>
</tr>
<tr>
<td></td>
<td>2020</td>
<td>1095 (37.5)</td>
<td>605 (40.2)</td>
<td>1502 (41.2)</td>
<td>3202 (39.7)</td>
</tr>
<tr>
<td></td>
<td>2021</td>
<td>1147 (40.8)</td>
<td>677 (46.1)</td>
<td>1557 (45.7)</td>
<td>3381 (44)</td>
</tr>
<tr>
<td></td>
<td>2022</td>
<td>1715 (46.8)</td>
<td>965 (49)</td>
<td>2411 (49.3)</td>
<td>5091 (48.4)</td>
</tr>
<tr>
<td></td>
<td>2023</td>
<td>601 (46.1)</td>
<td>291 (47.6)</td>
<td>752 (45.4)</td>
<td>1644 (46)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}OHCA: out-of-hospital cardiac arrests.
\textsuperscript{b}STA: survival to admission.
\textsuperscript{c}CPR: cardiopulmonary resuscitation.
\textsuperscript{d}HK: Hong Kong.
\textsuperscript{e}KW: Kowloon.
\textsuperscript{f}HKI: Hong Kong Island.
\textsuperscript{g}NT: New Territories and Outlying Islands.
Figure 2. Surge of OHCA incidence across HK districts pre- (2018 and 2019) and during pandemic (2020-2022). HK: Hong Kong; OHCA: out-of-hospital cardiac arrest.

Relationship Between STA and Patients’ Characteristics

Table 3 presents the results of the multivariate logistic regression analysis of the associations between the STA rate and various factors during the prepandemic and pandemic periods. The key findings included higher odds of OHCA among men during the pandemic period and higher odds of certain initial electrocardiogram rhythms during the same period. There were also higher odds of nontrauma causes of OHCA during both periods. There were other factors, such as the response time, whether the OHCA was witnessed, the location of the OHCA, and the region exhibited varying aORs between the 2 periods. Notably, we found that bystander CPR administration (aOR 1.48, 95% CI 1.17-1.86) and PAD use (aOR 1.16, 95% CI 1.05-1.28) were significantly associated with a high STA rate during the pandemic period.
Table 3. The associations of STA\(^a\) and patient’s characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Prepandemic period</th>
<th>Pandemic period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted OR(^b) (95% CI)</td>
<td>Adjusted OR (95% CI)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>—(^c)</td>
<td>—</td>
</tr>
<tr>
<td>Male</td>
<td>0.95 (0.87-1.05)</td>
<td>1.12(^*) (1.02-1.24)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.99(^***) (0.99-0.99)</td>
<td>0.99(^***) (0.99-0.99)</td>
</tr>
<tr>
<td><strong>Cause of arrest</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trauma</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Nontrauma</td>
<td>4.35(^***) (3.10-6.10)</td>
<td>4.48(^***) (3.05-6.56)</td>
</tr>
<tr>
<td><strong>Initial electrocardiogram rhythm</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asystole</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>VF/VT(^d)</td>
<td>2.92(^***) (2.49-3.43)</td>
<td>3.49(^***) (3.03-4.01)</td>
</tr>
<tr>
<td>PEA(^e)</td>
<td>2.31(^*) (2.05-2.61)</td>
<td>2.91(^***) (2.60-3.25)</td>
</tr>
<tr>
<td>Response time</td>
<td>0.99 (0.98-1.00)</td>
<td>0.97(^***) (0.96-0.98)</td>
</tr>
<tr>
<td><strong>Witnessed arrest</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No</td>
<td>0.48(^***) (0.43-0.53)</td>
<td>0.49(^***) (0.45-0.54)</td>
</tr>
<tr>
<td><strong>Location of OHCA(^f)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Street</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Home</td>
<td>0.47(^***) (0.37-0.58)</td>
<td>0.59(^***) (0.48-0.72)</td>
</tr>
<tr>
<td>En route to a hospital</td>
<td>0.78 (0.58-1.05)</td>
<td>1.56(^***) (1.19-2.05)</td>
</tr>
<tr>
<td>Home for aged</td>
<td>0.45(^***) (0.36-0.58)</td>
<td>0.53(^***) (0.42-0.67)</td>
</tr>
<tr>
<td>Public area excluding street</td>
<td>0.75(^*) (0.59-0.95)</td>
<td>0.97 (0.78-1.20)</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NT(^g)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>HKI(^b)</td>
<td>0.78 (0.68-0.88)</td>
<td>0.75(^***) (0.66-0.85)</td>
</tr>
<tr>
<td>KW(^i)</td>
<td>0.66(^***) (0.60-0.74)</td>
<td>0.86(^*) (0.78-0.95)</td>
</tr>
<tr>
<td><strong>Bystander PAD(^j)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>1.14 (0.86-1.51)</td>
<td>1.48(^*) (1.17-1.86)</td>
</tr>
<tr>
<td><strong>Bystander CPR(^k)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Yes</td>
<td>1.04 (0.93-1.16)</td>
<td>1.16(^**) (1.05-1.28)</td>
</tr>
<tr>
<td>Pseudo R(^2)</td>
<td>0.102</td>
<td>0.125</td>
</tr>
<tr>
<td>Akaike information criterion</td>
<td>11,597.9</td>
<td>14,440.9</td>
</tr>
<tr>
<td>Bayesian information criterion</td>
<td>11,725.9</td>
<td>14,581.6</td>
</tr>
</tbody>
</table>

\(^a\)STA: survival to admission.
\(^b\)OR: odds ratio.
\(^c\)Not applicable.
Correlation Between OHCA and COVID-19 Cases During the Pandemic

Figure 3 graphically presents the correlation between OHCA incidents and the progression of COVID-19 during the pandemic period in HK. Overall, there was a consistent relationship between these 2 variables, with a higher SARS-CoV-2 infection rate leading to an increased prevalence of OHCA. This correlation was particularly evident in 2022, which had the highest prevalence of COVID-19 and corresponded with the highest prevalence of OHCA. Table 4 provides a detailed overview of the OHCA and COVID-19 statistics from February 7, 2022, to April 24, 2022. This period marked the peak of COVID-19 cases, which coincided with a sharp increase in the number of OHCA incidents. During these 11 weeks, the number of OHCA cases increased dramatically, reaching a peak of approximately 114 cases per day in the fourth week. Concurrently, the number of COVID-19 cases increased, reaching a daily average of 67,583 cases during the same week. However, the highest daily death toll from COVID-19 was recorded in the fifth week. In contrast, the STA rate in cases of OHCA demonstrated a slightly different trend, with marginally higher numbers recorded during the most severe 5 weeks (weeks 3-7) of the COVID-19 outbreak.

Figure 3. Correlation of prevalence between OHCA and COVID-19 during pandemic (January 2020 to May 2023). OHCA: out-of-hospital cardiac arrests.

---

dVF/VT: ventricular fibrillation or ventricular tachycardia.
ePEA: pulseless electrical activity.
fOHCA: out-of-hospital cardiac arrests.
gNT: New Territories and Outlying Islands.
hHKI: Hong Kong Island.
iKW: Kowloon.
jPAD: public-access defibrillation.
kCPR: cardiopulmonary resuscitation.

*p<.05, **p<.01, ***p<.001.
During the pandemic, a significant shift in OHCA location was occurring in public places, likely due to the implementation of lockdowns and social distancing measures, which resulted in fewer people being out in public. This finding aligns with those of most previous studies [11-13], but a different trend has been reported in some Asian populations. For instance, Riyapan et al [14] reported no significant change in the occurrence of OHCA in public places in Thailand before and during the pandemic. Watanabe et al [15] also reported no decrease in the occurrence of OHCA in public places. However, the decrease observed in the number of OHCAs in public places does not necessarily indicate a reduction in the overall number of OHCAs. Instead, this may suggest a shift in the location of these incidents, with a potential increase in the number of OHCAs occurring at home.

There was a significant decrease in the number of OHCAs reported in nursing homes during the pandemic. This trend is surprising given the vulnerability of the older population to OHCA. The findings for this outcome in previous studies have been mixed. For example, a study in Portugal reported no difference between the number of OHCAs in nursing homes before and during the pandemic [13]. Another study in the United States found that the percentage of patients experiencing an OHCA in a nursing home was higher in 2020 than in 2019 [16]. A study in HK revealed that 96% of COVID-19–related deaths were among individuals aged 60 years and older, with 53% being nursing home residents [17]. Our findings suggest that, although the number of OHCAs in nursing homes decreased during the pandemic, this does not necessarily mean fewer older individuals experienced an OHCA. Many older individuals may have died from COVID-19, and their deaths were not classified as OHCA-related. This may result in an underestimation of the devastating impact of OHCA on the older population, particularly for those in nursing homes, during this public health emergency. The government should update and standardize mortality calculations during future public health emergencies [18].

Another reason for the decrease in the number of OHCAs in nursing homes during the pandemic may be the effectiveness of nonpharmaceutical interventions in reducing the number of COVID-19 cases and fatalities within long-term care facilities.
and nursing homes. Previous studies have shown an association between a decrease in the number of COVID-19 cases and the rigor of various containment measures, notably school and workplace closures, and public information campaigns globally [19,20]. In HK, border restrictions, quarantine and isolation, social distancing, and behavioral changes in the population were significantly linked to the control of the COVID-19 pandemic [21]. This may underscore the crucial role that nonpharmaceutical interventions can play in safeguarding vulnerable populations during pandemics.

Our study revealed an increase in the bystander CPR rate during both the prepandemic and pandemic periods, albeit at a slower rate during the pandemic. During the prepandemic period, the bystander CPR rate increased by approximately 10% from 2018 to 2019. However, this trend plateaued with the onset of the COVID-19 pandemic, resulting in a negligible difference in the CPR rate between 2019 and 2020. A Korean study reported a similar finding [22], but studies in some Western countries have reported notable decreases in CPR rates. For example, in France, the CPR rate dramatically fell from 63.9% to 47.8% between March and April 2020 [12]. A study in Spain also found a decrease in CPR rates from 51.1% to 42.6% between 2017/2018 and 2020 [23]. These decreases in CPR rates highlight the profound effects of the pandemic on critical medical procedures across different nations. Nevertheless, we found that, as the pandemic progressed, the CPR rate showed another increase, stabilizing between 44% and 48%. Data on bystander CPR rates during the pandemic are mixed. A systematic review of studies conducted in 2020 suggested a decline in the incidence of bystander CPR at the onset of the pandemic, with 1 study noting that this decline was observed only in patients diagnosed with COVID-19 [24]. Another review, including studies published in 2020 and 2021, indicated a change in the community response to OHCA, with fewer occurrences of bystander CPR [25]. However, a recent review of studies published in 2022 found no significant difference in bystander CPR rates between the COVID-19 pandemic and prepandemic periods [4]. It is important to note that the pandemic also affected other aspects of the CPR response and outcomes, such as EMS availability. A review of the relationship between OHCA and EMS indicated that difficulties encountered by the first responder system during the COVID-19 pandemic, such as dispatcher overload, increased response times, and adherence to personal protective equipment requirements, were superimposed on shortages in the supply of personal protection equipment [26]. Our findings highlight the complex impact of the pandemic on OHCA outcomes.

Although the frequency of witnesses of OHCA and bystander CPR significantly increased, we observed a drastic reduction in the rate of STA en route to the A&E department during the pandemic, which is consistent with the findings of previous studies [4,24,25]. During the prepandemic period, the overall STA rate in 2018 was approximately 16.9%, and this increased slightly to 18.5% in 2019. However, during the first year of the pandemic, the STA rate decreased by more than 25% and further decreased to approximately 5.6% at the peak of the pandemic in 2022. By the end of this study’s period, the STA rate had slightly increased to 6.7% in the first 5 months of 2023 but remained approximately 65% lower than the rate observed in the final year of the prepandemic period (2019). The STA rate was lower in HK than in other countries or regions during the pandemic [12,27,28]. Several factors may have accounted for this discrepancy. An overburdened health care system can lead to delayed recognition of and intervention in unstable patients. For instance, following the 2022 Chinese New Year holiday, public hospitals in HK were dealing with more than 1800 COVID-19 cases requiring EMS per day. Emergency medical systems, especially emergency departments, which serve as the frontline of the health care system, experienced severe overload and challenging working conditions. This was a global issue [29]. Moreover, there were fewer cardiac arrests in public places during the pandemic, which may have influenced the survival rate of patients with OHCA. Additionally, the lack of sufficient CPR training in HK [30,31] and the fear of COVID-19 may have affected the willingness of witnesses to perform CPR. Given that CPR is considered an aerosol-generating procedure with a significant risk of viral transmission, witnesses may have hesitated to perform it, potentially affecting the STA rate [32].

A strong correlation was observed between outbreaks of COVID-19 and surges in OHCA cases in HK over the past 4 years. In particular, the fifth wave of the pandemic, which coincided with the Chinese New Year holiday, resulted in a significant surge in OHCA cases. The older population, the most vulnerable group to both COVID-19 and OHCA, bore the brunt of the pandemic. The average age of our sample was over 75 years (SD 17.1 years), highlighting the insufficient protection provided by the government for this demographic during the pandemic. A previous review demonstrated a gradual age-associated increase in the COVID-19 mortality rate [33]. For example, 1 study showed a 10.5% fatality rate for older patients compared to 0.43% for younger patients [34]. As of April 2022, less than one-third of HK residents aged older than 80 years had been fully vaccinated. The vaccination rate was even lower among residents of care homes when the Omicron variant arrived in HK, at less than 20%. This is not surprising, given that vaccine hesitancy to both the first vaccine dose and the booster was widespread [35]. Another study underscored the urgency of vaccination for older people, revealing that approximately one-quarter of countries reported lower vaccination coverage among older adults compared to the overall population [36]. Additionally, as reported in a previous study [37], lockdown measures and no-visit policies resulted in older adults being unable to meet their families for extended periods, leading to heightened levels of anxiety and depression among this population, thereby increasing the risk of OHCA. It is crucial to develop comprehensive policies and strategies to mitigate the direct and indirect effects of public health emergencies on health outcomes in the older population.

In 2019, the HK Legislative Council suggested promoting bystander intervention to improve the survival rates for patients with OHCA. The suggestions included adopting technologies to reduce rescue time, providing education and training to enhance the skills of the public, and urging the government to develop laws and regulations to protect bystanders legally and limit their liability [38]. Thus, the outcomes of our regression analysis revealed that, during the pandemic, bystander CPR and PAD use significantly improved the STA rate for individuals...
with OHCA compared to the prepandemic period, during which the associations between STA rate and bystander interventions were insignificant. Furthermore, our regression analysis revealed a significant relationship between the response time and STA rate. To handle emergencies during the pandemic, the HKFSD adopted a target-oriented approach to devise various emergency response strategies and measures. In the meantime, the Director’s Command Post also established an artificial intelligence–powered social media service platform capable of providing automated inquiries and instant replies 24 h/d, 7 d/wk [39]. The importance of reform and collaboration in multiple social sectors is paramount to improving the STA rate of patients with OHCA during public health emergencies.

Limitations
Our study is among the first to delineate the relationship between COVID-19 and the prevalence of OHCA at the population level during both the prepandemic and pandemic periods. However, this study has certain limitations that need to be acknowledged. First, although we observed a decrease in the STA rate during the pandemic, we lacked information about patient survival rates upon discharge and their short- and long-term health outcomes, because patient data in A&E departments and hospital settings are managed by the Hospital Authority rather than the HKFSD. Matching OHCA patient data with medical records would provide a more comprehensive understanding of the impact of COVID-19 on OHCA in HK. Second, the HKFSD did not collect information about whether patients with OHCA had COVID-19 during the provision of services. Consequently, we lack knowledge about the prevalence of COVID-19 among patients with OHCA, which may have led to an inaccurate estimation of the impact of COVID-19 on the survival rate of patients with OHCA in HK.

Conclusions
In our study of the prevalence of OHCA from 2017 to 2023, we observed a significant increase in OHCA incidence and a considerable decrease in survival rates of patients with OHCA during the COVID-19 pandemic compared to the period before the pandemic. Moreover, we found that the spikes in the number of OHCA cases closely matched the increase in the number of confirmed COVID-19 cases in HK. Components of both the social sector, such as the Home Office, and the health care sector, such as EMS systems, should prepare for potential increases in the number of OHCA cases during future public health crises. It is necessary to reform these sectors to make them more resilient, efficient, and effective at meeting the needs of individuals with OHCA. Developing an effective multidisciplinary mechanism for leveraging resources, sharing knowledge and expertise, and coordinating efforts are key factors for improving the survival rates of patients with OHCA during future pandemics.

Acknowledgments
The authors appreciate the support from the SADS Hong Kong Foundation, Chan Sui Kau Family Benefits and Charitable Foundation, and the Hong Kong Fire Service Department. This study was supported by the Chan Sui Kau Family Benefits and Charitable Foundation (P0038183).

Data Availability
The data sets generated or analyzed during this study are not publicly available due to privacy concerns. However, they may be obtained from the corresponding author upon reasonable request.

Authors’ Contributions
RHX and AF did the conceptualization, project administration, handled resources, supervision, and data curation. RHX and RS worked on the formal analysis, software, validation, and visualization. RHX handled the methodology and writing of the original draft. RHX, RS, and AF did the review and editing of the writing.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Out-of-hospital cardiac arrest incidence, survival to admission, and cardiopulmonary resuscitation ratio in different demographic and geographic groups.

References


8. LKS Faculty of Medicine School of Public Health. HKU Med. URL: https://covid19.sph.hku.hk/ [accessed 2024-05-03]


**Abbreviations**

A&E: accident and emergency  
aOR: adjusted odds ratio  
CHP: Centre for Health Protection  
CPR: cardiopulmonary resuscitation  
EMS: emergency medical service  
HK: Hong Kong  
HKFSD: Hong Kong Fire Service Department  
OHCA: out-of-hospital cardiac arrest  
PAD: public-access defibrillator  
STA: survival to admission

Brady W Bennett¹, MPH; Stephanie DuBose², MPH; Ya-Lin A Huang³, PhD; Christopher H Johnson, MS; Karen W Hoover³, MPH, MD; Jeffrey Wiener³, PhD; David W Purcell¹, PhD, JD; Patrick S Sullivan¹,², DVM, PhD

¹Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States
²PRISM Health Research Group, Rollins School of Public Health, Emory University, Atlanta, GA, United States
³Division of HIV Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States

Corresponding Author:
Brady W Bennett, MPH
Department of Epidemiology
Rollins School of Public Health
Emory University
1518 Clifton Road
Atlanta, GA, 30322
United States
Phone: 1 706 766 1510
Email: brady.bennett@emory.edu

Abstract

Background: Male-to-male sexual transmission continues to account for the greatest proportion of new HIV diagnoses in the United States. However, calculating population-specific surveillance metrics for HIV and other sexually transmitted infections requires regularly updated estimates of the number and proportion of men who have sex with men (MSM) in the United States, which are not collected by census surveys.

Objective: The purpose of this analysis was to estimate the number and percentage of MSM in the United States from population-based surveys.

Methods: We used data from 5 population-based surveys to calculate weighted estimates of the proportion of MSM in the United States and pooled these estimates using meta-analytic procedures. We estimated the proportion of MSM using sexual behavior–based questions (encompassing anal or oral sex) for 3 recall periods—past 12 months, past 5 years, and lifetime. In addition, we estimated the proportion of MSM using self-reported identity and attraction survey responses. The total number of MSM and non-MSM in the United States were calculated from estimates of the percentage of MSM who reported sex with another man in the past 12 months.

Results: The percentage of MSM varied by recall period: 3.3% (95% CI 1.7%-4.9%) indicated sex with another male in the past 12 months, 4.7% (95% CI 0.0%-33.8%) in the past 5 years, and 6.2% (95% CI 2.9%-9.5%) in their lifetime. There were comparable percentages of men who identified as gay or bisexual (3.4%, 95% CI 2.2%-4.6%) or who indicated that they are attracted to other men (4.9%, 95% CI 3.1%-6.7%) based on pooled estimates. Our estimate of the total number of MSM in the United States is 4,230,000 (95% CI 2,179,000-6,281,000) based on the history of recent sexual behavior (sex with another man in the past 12 months).

Conclusions: We calculated the pooled percentage and number of MSM in the United States from a meta-analysis of population-based surveys collected from 2017 to 2021. These estimates update and expand upon those derived from the Centers for Disease Control and Prevention in 2012 by including estimates of the percentage of MSM based on sexual identity and sexual attraction. The percentage and number of MSM in the United States is an important indicator for calculating population-specific disease rates and eligibility for preventive interventions such as pre-exposure prophylaxis.

(JMIR Public Health Surveill 2024;10:e56643) doi:10.2196/56643
KEYWORDS
sexual behavior; sexual identity; sexual attraction; men who have sex with men; population estimates; MSM; men who have sex with other men; national surveys; census; United States

Introduction
Men have been the epicenter of HIV infections in the United States since the beginning of the epidemic, accounting for the largest proportion of new cases each year [1]. Most new HIV infections are among gay, bisexual, and other men who have sex with men (MSM) [1]. In 2021, 67% of new HIV infections in the United States were among MSM [2]. To better understand national rates of HIV infections, medication adherence, and preventative measures among MSM, it is imperative to have accurate estimates of the proportion and number of MSM in the United States [3]. However, the most recent published national estimates for MSM populations were published over a decade ago [4].

There is currently no census-based question on same-sex behavior that yields data to estimate a national proportion or count of MSM in the United States, although there is 1 question about being part of a same-sex household [5]. However, these data exclude MSM who are not domiciled with a male partner or who decline to report that their housemate is a partner. Therefore, previous estimations of MSM in the United States have focused on periodically conducted representative surveys, such as the General Social Survey (GSS) [6] or National Health and Nutrition Examination Survey (NHANES) [7]. In 2012, Centers for Disease Control and Prevention scientists used meta-analysis [4] to determine the population size of MSM in the United States and calculate disease rates and rate ratios for HIV and primary and secondary syphilis [4]. Subsequent analyses by Grey et al [8] extended national estimates of MSM to smaller area estimates such as states and counties.

These past analyses focused on sexual behaviors to determine MSM status. Because some national surveys ask about sexual orientation or attraction, it also is possible to examine prevalence using orientation or attraction and consider the concordance between self-reported sexual behavior and self-reported sexual identity or attraction. Currently, the GSS, NHANES, and the National Survey of Family Growth (NSFG) [9] contain both sexual behavior and sexual identity or attraction questions. These 2 measures are not always concordant, and the concordance can vary by age and race [10,11]. It has been argued that for some public health uses of estimates such as estimating populations of MSM in need of HIV prevention services or testing, behavior might be a better indication of need than identity [8]. However, willingness to report a same-sex orientation may affect willingness to seek prevention services.

It has been over a decade since Purcell et al [4] published their national estimates of MSM, and in that time, it has been cited over 200 times [12]. However, there have also been major policy decisions, such as the 2015 Supreme Court decision affirming the right for same-sex couples to marry in Obergefell v. Hodges that may affect our understanding of the proportion of the population who are MSM. Since the landmark 2015 decision, public support for marriage equality has grown consistently year-over-year, indicating larger acceptance of lesbian, gay, bisexual, transgender, queer or questioning, intersex, and asexual persons [13]. Therefore, we sought to update national behavior prevalence estimates of MSM and explore whether there are enough data to provide estimates of MSM based on orientation or attraction. Following the methods of Purcell et al [4], we used population-based surveys and meta-analysis to estimate separate estimates of each survey and year combination and then aggregate estimates into single, nationally representative proportions of MSM in the United States.

Methods

Study Selection
To determine which population-based surveys to include in our analysis, we started with the selection methods used by Purcell et al [4]. We examined the sources used in the previous estimates to identify whether the data collection was ongoing. We then completed a literature search using PubMed (National Institutes of Health) and Google Scholar to identify additional sources that include questions about sexual behavior, sexual identity, or attraction. We searched using key terms for measurement (prevalence and estimation), male-to-male sexual behavior or identity (MSM, male-to-male sexual contact, gay, and bisexual), geography (United States), and survey (population-based survey). Abstracts were screened by the lead author (BWB) in consultation with the coauthors for data sources that used population-based surveys for estimation.

Studies eligible for inclusion were population-based surveys with available complex survey methodology documentation to allow weighting to obtain estimated population proportions of MSM and 95% CIs. Surveys that recruited predominantly at sites that are likely to have a high proportion of MSM (ie, surveys among HIV-infected persons and surveys at sexually transmitted infection clinics) were excluded to minimize overestimation. We identified 4 ongoing population-based surveys from Purcell et al [4] that currently include questions about sexual behavior, sexual identity, or attraction: the GSS [6], NHANES [7], NSFG [9], and National Survey on Drug Use and Health [14]. In addition to this, we also identified the National Health Interview Survey [15] as another eligible study (Table 1). Gallup [16] and the US Census Pulse Household Survey [17] also estimate the proportion of men who identify as gay or bisexual; however, we were unable to gain access to CIs for reported point estimates and, therefore, excluded these data from our analysis.
To determine the study-level prevalence of MSM by sexual behavior, we considered questions about anal or oral sex with another male for 3 time periods: lifetime, in the past 5 years, and past 12 months. Survey-specific questions are outlined in Table S1 in Multimedia Appendix 1. For sexual identity, we categorized men who indicated that they are “gay,” “homosexual,” or “bisexual” best described them to be MSM. For sexual attraction, we considered men who indicated that they are “equally attracted to males and females,” “mostly attracted to the same sex,” or “only attracted to the same sex” to be MSM. Data were analyzed accounting for the complex sample design [18]; proportions of MSM and variances were estimated for each behavioral domain, for attraction and for orientation.

### Meta-Analysis

We applied Rao et al’s [19] meta-analytic method to pool survey-specific results into a single estimate with confidence bounds. First, for each recall period (eg, lifetime sex, past 5 years, and past 12 months) and for each identity or attraction category, we multiplied the population-level prevalence by the inverse of its variance. Then, we summed these weighted prevalence estimates across studies and then divided by the sum of the weights. Because surveys included in our analysis were conducted over several years and with differing sample designs and age ranges, we included a corresponding between-studies variance term [20] before deriving the overall prevalence estimates.

We examined heterogeneity of prevalence estimates across surveys using the Q-statistic [19] and Higgins I index [21]. Pooled estimates for the overall prevalence of MSM were based on random-effects models, which provide a more conservative estimate of the variance, generating potentially more accurate
inferences about a population of studies beyond what we present in this analysis [4]. All meta-analytic calculations were completed in R (R Foundation for Statistical Computing) using the “meta” package [22].

Calculation of Count of MSM and Other Men in the United States

To calculate the total number of MSM and non-MSM, aged 18 years and older, in the United States, we took our recently derived past 12-month estimate of the proportion of MSM and its 95% CI and multiplied it by the 2022 population estimate of men aged 18 years and older in the United States from the US Census Bureau [23]. The number of MSM was then subtracted from the total estimated number of men aged 18 years and older in the United States to compute the population size of other men.

Ethical Considerations

This project did not require review by an internal review board because it did not include human subjects, nor was it a clinical investigation as defined by federal regulations [24].

Results

MSM Estimates Based on Sexual Behavior

The estimated proportions of men who have sex with other men, by recall period, for each of the included population-based surveys that contained sexual behavior–based questions and the pooled estimates obtained from the meta-analysis are presented in Table 2. The pooled estimates were 3.3% (95% CI 1.7%-4.9%) in the last 12 months, 4.7% (95% CI 0.0%-33.8%) in the last 5 years, and 6.2% (95% CI 2.9%-9.5%) in their lifetime. The tests for heterogeneity were significant for both the “past 5 years” and “lifetime sex” recall periods but not for the past 12 months recall period.

Table 2. Estimated proportion of men who have sex with men for individual studies and combined meta-analysis by behavior, NSFG, NHANES, and GSS surveys, United States, 2017-2021.

| Study name | Time period | Sample size | Estimated prevalence (%; 95% CI) | Q statistic | P value | I² (%)
|------------|-------------|-------------|---------------------------------|-------------|---------|-------
| Past 12 months | | | | | | |
| GSS | 2018 | 502 | 3.3 (1.7-4.9) | 6.8 | .08 | 55.6 |
| GSS | 2021 | 809 | 2.9 (1.4-4.4) | | | |
| NSFG | 2017-2019 | 5206 | 3.1 (2.4-3.8) | | | |
| NHANES | 2017-2020 | 3338 | 2.7 (1.9-3.4) | | | |
| Past 5 years | | | | | | |
| GSS | 2018 | 559 | 4.7 (0.0-33.8) | 14.4 | <.001 | 93.1 |
| GSS | 2021 | 904 | 2.5 (1.2-3.8) | | | |
| Lifetime sex | | | | | | |
| GSS | 2018 | 500 | 6.2 (2.9-9.5) | 25.9 | <.001 | 88.4 |
| GSS | 2021 | 807 | 3.1 (1.6-4.6) | | | |
| NSFG | 2017-2019 | 5163 | 7.0 (6.9-7.0) | | | |
| NHANES | 2017-2020 | 3338 | 6.9 (5.6-8.2) | | | |

aNSFG: National Survey of Family Growth.
bNHANES: National Health and Nutrition Examination Survey.
cGSS: General Social Survey.
dQ-statistic follows a chi-square distribution to determine the presence or absence of heterogeneity in a set of studies in a meta-analysis.
eI² quantifies the degree of heterogeneity in a meta-analysis.

MSM Estimates Based on Sexual Identity and Sexual Attraction

Table 3 contains the estimated prevalence and 95% CIs, along with the pooled aggregate results, by self-reported sexual identity and attraction. The pooled estimated prevalence of gay and bisexual men by sexual identity was 3.4% (95% CI 2.2%-4.6%), which comes from 8 collection cycles across 5 different surveys. Both tests for heterogeneity (Q-statistic and I²) indicated heterogeneity across surveys (I²=95.4%; Q=152.4; P<.001). The pooled estimated prevalence of gay and bisexual men by sexual attraction was 4.9% (95% CI 3.1%-6.7%), with significant heterogeneity across the 3 collection periods from 2 surveys (I²=79.2; Q=9.6; P=.008). We calculated pooled estimates of the proportion of gay and bisexual men by including a broader definition that includes men who responded that they are “mostly attracted to women.” This broader definition resulted in a proportion of MSM of 9.3%, which is substantially higher than any other proportion calculated from this analysis. Therefore, for the purpose of this paper, we used the stricter definition of MSM based on attraction (ie, men who indicated...
that they are “equally attracted to males and females,” “mostly attracted to the same sex,” or “only attracted to the same sex.”

Our data indicate limited fluidity in respondents’ identification and attraction. For the GSS, which has both sexual identity or attraction questions and sexual behavior questions, approximately 0.6% of straight-identifying men in 2018 sample and 2.6% in the 2021 sample also reported anal or oral sex with another man in the past 5 years. Similarly, for the 2017-2019 NSFG, of men who identified as straight, 4.4%-5.9% also reported anal or oral sex with another man in their lifetime and 0.9%-2.2% reported anal or oral sex in the past 12 months.

Table 4 shows the estimated number of MSM and non-MSM men aged 18 years and older in the United States. These totals were derived by multiplying the proportion of MSM from the pooled estimate for sexual behavior in the past 12 months (95% CI 1.7-4.9) by the total number of men aged 18 years and older in the United States [25]. We estimated that there were 4,230,000 MSM (95% CI 2,179,000-6,281,000) aged ≥18 years in the United States.

Table 3. Estimated proportion of men who identify as gay or bisexual for individual studies and combined meta-analysis by identity and attraction (GSS\textsuperscript{a}, NSDUH\textsuperscript{b}, NSFG\textsuperscript{c}, NHANES\textsuperscript{d}, and NHIS\textsuperscript{e} surveys, United States, 2017-2021).

<table>
<thead>
<tr>
<th>Study name</th>
<th>Time period</th>
<th>Sample size</th>
<th>Estimated prevalence (%; 95% CI)</th>
<th>Q statistic</th>
<th>P value</th>
<th>$I^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSS</td>
<td>2018</td>
<td>629</td>
<td>3.4 (2.2-4.6)</td>
<td>152.4</td>
<td>&lt;.001</td>
<td>95.4</td>
</tr>
<tr>
<td>GSS</td>
<td>2021</td>
<td>1012</td>
<td>6.4 (4.7-8.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSDUH</td>
<td>2018</td>
<td>20,162</td>
<td>4.6 (4.1-5.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSDUH</td>
<td>2019</td>
<td>19,923</td>
<td>4.2 (3.8-4.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSFG\textsuperscript{f}</td>
<td>2017-2019</td>
<td>5206</td>
<td>2.5 (1.9-3.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSFG\textsuperscript{f}</td>
<td>2017-2019</td>
<td>5206</td>
<td>2.6 (1.9-3.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHIS</td>
<td>2019</td>
<td>14,733</td>
<td>2.1 (1.8-2.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHANES</td>
<td>2017-2020</td>
<td>2595</td>
<td>4.6 (3.5-5.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSDUH</td>
<td>2018</td>
<td>20,162</td>
<td>5.3 (4.8-5.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSDUH</td>
<td>2019</td>
<td>19,923</td>
<td>5.3 (4.9-5.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSFG</td>
<td>2017-2019</td>
<td>5206</td>
<td>4.0 (3.2-4.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}GSS: General Social Survey.
\textsuperscript{b}NSDUH: National Household Survey on Drug Use and Health.
\textsuperscript{c}NSFG: National Survey of Family Growth.
\textsuperscript{d}NHANES: National Health and Nutrition Examination Survey.
\textsuperscript{e}NHIS: National Health Interview Survey.
\textsuperscript{f}The NSFG contains 2 different questions about a respondent’s sexual identity; therefore, we included both estimates for 2017-2019.

Table 4. Population size of men aged 18 years and older in the United States, 2022 and number of MSM\textsuperscript{a} and non-MSM using past 12 months proportion estimate of MSM from meta-analysis.

<table>
<thead>
<tr>
<th>Population estimation (95% CI)</th>
<th>MSM proportion (%)</th>
<th>MSM population size, n</th>
<th>Non-MSM population proportion (%)</th>
<th>Non-MSM population size, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3 (1.7-4.9)</td>
<td>4,230,000 (2,179,000-6,281,000)</td>
<td>96.7 (95.1-98.3)</td>
<td>123,945,000 (121,894,000-125,996,000)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}MSM: men who have sex with men.

Discussion

Principal Findings

Our meta-analysis of data from 5 population-based surveys indicated that 3.3% of the US male population report recent sex with men, and 3.4%-4.9% of the US male population report identifying as gay, bisexual, or being sexual attracted to men. We found substantial overlap between estimates based on behavioral measures and measures based on orientation or attraction. Our behavioral estimates are all within 1 percentage point of the Centers for Disease Control and Prevention’s 3 behavior estimates from over a decade ago [4]. Additionally, compared to this analysis, they include more national surveys that asked about sexual behavior or identification [4]. Although there is growing interest in understanding the experiences and risks of MSM, there is a smaller number of ongoing population-based studies that examine these issues.
Previous estimates by Purcell et al [4] have also been used extensively in calculating the rates of disease among MSM. We hope that these updated percentages of MSM will be used in future analyses to calculate the rates of health states and conditions among MSM. For example, time periods for syphilis diagnoses can be matched to time periods for the estimated MSM population size. Then, the count of syphilis diagnoses among MSM can be divided by the estimated population size of MSM for that same recall period and transformed into rates per 100,000 population.

For our calculation of the total number of MSM in the United States, we chose to use the prevalence of male-male sex during the past 12 months. This differs from the previous paper by Purcell et al [4], which used the estimated prevalence from self-reported sex with a man in the past 5 years. Our decision was both practical and methodical. \( Q \)-statistics and \( P \) for the pooled “past 5 years” estimates indicated heterogeneity across surveys (\( Q=14.4, P=.001; I^2=93.1\% \)). In addition, our “past 5 years” estimates also had very wide 95% CIs (0.0-33.8), indicating a substantial amount of instability in the estimate. Out of the behavior-based estimates in our analysis, self-reported sex in the past 12 months had the lowest heterogeneity. In addition to considering current risk for sexually transmitted infections, using an estimate based upon recent sexual activity may allow us to enumerate those MSM who are currently most at risk for disease acquisition or most appropriate for prevention services.

Estimates from “past 5 years” or “lifetime” behavioral data might be used in other circumstances. For example, Purcell et al [4] use of the “past 5 years” measure to estimate number of MSM in the United States in 2012 was designed to include a broader section of MSM, including those who were not recently sexually active. Such denominators might be relevant for characterizing the impact of MSM on other types of diseases, depending on the timelines for exposure and pathogenesis. Having separate behavioral measurements with different recall periods allow users to determine which estimate makes the most sense for their public health purposes.

We expanded on previous work by calculating the prevalence of self-reported sexual identity and attraction to update and to understand the percentage of MSM in the United States. Identity and attraction are neither mutually exclusive nor perfectly concordant, and they do not necessarily overlap with behavior. The term MSM specifically refers to sexual behavior, but identity and attraction refer to how men classify themselves—an important distinction for epidemiologists and sexual health researchers who seek to avert disease transmission associated with male-male sex. Identity and attraction are not risk factors for sexually transmitted infections but may be proxies for current or future risks and for prevention service needs. However, as our data show, identity and attraction are not necessarily specific or sensitive [26]: a person who identifies as straight may still engage in anal or oral sex with another man, a man identifying as gay or bisexual may have anal or vaginal sex with a female. Providing additional prevalence estimates for orientation and attraction gives users more choices to fit their needs when trying to estimate these proportion of men who identify as gay or bisexual or are attracted to men, based on their programmatic goal or research question. Furthermore, including this new estimate of sexual identity and attraction aligns with new reports form the US Census Bureau, which began collecting data on sexual orientation in 2021. According to these data, 6.5% of cisgendered US men reported identifying as gay or bisexual [27].

Here, we calculated the 3 updated prevalence estimates for MSM, a new estimate for gay or bisexual men, and a new estimate for men who are attracted to other men. Going forward, we plan to update these estimates at regular intervals and to share public use data sets through AIDSVu [28]. We propose to update these estimates every 5 years, given the pace at which the national population-based surveys are implemented and published. We do not anticipate a substantial change in the estimates in the coming 5 years. Comparing the newly calculated “past 12 months” estimate to the “past 12 months” estimate from Purcell et al [4], we found a relatively small change over a decade of time (3.3% vs 2.9%, respectively). However, the visibility of sexual behavior and identity in our culture has shifted rapidly over the past few decades, and regular updated analyses will allow us to identify any unexpected shifts should they emerge and to provide estimates that always include the most recently collected data.

Limitations

This analysis has important limitations. First, meta-analyses rely on the strengths of the underlying studies. The surveys that we used are minimally biased in terms of sampling—5 population-based surveys—but they have varying sample sizes of MSM (which comprise a small subpopulation of the sampled population). Thus, the individual survey estimates that feed into our pooled estimates have varying confidence limits around their point estimates. Furthermore, these surveys have varying age ranges for inclusion in the survey and thus may not be exactly transferrable across surveys. Similar to Purcell et al [4], we were unable to directly calculate stratified estimates of the proportion of MSM by race, ethnicity, or age because of small sample sizes within stratifications.

There is also the potential for misclassification of a respondent’s true sexual behavior, identity, or attraction, leading to an underestimation of the count and percentage of MSM in the United States from our calculation. A recent analysis in Canada found that 13.5% of the gay and bisexual male population (based upon sexual behavior) reported being unlikely to disclose their sexual identity on government surveys [29]. However, it is unclear whether these estimates of misclassification are transferrable to the US population. Additionally, although we provide estimated percentages of MSM by sexual attraction and sexual identity, our final estimation of the total MSM population for the United States was solely based on sexual behavior (anal or oral sex) in the past 12 months rather than identity or attraction alone or a combination of identity or attraction and behavior.

Finally, we did not adjust for differing age ranges between individual surveys. This may have implications for our outcomes; however, only 1 survey, the NSFG, included men younger than 18 years in their estimate, and the number of
respondents younger than 18 years old who answered that they had ever had sex with another man was <0.5% of the total respondents for NSFG and is, therefore, unlikely to have a significant impact on the results. Only the NSFG included data from men aged younger than 18 years; all other estimates for sexual behavior or identity or attraction used men older than 18 years of age.

Conclusions
The National HIV or AIDS strategy and the Ending the HIV Epidemic Initiative in the United States each highlight the prevention of HIV among MSM as key to meeting the goal of decreasing new HIV infections by 90% by 2030 [30]. Developing updated, accurate estimates of this population at risk is critical to better understanding the disproportionate burden of HIV and risk for HIV among MSM. Updated estimates of the MSM population sizes can help to guide resource allocation and programmatic efforts and support key benchmarks for progress as we seek to end the HIV epidemic and other health concerns that disproportionately impact MSM. Providing public, updated estimates of behaviors across different time periods and data on orientation and attraction can offer choices to researchers and health departments seeking to serve these populations.

Acknowledgments
This work was supported by the Center for AIDS Research at Emory University (P30 AI050409). The authors acknowledge receipt of financial support for this project from Gilead Sciences. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Data Availability
Data from General Social Survey, National Survey of Family Growth, National Health Interview Survey, and National Household Survey on Drug Use and Health are publicly available at their respective survey websites [4,6,14,15]. The 2017 to 2020 National Health and Nutrition Examination Survey sexual behavior data are not publicly available due to being “limited access data,” containing sensitive private information, but they can be accessed at a National Center for Health Statistics Research Data Center [7].

Conflicts of Interest
PSS reports a relationship with Merck & Co that includes consulting and advisory and funding grants.

Multimedia Appendix 1
Survey-specific questions.
[DOCX File, 20 KB - publichealth_v10i1e56643_app1.docx ]

References


13. Fields A. Public support for marriage equality at all-time high approaching the anniversaries of landmark Supreme Court decisions that contributed to LGBTQ+ equality. Human Rights Campain. 2022. URL: https://tinyurl.com/wj2emmb [accessed 2024-05-14]


Abbreviations

GSS: General Social Survey
MM: men who have sex with men
NHANES: National Health and Nutrition Examination Survey
NSFG: National Survey of Family Growth
Corrigenda and Addenda

Correction: Dashboard With Bump Charts to Visualize the Changes in the Rankings of Leading Causes of Death According to Two Lists: National Population-Based Time-Series Cross-Sectional Study

Shu-Yu Tai, MD, DPhil; Ying-Chen Chi, DPhil; Yu-Wen Chien, MD, DPhil; Ichiro Kawachi, MD, DPhil; Tsung-Hsueh Lu, MD, DPhil

1Department of Family Medicine, Kaohsiung Municipal Ta-Tung Hospital and Kaohsiung Medical University Hospital; School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan
2Department of Healthcare Information and Management, School of Health Technology, Ming Chuan University, Taoyuan, Taiwan
3Department of Public Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan
4Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, United States

Corresponding Author:
Tsung-Hsueh Lu, MD, DPhil
Department of Public Health
College of Medicine
National Cheng Kung University
No. 1, Dah Hsueh Road, East District
Tainan, 701
Taiwan
Phone: 886 0928389971
Fax: 886 6 2359033
Email: robertlu@mail.ncku.edu.tw

Related Article:
Correction of: https://publichealth.jmir.org/2023/1/e42149

In “Dashboard With Bump Charts to Visualize the Changes in the Rankings of Leading Causes of Death According to Two Lists: National Population-Based Time-Series Cross-Sectional Study” (JMIR Public Health Surveill 2023;9(1):e42149) the authors noted one error.

In the originally published article, the first affiliation incorrectly appeared as:

Department of Family Medicine, Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung, Taiwan.

This has been changed to the following:

Department of Family Medicine, Kaohsiung Municipal Ta-Tung Hospital and Kaohsiung Medical University Hospital; School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan.

The correction will appear in the online version of the paper on the JMIR Publications website on January 9, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Correction: Comparison of Four Active SARS-CoV-2 Surveillance Strategies in Representative Population Sample Points: Two-Factor Factorial Randomized Controlled Trial

Andreas Deckert¹, PhD; Simon Anders², Prof Dr; Ivonne Morales³, PhD; Manuela De Allegri¹, Prof Dr; Hoa Thi Nguyen¹, PhD; Aurélia Souares¹, PhD; Shannon McMahon¹, PhD; Matthias Meurer², Dipl Ing; Robin Burk², PhD; Dan Lou², PhD; Lucia Brugnara⁴, MSc; Matthias Sand⁵, PhD; Lisa Koeppel⁶, PhD; Lena Maier-Hein⁶, Prof Dr; Tobias Ross⁶; Tim J Adler⁶, PhD; Stephan Brenner¹, PhD; Christopher Dyer⁷, PhD; Konrad Herbst², PhD; Svetlana Ovchinnikova², PhD; Michael Marx⁴, Prof Dr; Paul Schnitzler⁸, Prof Dr; Michael Knop², Prof Dr; Till Bärnighausen¹, Prof Dr; Claudia M Denkinger³, PD Dr

¹Heidelberg Institute of Global Health, Heidelberg, Germany
²Center for Molecular Biology Heidelberg, Heidelberg, Germany
³Division of Infectious Disease and Tropical Medicine, Heidelberg University Hospital, Heidelberg, Germany
⁴evaplan GmbH at the University Hospital, Heidelberg, Germany
⁵GESIS Leibniz-Institute for the Social Sciences, Mannheim, Germany
⁶Division of Computer Assisted Medical Interventions, German Cancer Research Centre, Heidelberg, Germany
⁷Gesundheitsamt des Rhein-Neckar-Kreises, Heidelberg, Germany
⁸Center of Infectious Diseases, Virology, Heidelberg University Hospital, Heidelberg, Germany

Corresponding Author:
Andreas Deckert, PhD
Heidelberg Institute of Global Health
INF 324
Heidelberg, 69120
Germany
Phone: 49 6221 4537260
Email: a.deckert@uni-heidelberg.de

Related Article:
Correction of: https://publichealth.jmir.org/2023/1/e44204
doi:10.2196/57203

In “Comparison of Four Active SARS-CoV-2 Surveillance Strategies in Representative Population Sample Points: Two-Factor Factorial Randomized Controlled Trial” (JMIR Public Health Surveill 2023;9:e44204) the authors noted one error.

In the originally published manuscript, the following sentence appeared in the “Study Design” section:

The 4 study arms represented all combinations of the two factors: (i) testing without condition (A) versus testing under the condition of upstream COVID-19 symptom prescreening (B) and (ii) testing individuals (1) versus households (2).

This sentence has been corrected to:

The 4 study arms represented all combinations of two factors: i) testing unconditional (A) versus testing under the condition of upstream COVID-19 symptom prescreening (B), and ii) testing individuals (1) versus households (2).

The correction will appear in the online version of the paper on the JMIR Publications website on February 16, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Corrigenda and Addenda

Corrigendum to: COVID-19 Vaccine Hesitancy: Umbrella Review of Systematic Reviews and Meta-Analysis

Tiffany Leung¹, MPH, MD; Travis Sanchez², MPH, DVM; Gunther Eysenbach¹, MPH, MD

¹JMIR Publications, Toronto, ON, Canada
²Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Tiffany Leung, MPH, MD
JMIR Publications
130 Queens Quay East, Unit 1100
Toronto, ON, M5A 0P6
Canada
Phone: 1 4165832040
Email: tiffany.leung@jmir.org

Related Article:
Correction of: https://publichealth.jmir.org/2024/1/e54769
doi:10.2196/64080

Shortly after the publication of the article titled “COVID-19 Vaccine Hesitancy: Umbrella Review of Systematic Reviews and Meta-Analysis” in JMIR Public Health and Surveillance [1], the editors discovered processing errors that occurred during the peer review process. Specifically, Roy Rillera Marzo, MD, acted as guest editor for the theme issue titled “Preventive Strategies: Scaling Up Effective Public Health Interventions for Long-Term Population Health Benefits” and made a decision for this paper where they are also a coauthor, which is against the Committee on Publication Ethics (COPE) [2] and JMIR Publications’ Conflict of Interest policies [3].

As a result, the paper underwent additional post-publication editorial review to address the question whether the published version should be retracted or can stand. During the post-publication review process, the editor in chief of JMIR Public Health and Surveillance, Travis Sanchez, DVM, MPH, reviewed the manuscript drafts, peer review comments, and author responses. He noted that the paper adheres to standards for systematic reviews. There were two reviews on record, including one peer review that identified methodological issues that the authors addressed in a subsequent revision.

After the first round of reviews, guest editor Marzo did not adhere to JMIR Publications’ best practices and decision-making guidelines for editors by issuing a “B” decision (ie, minor revisions, no further peer review required) instead of a “C” decision (ie, revise and re-review), which would have presented the revised manuscript to the peer reviewers for re-review [4]. Upon editorial review of the revised manuscript and the authors’ point-by-point responses to the peer reviewers’ comments, it was noted that there could have been additional clarifications in the author responses, including more explicit comments on impacts of the changes made on the results. However, on examination of the edits that the authors made in the final draft of the manuscript by the editor in chief, it appears that the authors appropriately addressed the most substantive methodology concerns noted by the peer reviewers. Although minor revisions or refinements could have been made, the validity of the study appears to be sound and would not be impacted by minor revisions. No further revisions, corrections, or addenda are required at this time.

As a result of our investigations and a de novo evaluation of the acceptability of the paper by the editor in chief Travis Sanchez, the determination is that the editor of record for this manuscript has been corrected to Travis Sanchez. The editor information in the published paper will be changed from:

Edited by R Marzo

to:

Edited by T Sanchez

At the time of this manuscript publication, Marzo had not made other decisions for this theme issue. Henceforward, Marzo and other guest editors will only make future decisions for this theme issue together with editor in chief Sanchez.

JMIR Publications adheres to COPE guidance, including the 2023 published guidance on “Best practices for guest edited collections” [2]. JMIR Publications also routinely reviews and updates its Conflict of Interest disclosure policies as part of a larger effort to promote integrity of the published scientific record [3]. JMIR Publications has adhered to all key points from the COPE guidance on guest edited collections, and processes have been further optimized to mitigate risks and prevent further similar occurrences.
The correction of the editor’s name will appear in the online version of the paper on the JMIR Publications website on July 10, 2024, together with the publication of this correction notice. Because this correction was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

References
2. Committee on Publication Ethics. Best practices for guest edited collections. 2023. URL: https://doi.org/10.24318/7cKLaia0 [accessed 2024-06-10]

©Tiffany Leung, Travis Sanchez, Gunther Eysenbach. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 10.07.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Letter to the Editor

The Need for a Bleed Type–Specific Annual Bleeding Rate in Hemophilia Studies

Kun Huang, MD
Department of Pediatrics, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

Corresponding Author:
Kun Huang, MD
Department of Pediatrics
The First Affiliated Hospital
Zhejiang University School of Medicine
1367 Wenyi West Road
Yuhang District
Hangzhou, 311121
China
Phone: 86 166 1994 4294
Email: aerohit@163.com

Related Articles:
Comment on: https://publichealth.jmir.org/2023/1/e45747

Comment in: https://publichealth.jmir.org/2024/1/e54756

(JMIR Public Health Surveill 2024;10:e51372) doi:10.2196/51372

KEYWORDS
benefit-risk assessment; discrete choice experiment; hemophilia A; patient preference; prophylactic treatment

We read with great interest the paper “Quantifying benefit-risk trade-offs toward prophylactic treatment among adult patients with hemophilia A in China: discrete choice experiment study” by Wang et al [1]. The results were unsurprising and aligned with clinical practice and provided valuable insights into the preferences of patients for prophylaxis in hemophilia A treatment. The authors designated the annual bleeding rate as an important item in the questionnaire used. This is rational as most clinical studies on hemophilia consider it a primary clinical outcome. However, there is a need for greater specificity in the annual bleeding rate, particularly with regard to joint bleeding and spontaneous bleeding.

As the standard treatment for patients with hemophilia A, prophylaxis with coagulation factor VIII (FVIII) was proposed to maintain a steady trough FVIII level mainly to prevent spontaneous and breakthrough bleeds [2]. In the study, the annual bleeding rate was calculated using all types of bleeds, including injury-caused, spontaneous, joint, and subcutaneous bleeds. However, a great difference exists across these bleeding types. For example, compared with subcutaneous bleeds, joint bleeds typically result in more serious consequences and require greater attention and more effective management. The definition of “target joint” only needs 3 spontaneous bleeds to occur in a single joint [2]. However, patients with target joints will need a substantially higher FVIII level to ensure their safety [3]. On the other hand, it is common for patients with active lifestyles to report more mild injury-caused bleeds (e.g., bleeding resulting from a sports-related injury) but a good quality of life and normal joint structure or function [4]. As reported before, the primary aim of prophylaxis is to prevent spontaneous bleeds, which could be completely different from injury-caused bleeds [2]. All individuals can bleed if an injury occurs while taking part in sports; thus, sports-related bleeds are not equivalent to spontaneous bleeds [5]. For example, there are substantial differences in treatment or clinical outcome between patients with 6 instances of spontaneous joint bleeds and those with 6 instances of injury-caused bleeds in 1 year. Thus, Wang et al [1] could consider distinguishing bleed types, at least for spontaneous, injury-caused, joint, and other mild bleeds. With reference to the attributes of efficacy factor among the authors’ citations, 3 studies (references 22, 23, and 25) picked reduction of bleeds and 2 studies (references 25 and 26) selected breakthrough bleeds instead of annual bleeding rate. In 2 studies that did not involve specific bleed types, one (reference 28) added joint evaluation to make the results clearer.

In conclusion, although the annual bleeding rate is a commonly used indicator for clinical outcomes among patients with hemophilia, it is better to distinguish different kinds of bleeds or simply choose a more detailed and suitable one for analysis in a discrete choice experiment. This will eliminate potential
confusion during the interpretation of results and make study findings and conclusions more detailed and valid, thus enhancing their clinical application.

Conflicts of Interest
None declared.

References

Abbreviations
FVIII: factor VIII

©Kun Huang. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 13.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Authors’ Reply: “The Need for a Bleed Type–Specific Annual Bleeding Rate in Hemophilia Studies”

Limin Wang1,2,3, MPH; Shimeng Liu4,5, PhD; Shan Jiang6, PhD; Chaofan Li1,2,3, PhD; Liyong Lu1,2,3, PhD; Yunhai Fang7, PhD; Shunping Li1,2,3, PhD

1Centre for Health Management and Policy Research, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China
2NHC Key Laboratory of Health Economics and Policy Research, Shandong University, Jinan, China
3Centre for Health Preference Research, Shandong University, Jinan, China
4School of Public Health, Fudan University, Shanghai, China
5NHC Key Laboratory of Health Technology Assessment, Fudan University, Shanghai, China
6Macquarie University Centre for the Health Economy, Macquarie Business School and Australian Institute of Health Innovation, Macquarie University, Sydney, Australia
7Shandong Hemophilia Treatment Center, Shandong Blood Center, Jinan, China

Corresponding Author:
Shunping Li, PhD
Centre for Health Management and Policy Research
School of Public Health, Cheeloo College of Medicine
Shandong University
44 Wenhuaxi Road
Jinan, 250014
China
Phone: 86 13188934998
Fax: 86 053188380161
Email: lishunping@sdu.edu.cn

Related Articles:
Comment on: http://www.jmir.org/2024/1/e51372/
Comment on: http://www.jmir.org/2023/1/e45747/

(JMIR Public Health Surveill 2024;10:e54756) doi:10.2196/54756

KEYWORDS
benefit-risk assessment; discrete choice experiment; hemophilia A; patient preference; prophylactic treatment

We appreciate Huang’s [1] thoughtful consideration of our article [2]. Their suggestion to refine the attribute of “annual bleeding rate” in our discrete choice experiment (DCE) by distinguishing between spontaneous versus injury-induced bleeding, and further categorizing bleeds as joint or mild, is acknowledged with appreciation. In the formulation of our DCE questionnaire, considerable attention was devoted to addressing this aspect. However, the integration of a comprehensive spectrum of bleed types into the DCE was met with challenges, particularly within the context of the current hemophilia A scenario in China and the specific focus of our investigation.

Our study focused on adult patients with hemophilia. With the notable increase in the adoption of low-dose prophylactic treatment and advancements in diagnostic and therapeutic capabilities, there has been a significant rise in pediatric patients benefiting from prophylaxis with coagulation factor [3]. However, it is important to highlight that the current percentage of adult patients undergoing prophylaxis with coagulation factor remains relatively low [4]. In our DCE, the number of annual bleeds attribute was operationalized as spontaneous bleeding events. The decision not to delve into further details of bleeding types in the DCE was primarily influenced by the reasons listed below.

First, patients have limited ability to assess their own bleeding, and misrecording or missing records often occur [5]. Unlike during a clinical trial, it is challenging for them to obtain detailed descriptions of all types of bleeds per year via questionnaire. Second, while adult patients with hemophilia in developed countries engage in physical activities at rates comparable to the general population, the situation in China is markedly different [6]. With age, patients often exhibit varying degrees of joint damage or deformities, particularly in the knees, ankles,
and elbows, resulting in a generally reduced activity level [7]. Consequently, the incidence of activity- or injury-induced bleeding is relatively low. Moreover, a significant portion of adult patients in China may already be experiencing joint deformities or at a stage requiring joint replacement [4], rendering the notion of “target joints” less relevant.

In the DCE context, it is crucial that the chosen attribute and levels align with practical conditions, ensuring ease of investigation and data collection from the participants. Therefore, for this study, we selected the number of annual bleeds (spontaneous bleeding) as the attribute. Our DCE design, particularly in determining attributes and levels, entailed multiple rounds of consultations and interviews with clinical experts, including hemophilia specialists from the Shandong Hemophilia Treatment Center (the core center of the Hemophilia Treatment Center Collaborative Network of China and the World Federation of Hemophilia). These consultations were instrumental in providing valuable professional guidance.

We believe that these additional details provide a comprehensive background of our DCE design. Furthermore, our aspiration is to enhance access to prophylaxis with coagulation factor for adult patients in China, accompanied by a commitment to maintain precise bleeding records. This commitment would not only benefit our study but also pave the way for future research to refine the measurement of annual bleeding frequency by incorporating additional indicators like bleeding types.

**Conflicts of Interest**

None declared.

**References**

1. Huang K. The Need for a Bleed Type–Specific Annual Bleeding Rate in Hemophilia Studies. JMIR Public Health Surveill 2024. [doi: 10.2196/51372]


**Abbreviations**

DCE: discrete choice experiment
Generating Contextual Variables From Web-Based Data for Health Research: Tutorial on Web Scraping, Text Mining, and Spatial Overlay Analysis

Pablo Galvez-Hernandez1,2, PhD; Angelina Gonzalez-Viana3, PhD; Luis Gonzalez-de Paz4,5, PhD; Ketan Shankardass6,7, PhD; Carles Muntaner1,8, PhD

1Lawrence S Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada
2Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
3Public Health Agency of Catalonia, Health Department, Barcelona, Spain
4Primary Healthcare Transversal Research Group, Institut d’Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain
5Consorci d’Atenció Primària de Salut Barcelona Esquerra, Barcelona, Spain
6Department of Health Sciences, Wilfrid Laurier University, Waterloo, ON, Canada
7MAP Centre for Urban Health Solutions, Li Ka Shing Knowledge Institute, St Michael’s Hospital, Toronto, ON, Canada
8Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Pablo Galvez-Hernandez, PhD
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
Health Sciences Building, 4th Fl.
155 College St
Toronto, ON, M5T 3M6
Canada
Phone: 1 6475752195
Email: pau.galvez@utoronto.ca

Abstract

Background: Contextual variables that capture the characteristics of delimited geographic or jurisdictional areas are vital for health and social research. However, obtaining data sets with contextual-level data can be challenging in the absence of monitoring systems or public census data.

Objective: We describe and implement an 8-step method that combines web scraping, text mining, and spatial overlay analysis (WeTMS) to transform extensive text data from government websites into analyzable data sets containing contextual data for jurisdictional areas.

Methods: This tutorial describes the method and provides resources for its application by health and social researchers. We used this method to create data sets of health assets aimed at enhancing older adults’ social connections (eg, activities and resources such as walking groups and senior clubs) across the 374 health jurisdictions in Catalonia from 2015 to 2022. These assets are registered on a web-based government platform by local stakeholders from various health and nonhealth organizations as part of a national public health program. Steps 1 to 3 involved defining the variables of interest, identifying data sources, and using Python to extract information from 50,000 websites linked to the platform. Steps 4 to 6 comprised preprocessing the scraped text, defining new variables to classify health assets based on social connection constructs, analyzing word frequencies in titles and descriptions of the assets, creating topic-specific dictionaries, implementing a rule-based classifier in R, and verifying the results. Steps 7 and 8 integrate the spatial overlay analysis to determine the geographic location of each asset. We conducted a descriptive analysis of the data sets to report the characteristics of the assets identified and the patterns of asset registrations across areas.

Results: We identified and extracted data from 17,305 websites describing health assets. The titles and descriptions of the activities and resources contained 12,560 and 7301 unique words, respectively. After applying our classifier and spatial analysis algorithm, we generated 2 data sets containing 9546 health assets (5022 activities and 4524 resources) with the potential to enhance social connections among older adults. Stakeholders from 318 health jurisdictions registered identified assets on the platform.
between July 2015 and December 2022. The agreement rate between the classification algorithm and verified data sets ranged from 62.02% to 99.47% across variables. Leisure and skill development activities were the most prevalent (1844/5022, 36.72%). Leisure and cultural associations, such as social clubs for older adults, were the most common resources (878/4524, 19.41%). Health asset registration varied across areas, ranging between 0 and 263 activities and 0 and 265 resources.

**Conclusions:** The sequential use of WeTMS offers a robust method for generating data sets containing contextual-level variables from internet text data. This study can guide health and social researchers in efficiently generating ready-to-analyze data sets containing contextual variables.

(JMIR Public Health Surveill 2024;10:e50379) doi:10.2196/50379

**KEYWORDS**
web scraping; text mining; spatial overlay analysis; program evaluation; social environment; contextual variables; health assets; social connection; multilevel analysis; health services research

**Introduction**

**Background**

Contextual variables refer to the social or physical attributes of geographic or jurisdictional areas (eg, country, city, neighborhood, and administrative health area) that are not derived from the characteristics of their members [1]. Common examples include social cohesion [2], social capital [3], and presence of green spaces [4]. Contextual variables have multiple applications in health and social research. As people living in the same community or context are likely to be exposed to a similar environment, contextual variables can be used in multilevel models to explain variability in health outcomes [5].

Although information on some contextual variables, such as census data, is widely available, accessing context-level data in emerging research fields can pose significant challenges. For example, monitoring systems may not exist yet to fully capture the social determinants of health (SDOH) across delimited areas. In addition, there may not be data available on the exposure and implementation of large-scale interventions targeting SDOH, making program and implementation evaluation studies challenging or impossible [6]. This could be the case for regional or state public policies and public health programs, such as provincial public health programs to promote local intersectoral collaborations to tackle SDOH [7] or national legislation to promote healthy nutrition to prevent obesity [8]. As these policies and programs can be implemented without an evaluation plan, and data might be complex or unavailable, they often remain unevaluated [9].

When structured databases or primary data gathering are not feasible, the internet can be a valuable resource for compiling information to define contextual variables. However, this presents several challenges: internet data are often cluttered, fragmented, and spread over multiple websites [10]. Moreover, the content of most websites is not designed for use by health researchers nor is it grouped by relevant contextual areas. To overcome these challenges, we developed a novel 8-step method, which we have termed web scraping, text mining, and spatial overlay analysis (WeTMS) to collect large amounts of internet data from websites, transforming it into meaningful data sets containing research-relevant variables, and classifying them based on delimited geographical or jurisdictional areas.

This method combines the techniques used in web scraping, text processing and mining, and spatial analysis. Web scraping, also known as web data mining, involves the creation of programs that can automatically download, parse, organize, and store information collected from the web in structured data sets [11]. This process is more efficient and less prone to errors compared with the traditional and laborious process of manually copying and pasting internet information into a spreadsheet [11]. Web scraping has been gaining traction in health research, fueling the rise of infodemiology, which analyzes the spread and impact of web-based information to inform public health and policy [12]. As of January 2023, a search of the keyword “web scraping” in Medline yielded 105 records, 95 of which were published starting from 2019. Articles using web scraping in health and social research mostly used information from social media [13,14] (eg, Twitter, Instagram, and TikTok), forums [15,16], business and review websites [17], and news web pages [18]. Similarly, text mining has been increasingly applied in health and social research [19]. Text mining is the process of extracting meaningful information from large volumes of unstructured text data using techniques such as text classification, sentiment analysis, and pattern recognition [19]. Examples include using sentiment analysis on social media posts to identify health and mental well-being issues [20] and characterizing mental health problems [21]. In addition, topic modeling has been used to understand public perceptions of the COVID-19 pandemic on Twitter [22] and to uncover health-related topics on social media [23].

Spatial overlay analysis is a group of methodologies used in geographic information systems to simultaneously display multiple layers of spatial information and assess the relationships between different geographic features and attributes [24]. Spatial overlay analysis can be used to examine the relationships between multiple layers of geospatial data to locate spatial points (eg, coordinates) in delimited geographic or jurisdictional areas. Geographic information system methods have been used in health geography and environmental epidemiology to study the geographic incidence or distribution of diseases [25]. When census data or data sets containing contextual variables are unavailable, researchers may have to engage in laborious manual extraction of web-based data, which can be time-consuming and susceptible to inaccuracies [11]. Currently,
there is a gap in the literature regarding methods that enable researchers to automatically convert large volumes of internet text information into meaningful, ready-to-analyze data sets containing contextual data. We propose that by combining techniques used in WeTMS, researchers can efficiently extract, process, and geolocate vast amounts of internet text data to produce structured data sets that encompass variables reflecting the contextual characteristics of specific geographic or jurisdictional areas.

**Objectives**

The aims of this study are 2-fold. First, we outline the implementation of the WeTMS method through a research case, creating data sets with contextual variables on health assets that could improve social connections for older adults across various health jurisdictions in Catalonia, Spain. Second, we analyze these data sets to describe the characteristics and registration trends of these health assets by local stakeholders.

In this tutorial, we first introduce the WeTMS method and describe its application to a research case for compiling data sets of health assets that could enhance social connections among older adults in the health jurisdictions of Catalonia. These assets include activities and resources in the community that can facilitate social interaction, such as social activities, walking groups for retirees, libraries, and senior community centers [26].

Next, we use these new data sets to extract assets with the potential to foster older adults’ social connections and conduct a descriptive analysis to explore their characteristics and asset registration trends across jurisdictions. This analysis demonstrates the potential application of this method in program evaluation. In addition, we discuss the challenges that health and social researchers may face during the WeTMS process and provide resources and programming codes to facilitate its application in other areas of research.

**Methods**

**Context and Data Sources**

In 2015, the government of Catalonia launched the Assets and Health platform as a component of 2 provincial public health programs that aimed to promote intersectoral collaborations among health and nonhealth organizations to tackle complex public health issues, including older adults’ lack of social connections [7]. The Assets and Health platform (created by the Asturias Health Observatory and shared through the Spanish Community Health Alliance) is a search engine and repository where stakeholders from multiple local organizations can register community health assets [27]. Health assets are activities and resources within the community that contribute to maintaining the health and well-being of individuals and groups [28].

Health assets were registered as “activities” (time-bound initiatives and structured interventions, like arts and crafts or supervised walking outings) and “resources” (permanent community fixtures such as associations, parks, and civic centers). Once registered, each health asset is stored on an individual website detailing characteristics, such as its title, description, location, and target population. These individual websites are linked to a search engine, enabling stakeholders to locate assets available in their basic health areas (BHAs), which can be used in collaborative interventions to address public health problems. Each BHA in Catalonia is a local health jurisdiction that functions as an administrative unit within the Catalan healthcare system [29]. In urban settings, BHAs typically cover specific neighborhoods or districts, whereas in rural areas, they may span one or more municipalities, as determined by demographic, epidemiological, and accessibility considerations.

**Overview of the WeTMS Method**

The 8 steps of the proposed method are summarized in Figure 1. The first 3 steps involve identifying and extracting website data through web scraping, and then storing the information in structured data sets to facilitate their analysis. Steps 4-6 describe the application of text processing and mining techniques to analyze the scraped data, identify patterns in the text content, and classify the data into new variables and categories. Steps 7 and 8 elaborate on the use of spatial overlay analysis to locate data within delimited geographic or jurisdictional areas.
Steps 1 to 3: Using Web Scraping for Data Extraction

**Step 1: Defining Variables of Interest**

The target information, including the type of data and the desired outputs for web scraping, was first outlined to avoid extracting irrelevant information. We aimed to generate context-level variables capturing the attributes and registration dates of community health assets to enhance social connections among older adults in Catalonia's BHAs. The data to be extracted from each website detailing a health asset included text elements such as title, description, target population, location, asset registration date, cost, duration, and activity topics. Other data types that could be targeted for extraction include images, links, and metadata, while outputs might be structured data sets such as CSV files, which contain prespecified variables.

**Step 2: Identifying and Exploring the Data Sources**

The second step involves identifying the URLs or web addresses containing the target information and understanding how their content is structured. Identifying URLs can present challenges such as information being dispersed across multiple websites or URLs being hidden or changing [30]. Consequently, sites may be missed because the web scraping “crawler” (the portion of code responsible for finding each URL) requires exact web addresses [30].

In our initial exploration of the Assets and Health search engine, only the last 100 registered health assets were displayed, and those URLs were hidden. A pattern in the URLs for websites describing each health asset, comprising a fixed segment and variable reference number, was identified using Chrome DevTools for network inspection.

The source code of the target websites, usually HTML, was examined to discern their organization and structure. The attributes of HTML elements containing target data (eg, asset titles and descriptions) were identified and used to program the web scraper. Additional information and resources required to implement step 2 are available in Multimedia Appendix 1 [11,31-36].

**Step 3: Extracting and Parsing URL Source Data**

After identifying the relevant URLs and HTML elements, a web scraper comprising a crawler, parser, and data handler was developed using Python 3.10 [37] on the PyCharm 2022.2.2 environment, and the libraries “requests” [38], “beautifulsoup4” [39], and “pandas” [40]. The web scraper also incorporated error-handling mechanisms to manage potential issues, such as connection failures when URLs were nonexistent. The code is explained in detail in Multimedia Appendix 2 and is publicly accessible on GitHub [41].

The web crawler requested 50,000 URLs from websites linked to the Assets and Health platform, comprising 25,000 activities and 25,000 resources, to capture reference numbers from the onset of the program, from July 2015 to December 23, 2022, which was the day when the data were scraped. The program “parser” then analyzed the HTML code of each existing URL and extracted the desired elements, stripping the text, which was automatically stored in 2 CSV data sets for activities and resources. The encoding of the data sets was revised to avoid mismatches between the character set used to represent the text data and that of the scraped text, as this can result in certain characters being displayed as symbols. An initial review of scrapped health assets was conducted to exclude irrelevant observations. We filtered out assets registered outside Catalonia or targeted solely at children and youth before proceeding with the text processing and mining steps.
Steps 4 to 6: Text Processing and Mining to Generate Meaningful Contextual Variables

Step 4: Cleaning and Preprocessing
Cluttered and inconsistent text data obtained from web scraping were preprocessed for analysis [31]. We used RStudio (version 2022.12.0), with the “tm” [42] and “qdap” [43] libraries. The “tm” library provides functions for cleaning, preprocessing, and analyzing text data. The “qdap” library allows text categorization, word frequency calculation, tokenization, and clustering.

The first author manually examined a set of activities and resources from scraped text to assess the quality and structure of the text data. This step was crucial for identifying inconsistencies, such as assigning different age ranges (e.g., 60 and 65 years) to older adults simultaneously.

To preprocess the text data, the columns containing free text, namely titles and descriptions of the health assets, were merged and converted to “corpus” objects—a data structure for text data in R. Columns with text derived from fixed responses were not preprocessed. Irrelevant stop words were then removed, text data were segmented into individual units that could be transformed into numerical variables (tokenization), and words were normalized to their root form (stemming) [44]. The 2 data sets, containing health assets registered as activities and resources, were processed independently. The code with explanations for this step can be accessed through GitHub [45].

Step 5: Defining Dictionaries and Categorizing Text
We used text mining techniques to develop a classification system to filter and categorize health assets for older adults to enhance social connections from all other registered activities and resources on the platform. Text classification is pivotal for the generation of new variables of interest from unstructured data, because it can categorize text into predefined classes or labels.

First, to classify health assets, we predefined new variables and categories created through a deductive approach, based on the literature on social connections [46,47]. We also used inductive processes to create new variables and categories based on patterns identified during the text analysis and discussion among the research team. Table 1 lists the new variables, categories, type of creation process, and literature sources. Detailed definitions of the new variables and categories are provided in Multimedia Appendix 3 [46,48-51].

Second, we created document-term matrixes from the corpus of preprocessed text data containing health asset titles and descriptions. A document-term matrix is a mathematical matrix that describes the frequency of terms in a collection of textual data [31]. The frequency of each word was calculated and sorted based on their frequency values, representing the number of times a term appeared in the title and description of health assets.

Third, over the course of 3 meetings, 2 researchers (PG-H and CM) identified and selected high-frequency words that were repeated 15 times or more in the scraped data, grouped them into topic-specific dictionaries, and refined the list. Eligibility criteria, informed by the definitions of each new variable category, were developed to determine which words to include in each dictionary.

Finally, a classification system was developed using a rule-based classifier. A rule-based classifier categorizes data into predefined classes by applying a set of human-defined rules and conditions based on the features and attributes of the data [52]. We opted for a rule-based system over more complex machine learning classifiers, as this approach is better suited for scenarios with a limited number of specific labels and smaller data sets and ensures efficiency and interpretability without the need for extensive training data [52,53]. Topic-specific dictionaries consist of lists of words related to the definitions of the predefined variable categories as conditions to classify the text data [54]. Finally, an R function was developed to automatically generate a new column for each new variable, search for dictionary words in the scraped data, and assign a new category value if a word was found. The classifier system, including topic-specific dictionaries, is accessible on GitHub [45].
<table>
<thead>
<tr>
<th>New variables</th>
<th>Categories within each variable</th>
<th>Source columns from scraped text data</th>
<th>Creation process and literature sources</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Activities</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity type</td>
<td>Leisure and skill development, physical activity, social facilitation, psychological therapies, awareness campaigns, health and social care, and befriending</td>
<td>Title and description</td>
<td>Deductive [46,48,49]</td>
</tr>
<tr>
<td>Format</td>
<td>Group and individual</td>
<td></td>
<td>Deductive [49]</td>
</tr>
<tr>
<td>Focus</td>
<td>Direct and indirect</td>
<td></td>
<td>Deductive [49]</td>
</tr>
<tr>
<td>Age</td>
<td>Children, youth, adults, older adults, general population, minors unspecified, and adults unspecified</td>
<td>Description, target population, and topics</td>
<td>Deductive [50]</td>
</tr>
<tr>
<td>Gender</td>
<td>Women, men, nonbinary, and any</td>
<td>Description, target population, and topics</td>
<td>Deductive [51]</td>
</tr>
<tr>
<td>Vulnerable populations</td>
<td>Migrants, caregivers, substance use, physical diseases, risk social exclusion, mental diseases, and all(^a)</td>
<td>Title, description, target population, and activity topics</td>
<td>Inductive</td>
</tr>
<tr>
<td><strong>Resources</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource type</td>
<td>Municipal natural and green space, health institution, social welfare institution, education institution, patient advocacy group, charitable and voluntary organization, faith-based organization, parent school associations, public library, civic center, sports institution, leisure and cultural association, neighborhood association, and cultural institution</td>
<td>Title and description</td>
<td>Inductive</td>
</tr>
<tr>
<td>Focus</td>
<td>Direct and indirect</td>
<td></td>
<td>Deductive [49]</td>
</tr>
<tr>
<td>Age</td>
<td>Children, youth, adults, older adults, general population, minors unspecified, and adults unspecified</td>
<td>Title, description, and topics</td>
<td>Deductive [50]</td>
</tr>
<tr>
<td>Gender</td>
<td>Women, men, nonbinary, and any</td>
<td>Description and topics</td>
<td>Deductive [51]</td>
</tr>
<tr>
<td>Vulnerable populations</td>
<td>Migrants, caregivers, substance use, physical diseases, risk social exclusion, mental diseases, and all(^a)</td>
<td>Title, description, and topics</td>
<td>Inductive</td>
</tr>
</tbody>
</table>

\(^a\)“Substance use,” “physical diseases,” “risk social exclusion,” “mental diseases,” and “all” are simplified terms for target populations experiencing substance use, physical diseases, mental diseases, those at risk of social exclusion, and the general population.

**Step 6: Manual Verification**

The categories assigned to the new variables for each health asset were reviewed for inconsistencies by 2 researchers with expertise in the topic (PG-H and Angeli Chaculiaz). Manual verification refers to a one-by-one examination of the classified data by human reviewers to ensure the accuracy of the new variables created [55]. Verification involved an independent review of 200 health assets by 2 researchers to assess new variable categories based on eligibility criteria. Discrepancies were resolved through web meetings with manual reclassification if necessary. The data sets were then divided into groups of 500 health assets for independent review. Agreement rates between the verified and automatically generated variables were computed using the Excel software.

**Step 7: Identifying the Layers of Spatial Data**

We used a spatial overlay analysis to generate a new variable that identified the BHA in which each health asset was located. The analysis was conducted in RStudio (version 2022.12.0) because of its many packages specifically designed for spatial overlay analysis, such as “sp,” [56] “sf,” [57] “rgdal,” [58] “rgeos,” [59] and “ggplot2” [60]. In this step, 2 spatial layers were identified. The first layer consisted of polygonal data depicting 374 BHAs in Catalonia. The data were obtained from the open database of the General Directorate of Health Planning and Research in Catalonia. Polygonal data can represent geographic or jurisdictional regions by defining their boundaries [61]. The second layer comprises a vector of geographic point data for each health asset. Point data consisted of longitude and latitude coordinates obtained from the addresses scraped for each activity and resource using Excel add-on GeoCode, a map tool that uses Google services to automatically retrieve longitudes and latitudes from addresses.

**Step 8: Matching Coordinate Reference System and Finding Intersection Points**

Step 8 involves transforming the spatial data layers into a common coordinate reference system (CRS) and identifying the intersecting points. The CRS of a spatial object determines its location on the Earth’s surface. Thus, analyzing 2 or more spatial layers with different CRS can produce misleading outcomes [61]. To identify areas of overlap between health asset coordinates and BHAs, the following steps were taken: (1) coordinates were transformed to a simple feature object format, (2) simple feature objects were converted into single points using the “st_point” function, and (3) both spatial data layers were converted to a common CRS.
Finally, a spatial overlay analysis was performed using the “st_intersects” function to determine the BHA polygons with which each health asset point data intersected. The function was applied in a loop to each row of the activity and resource data sets. The resulting outputs are stored in new columns named “Code_BHA” and “Name_BHA.” The code, along with explanations for steps 7 and 8, is available in GitHub [45].

Data Set Filtering and Descriptive Analysis

The new data sets were filtered using the new variables and categories to select health assets with the potential to foster social connections among older adults from all scrapped assets. Eligible health assets registered as activities and resources were included if (1) the target population included older adults, (2) the format was either group activities or individual activities fostering social connections (eg, befriending), and (3) they were located in Catalonia.

A descriptive analysis was conducted in RStudio (version 2022.12.0), to understand the characteristics and asset registration trends of stakeholders across BHAs. Frequencies and proportions were calculated for each category of the new variables (activity type, format, focus, age, sex, and vulnerable populations). Temporal registration trends of activities and resources were analyzed using time-series graphs with local polynomial regression fitting lines, a nonparametric method used to describe the deterministic variation in data [62]. We also computed the average weekly registration of activities and resources in each BHA, assuming a Poisson distribution, where \( \lambda \) represents the weekly health asset registrations per area. Finally, visualization techniques were used to analyze the temporal evolution of the registration of activities and resources on the Assets and Health websites across BHAs, as well as their geographic distribution.

Ethical Considerations

The data collected in this study were publicly accessible and did not contain any personal or sensitive information. Thus, ethical approval and participant consent were not required for this study. In addition, before data collection, we verified that the websites of interest did not have any explicit prohibitions against automatic web scraping, such as a “robots.txt” file or similar declarations.

Results

Results From WeTMS

Web Scraping

Of the 50,000 URLs inspected, 17,305 contained websites describing health assets (9558 activities and 7747 resources) registered with local stakeholders from July 2015 to December 2022. The number of observations obtained through web scraping matched the total number of assets reported on the Assets and Health platform, thus demonstrating the efficacy of the web scraper. No missing values were detected for the main variables (eg, title, description, location, and date of asset registration). In the activity data set, 9.56% (480/5022) of observations did not disclose the activity cost, and 49.04% (2463/5022) did not report the activity duration. An example of an activity and resource, as they appear in the scraped data sets, is provided in Textbox 1.

Textbox 1. Example of a health asset registered as activity and resource extracted from the scraped text (English translation).

<table>
<thead>
<tr>
<th>Activity row #860</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title: School for Adults</td>
</tr>
<tr>
<td>Description: Reading and writing classes</td>
</tr>
<tr>
<td>Population: Over 65 years old—anyone (district neighbors over 65 years old)</td>
</tr>
<tr>
<td>Location: Campoamor Street 92, 08204, Civic Center Rogelio Soto, Sabadell, Barcelona, Catalonia, Spain</td>
</tr>
<tr>
<td>Organizations: Civic Center Rogelio Soto, Campoamor Neighborhood Association</td>
</tr>
<tr>
<td>Registration date: February 6, 2020</td>
</tr>
<tr>
<td>Is free: Yes</td>
</tr>
<tr>
<td>Categories: Women, older adults, people at risk of exclusion, school of health, mental health, or emotional well-being</td>
</tr>
<tr>
<td>Time activity: From September 13, 2019, to June 30, 2020</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resource row #1961</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title: Association of Retirees and Pensioners, La Pineda</td>
</tr>
<tr>
<td>Description: Association that aims to promote cultural training and sports activities for older adults, as well as avoiding loneliness and social isolation, fostering relationships between them</td>
</tr>
<tr>
<td>Registration date: May 15, 2017</td>
</tr>
<tr>
<td>Location: Alfredo Kraus Street 20, 43481, La Pineda Vila-seca, Tarragona, Catalonia, Spain</td>
</tr>
<tr>
<td>Categories: Older adults, mental health or emotional well-being, physical activity, community health</td>
</tr>
</tbody>
</table>
Text Mining

From the text processing of the corpus of titles and descriptions, a total of 12,560 tokens (or raw words) were identified for activities and 7301 for resources, of which 996 (7.9%) and 594 (8.1%) words had a frequency >15. Using words with a frequency of >15, we constructed 73 topic-specific dictionaries corresponding to each category of the new variables. For instance, for the physical activity category under the activity type variable, the topic-specific dictionary included words such as “physical,” “exercise,” “gym,” “yoga,” and “sport.” Figure 2 presents popular dictionary words for each category within the activity type variable.

Figure 2. Categories within the “activity type” variable showcasing popular words derived from topic-specific dictionaries (English translation).

After applying the rule-based classifier using topic-specific dictionaries, manual verification of the output yielded variable levels of agreement ranging from 62.02% (3417/5509) to 99.47% (4886/4912) across variables. For instance, variables with lower classification accuracy had a larger number of possible categories, a more evenly distributed number of observations across categories, or words repeated fewer than 15 times within the title and description corpus. The agreement rates between the verified and automatically generated databases are presented in Table 2.
Table 2. Agreement rate between manually verified and automatically classified data sets.

<table>
<thead>
<tr>
<th>New variables generated</th>
<th>Correctly assigned categories, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Activities data set (n=6260)</strong></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>4855 (77.55)</td>
</tr>
<tr>
<td>Gender</td>
<td>6215 (99.28)</td>
</tr>
<tr>
<td>Vulnerable populations</td>
<td>5525 (88.26)</td>
</tr>
<tr>
<td>Activity type&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3417 (62.02)</td>
</tr>
<tr>
<td>Format&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5326 (96.67)</td>
</tr>
<tr>
<td>Focus&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5342 (96.97)</td>
</tr>
<tr>
<td><strong>Resources data set (n=4912)</strong></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>4029 (82.02)</td>
</tr>
<tr>
<td>Gender</td>
<td>4843 (98.59)</td>
</tr>
<tr>
<td>Vulnerable populations</td>
<td>3883 (79.05)</td>
</tr>
<tr>
<td>Resource type</td>
<td>3845 (78.28)</td>
</tr>
<tr>
<td>Focus</td>
<td>4886 (99.47)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Automatic classification of the categories for variables activity type, format, and focus was performed only for activities targeting older adults (n=5509).

Spatial Overlay Analysis

Coordinates for the locations of 0.36% (18/5055) of activities and 2.41% (109/4530) of resources were not identified. Manual searches on Google Maps using addresses allowed us to locate the coordinates for all but 7 activities and 2 resources. Through spatial overlay analysis, intersections between spatial points and BHAs were not identified for 26 activities or 4 resources. The newly generated columns for the variables Code_BHA and Name_BHA encompassed values for 318 distinct BHAs, representing 85% of the 374 BHAs.

Results From Data Set Filtering and Descriptive Analysis

Filtering Health Assets With Potential to Enhance Older Adults’ Social Connections

Using the newly generated contextual variables, we filtered the data sets of activities and resources to identify those with the potential to foster social connections among older adults. From the initial 17,305 health assets identified, we obtained 9546 eligible health assets, comprising 5022 activities and 4524 resources. The reasons for exclusion and the stages in which health assets were discarded are shown in Figure 3.
Figure 3. Flowchart for the filtering of health asset data sets by contextual variables related to social connection constructs generated using the web scraping, text mining, and spatial overlay (WeTMS) method. BHA: basic health area.

Characteristics of Eligible Health Assets

Of the health assets registered as activities, 24.59% (1235/5022) specifically targeted older adults, whereas 75.41% (3787/5022) targeted broader age ranges, including the older population. Most resources targeted the general population and included older adults, with only 4.12% (207/5022) being exclusively for older adults, such as civic centers for retirees. Only 2.49% (238/9546) of the health assets had a sex-specific target; these were predominantly women (n=212). Among all health assets, 13.5% (678/5022) of activities and 7.98% (361/4524) of resources were tailored for specific vulnerable groups, with physical or mental illness being the primary focus.

Group-oriented activities promoting social interactions accounted for 99.56% (5000/5022) of the eligible activities. However, only 4.36% (219/5022) explicitly used concepts related to social connections (eg, loneliness and social isolation) in titles and descriptions. Over 57% (2862/5022) of the activities were cost-free, and the most common activity duration was 1 to 3 months (975/5022, 19.41%). Data on format, duration, and cost are not available for resources.

The analysis of the new variable *activity type* showed that leisure and skill development activities were most common (1844/5022, 36.72%). This included group handcrafts, dance, painting, theater, cooking, choir courses, and conversation groups that focused on shared-interest topics. Group exercise activities (eg, walking groups) accounted for 31.08% (1561/5022) of the activities. Over 22% (1103/5022) of the activities involved group activities with health and social professionals outside the health care center, including psychological therapies and health and social care. Finally, 8.46% (425/5022) were social facilitation activities such as group meetings to share common interests (eg, film forums). Overall, more than half of these activities were registered between 2021 and 2022 (Figure 4).
Almost 61.49% (2782/4524) of the registered resources facilitated exchange of knowledge and interests among older adults. These resources included leisure and cultural associations; public libraries; civic centers; and cultural, sports, and educational institutions. Municipal natural and green spaces where adults can gather accounted for 17.28% (782/4524) of the resources. A total of 595 (13.1%) health institutions and 140 (3.1%) social welfare institutions were found, including primary care centers, health and social foundations, and advocacy institutions promoting social inclusion. Other resources linked to health and social welfare include patient advocacy groups, faith-based organizations, and charitable and voluntary organizations. In contrast, most resources were registered before 2019 (Figure 5). A detailed descriptive analysis of the type, target population, focus, cost, format, and duration of health assets is included in Multimedia Appendix 4.

Figure 4. Number of activities with the potential to enhance older adults’ social connections by type and year of registration (2015-2022).

Figure 5. Number of resources with potential to enhance older adults’ social connections, by type and year of registration (2015-2022).
Overview of Registration Trends of Eligible Health Assets Across BHAs

The first registry of a health asset on the Assets and Health websites occurred on July 23, 2015, and the last on December 23, 2022, the day when web scraping was conducted. Total registration of activities remained consistently low from the start of the program until early 2018, whereas for resources, a registration peak was observed in late 2016. Activity and resource registrations have increased from 2018 to mid-2020. A decline in registration was observed from early 2020 to mid-2021, coinciding with the outbreak of the COVID-19 pandemic. Local polynomial regression fitting lines showed a growing pattern in the registration of activities from 2021 onwards, whereas resource registration remained low (Figure 6).

On the basis of the observed trends, 4 implementation periods were defined to better understand registration trends: period 1, from July 2015 to January 2018; period 2, from February 2018 to February 2020; period 3, from March 2020 to May 2021; and period 4, from the end of June 2021 to December 2022. During the first two and a half years of the program, the average number of activities and resources registered per week across all BHAs was 0.37 and 3.47, respectively, increasing to 11.83 and 25.27 in period 2. During the COVID-19 pandemic in period 3, these figures decreased to an average of 3.19 activities and 20.64 resources per week. Period 4 had the highest registration rate for activities (38.52/wk).

To calculate the registration trends in individual BHAs, we divided the number of BHAs with one or more activities registered by the total number of BHAs (n=374) for each period. We did not consider the resource data set because of the observed patterns suggesting centralized registration, rather than local registration. For instance, in late 2016, resources in 237 BHAs were registered in a single day. At the end of period 1, 8% (30/374) of the BHAs had one or more registered activities, which increased to 85% (318/374) by the end of period 4 (Figure 7; Table 3). The number of health assets registered per BHA varied significantly, ranging from 0 to 263 activities and 0 to 265 resources. The median number of activities registered per BHA from 2015 to 2022 was 5 (IQR 13.75) and 9 (IQR 10) for resources. Figure 8 illustrates the geographic distribution of the activities and resources registered in each period.

Figure 6. Weekly registration trends of health assets from July 2015 to December 2022. The y-axis represents the number of registered activities and resources per week. The x-axis represents time, labeled by years for clarity.
Figure 7. Cumulative frequency of basic health areas (BHAs) with registered health assets aimed at enhancing older adults’ social connections from July 2015 to December 2022. Each point represents a BHA at the time of its first asset registration on the Assets and Health platform, cumulative.

Table 3. Average number of health assets registered per week and proportion of basic health areas (BHAs) with one or more activities registered per period.

<table>
<thead>
<tr>
<th>Time periods</th>
<th>Number of health assets</th>
<th>λ (95% CI)a</th>
<th>Resources</th>
<th>BHAs with registered activities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Activities</td>
<td>Resources</td>
<td>Activities</td>
<td>Resources</td>
</tr>
<tr>
<td>Period 1: July 2015 to January 2018</td>
<td>50</td>
<td>462</td>
<td>0.37 (0.27-0.48)</td>
<td>3.47 (3.16-3.79)</td>
</tr>
<tr>
<td>Period 3: March 2020 to May 2021</td>
<td>562</td>
<td>301</td>
<td>8.64 (7.93-9.36)</td>
<td>4.63 (4.11-5.15)</td>
</tr>
<tr>
<td>Period 4: June 2021 to December 2022</td>
<td>3120</td>
<td>941</td>
<td>38.52 (37.16-39.87)</td>
<td>11.62 (10.87-12.36)</td>
</tr>
</tbody>
</table>

aλ denotes the average number of health assets registered per week in each period.
bUnique BHAs that registered activities targeting social connections in older adults for the first time in the specified period.
cBHAs that registered such activities up to and including each period with earlier registrations.
Discussion

Innovation: Generating Area-Specific Contextual Variables From Unstructured Web-Based Data

We introduce a novel approach for generating area-specific contextual variables from unstructured website data using WeTMS. By combining the methods commonly used in computer and data science, we were able to efficiently gather and transform large amounts of website data into comprehensive data sets of theoretically informed variables. The resulting data sets enabled us to identify and characterize health assets with the potential to enhance social connections among older adults registered within health jurisdictional areas from 2015 to 2022. In addition, this approach allowed us to examine area-specific registration trends for health assets, showing the use of the Assets and Health platform developed as part of a public health strategy in Catalonia. We provided detailed explanations of concepts, steps, and the code used, and included supplementary information to facilitate the replication of the steps, attempting to familiarize novice readers with these techniques.

Applications of the WeTMS Method

Our method provides a tool for researchers interested in developing new contextual variables when data are scarce or difficult to obtain using traditional means. Researchers in fields such as public health, nursing, and social epidemiology who study the impact of emerging health and social phenomena on health outcomes and determinants of health can benefit from this method. A practical example of an emerging social determinant of health, such as precarious employment [63], can consist of applying the web-scraping steps to obtain website data from employment portals, text mining to analyze posts, identifying precarious job offers, and spatial overlay analysis to locate them into geographic areas and study the effect on population outcomes using multilevel modeling. Researchers and program evaluators in health services research and implementation science can use this method to obtain data to conduct descriptive analyses explaining policy adoption within jurisdictional or geographic areas, following the research case outlined in this study.

A key feature of this method is that its steps can be implemented in sequence or independently, depending on the research goals. For example, researchers interested in generating new variables from text data without locating them in specific geographic areas can follow steps 1 to 6, which involve web scraping, text processing, and mining. If a data set is already available and researchers want to group the data by geographic settings, they can follow steps 7 and 8, which involve overlay spatial analysis.

Challenges and Limitations

There are challenges with this method that can limit its feasibility and application. In our example, extracting comparable data from multiple URLs was feasible because the websites associated with the Assets and Health platforms had similar HTML structures. The consistent placement of targeted information across websites simplifies the complexity of the web-scraping program. Thus, the attached web scraper code is only suitable for single or multiple websites with a limited number of distinct HTML structures (eg, forums, social media, and employment portals). Studies in which the target data are spread over different websites with varied designs require more advanced programming [11].

A key step in the process—the creation of topic-specific dictionaries for the classification of observations—necessitates a deep understanding of the field and the terminology used in the data. Overall, the rule-based classifier demonstrated high
accuracy. However, some variables, such as activity type, showed a higher rate of errors, in part because the dictionaries used to classify them contained only high-frequency words found in the titles and descriptions. Thus, our experience suggests that manual verification of new variables and categories by researchers with a comprehensive understanding of the data and subject matter is essential to ensure data validity before statistical analysis. However, this can be unfeasible for large data sets. In such scenarios, the impracticality of manual verification may necessitate the use of complex machine learning classifiers, presenting a trade-off in the confidence of the data that potentially compromises the robustness of the resulting variables [52].

Spatial overlay analysis effectively localizes health assets to their respective health jurisdictions, facilitated by the acquisition of complete addresses during the web scraping phase and the availability of a high-quality polygon map for analysis. Geospatial maps can be obtained from government agencies, nonprofit organizations, and commercial providers. If maps are unavailable, they can be created using accessible satellite imagery [64]. However, the necessity for location-specific data (eg, addresses, postal codes, and cities) for each observation to generate contextual-level variables limits the range of suitable data sources available to researchers.

Ethical and data protection considerations are important. Web scraping is typically permitted when data are publicly accessible and not subject to international legislation concerning personal data, trademarks, copyrights, or private information [30]. Automatic extraction of internet data might be unfeasible if the data are not publicly available or if a website’s terms of service restrict automated collection and analysis [65]. Researchers may consult ethics bodies to ensure that the methodology adheres to ethical standards when dealing with sensitive topics and personal information, even when relying on publicly available sources.

In addition to these challenges, the method and data sets that it produces have limitations. The complexity of these steps requires introductory technical knowledge. Thus, we have provided detailed explanations and supplementary information that can support researchers, as they familiarize themselves with the steps. We anticipate that the compendium of concepts, code, software packages, and references gathered from trustworthy sources will serve as a resource for those interested in these techniques.

Another limitation is the bias associated with the classifier system. The development of classification systems inherently relies on the subjective judgment of researchers. This can result in misclassifications, particularly those related to assumptions about race, gender, or social exclusion factors, especially within machine learning classifiers [66]. It is advisable for researchers to engage in a reflexive process, carefully considering their assumptions in the definition and selection of dictionary words and to critically evaluate how these decisions may influence the investigation [67].

Finally, although the data sets generated are robust for descriptive analysis, researchers should proceed with clearly defined assumptions when using new context-level variables in statistical analyses, particularly in multilevel modeling or ecologic studies that aim to draw inferences. Although we successfully compiled 2 data sets of health assets from targeted websites, their comprehensiveness and accuracy in reflecting all identified assets across BHAs remain unknown. It is possible that some local organizations in BHAs were more or less likely to register assets on Assets and Health websites, influenced by context-specific factors such as management support and training on the platform [68]. If feasible and ethical, it would be advisable for researchers to triangulate data to validate the data sets, thus verifying web-scraped data with secondary sources or direct inputs from stakeholders [69].

Conclusions

The sequential use of WeTMS enabled the efficient creation of data sets of health assets registered with the Assets and Health websites in Catalonia, Spain, which aimed to enhance the social connections of older adults in local health jurisdictions. Our descriptive analysis demonstrated the usefulness of the data sets in exploring the characteristics of contextual variables, as well as in understanding temporal patterns and spatial distributions. Contextual-level variables generated via WeTMS may also be used in hierarchical analyses to evaluate the impact of contextual factors on health outcomes when more robust sources, such as census data, are not available. Adherence to data protection standards and ethical considerations should also guide this process. Although WeTMS has potential value for multiple research disciplines, it presents challenges and limitations, including the need for internet data sources to have comparable structures, a dependence on location data, the potential lack of representativeness in website content, the requirement for technical expertise, and a significant time investment for manual verification.

Acknowledgments

The authors thank Angeli Chacaliaza for his valuable contribution as a second reviewer in the manual verification stage and Ann Tourangeau, Martine Puts, and Tina Behrouzi for their comments and suggestions on the first draft. The authors also acknowledge Charles Anderson for his technical assistance. This study was not supported by any funding. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content in this manuscript.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.
Authors' Contributions
PG-H and CM conceptualized the study; PG-H wrote the study protocol, developed the method, and drafted the manuscript; and AG-V, LG-P, and CM provided expert input and conducted manuscript review and editing. All the authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Additional details in step 2: identifying hidden URLs, finding HTML elements, and further references.

Multimedia Appendix 2
Detailed explanation of Python libraries and web scraper code.

Multimedia Appendix 3
Definitions of new variables and categories for text classification.

Multimedia Appendix 4
Extended descriptive analysis of the type, target population, focus, cost, format, and duration of health assets.

References
33. HTML Tags. W3 Schools. 2023. URL: https://www.w3schools.com/tags/ [accessed 2023-12-25]
37. Python homepage. Python. URL: https://www.python.org/ [accessed 2023-12-18]
41. Targeted web scraping implementation evaluation. GitHub. URL: https://github.com/paugalvez/Targeted_webscraping_implementation_evaluation.git [accessed 2023-12-18]

42. Feinerer I. Introduction to the tm Package Text Mining in R. The Comprehensive R Archive Network. 2023 Feb 5. URL: https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf [accessed 2023-12-18]


47. National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Health and Medicine Division, Board on Behavioral, Cognitive, and Sensory Sciences, Board on Health Sciences Policy, Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults. Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System. Washington, DC: National Academies Press; 2020.


55. Galvez-Hernandez et alJMIR PUBLIC HEALTH AND SURVEILLANCE


Abbreviations
- **BHA**: basic health area
- **CRS**: coordinate reference system
- **SDOH**: social determinants of health
- **WeTMS**: web scraping, text mining, and spatial overlay analysis

©Pablo Galvez-Hernandez, Angelina Gonzalez-Viana, Luis Gonzalez-de Paz, Ketan Shankardass, Carles Muntaner. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 08.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Tutorial

Prospective Spatiotemporal Cluster Detection Using SaTScan: Tutorial for Designing and Fine-Tuning a System to Detect Reportable Communicable Disease Outbreaks

Alison Levin-Rector¹, MPH; Martin Kulldorff, PhD; Eric R Peterson¹, MPH; Scott Hostovich², BS; Sharon K Greene¹, PhD, MPH

¹Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene, Long Island City, NY, United States
²Information Management Services, Inc, Calverton, MD, United States

Corresponding Author:
Alison Levin-Rector, MPH
Bureau of Communicable Disease
New York City Department of Health and Mental Hygiene
42-09 28th St
Long Island City, NY, 11101
United States
Phone: 1 347 396 2600
Email: alevinrector@health.nyc.gov

Abstract

Staff at public health departments have few training materials to learn how to design and fine-tune systems to quickly detect acute, localized, community-acquired outbreaks of infectious diseases. Since 2014, the Bureau of Communicable Disease at the New York City Department of Health and Mental Hygiene has analyzed reportable communicable diseases daily using SaTScan. SaTScan is a free software that analyzes data using scan statistics, which can detect increasing disease activity without a priori specification of temporal period, geographic location, or size. The Bureau of Communicable Disease’s systems have quickly detected outbreaks of salmonellosis, legionellosis, shigellosis, and COVID-19. This tutorial details system design considerations, including geographic and temporal data aggregation, study period length, inclusion criteria, whether to account for population size, network location file setup to account for natural boundaries, probability model (eg, space-time permutation), day-of-week effects, minimum and maximum spatial and temporal cluster sizes, secondary cluster reporting criteria, signaling criteria, and distinguishing new clusters versus ongoing clusters with additional events. We illustrate how to support health equity by minimizing analytic exclusions of patients with reportable diseases (eg, persons experiencing homelessness who are unsheltered) and accounting for purely spatial patterns, such as adjusting nonparametrically for areas with lower access to care and testing for reportable diseases. We describe how to fine-tune the system when the detected clusters are too large to be of interest or when signals of clusters are delayed, missed, too numerous, or false. We demonstrate low-code techniques for automating analyses and interpreting results through built-in features on the user interface (eg, patient line lists, temporal graphs, and dynamic maps), which became newly available with the July 2022 release of SaTScan version 10.1. This tutorial is the first comprehensive resource for health department staff to design and maintain a reportable communicable disease outbreak detection system using SaTScan to catalyze field investigations as well as develop intuition for interpreting results and fine-tuning the system. While our practical experience is limited to monitoring certain reportable diseases in a dense, urban area, we believe that most recommendations are generalizable to other jurisdictions in the United States and internationally. Additional analytic technical support for detecting outbreaks would benefit state, tribal, local, and territorial public health departments and the populations they serve.

(JMIR Public Health Surveill 2024;10:e50653) doi:10.2196/50653

KEYWORDS
communicable diseases; disease outbreaks; disease surveillance; epidemiology; infectious disease; outbreak detection; public health practice; SaTScan; spatiotemporal; urban health
Introduction

Background
The Bureau of Communicable Disease (BCD) at the New York City (NYC) Department of Health and Mental Hygiene monitors electronic reports daily of >70 reportable infectious diseases for an urban population of approximately 8.5 million residents [1]. Since 2014, to help prioritize resources for case and outbreak investigations and response activities, the BCD has automated daily analyses to prospectively detect and monitor spatiotemporal clusters of reportable communicable diseases using SaTScan [2,3]. SaTScan (an abbreviation of Space and Time Scan Statistics) is a free software that analyzes data using scan statistics [4], which can detect increased disease activity without a priori specification of temporal period, geographic location, or size. SaTScan has effectively detected clusters of enteric and respiratory diseases and of antimicrobial-resistant infections using varied data sources from settings worldwide [5-11]. The BCD has applied several prospective SaTScan analyses to address varied surveillance needs in NYC, with early detection and near–real-time monitoring of salmonellosis, legionellosis, shigellosis, and COVID-19 outbreaks [2,12-14].

For example, in 2015, SaTScan analyses provided the first signal for the second largest outbreak of community-acquired legionellosis in the United States [13]. Detecting serious outbreaks even a day or 2 earlier can potentially save lives.

In the absence of centralized training and technical support for designing and fine-tuning such systems, we have provided informal consultations for other health departments. While there are scientific papers describing the statistical methods [4,15-19] and the SaTScan user guide [20] explains input data requirements and various software features, this is the first such practical resource for health department staff.

Objectives
We aimed to share our recommendations for health departments to design and fine-tune a system to detect community-acquired reportable communicable disease outbreaks using SaTScan (Tables 1-3). This guide describes requirements and parameter settings for a variety of analytic aims and how to use built-in features to interpret results. Methods for detecting building-level outbreaks, hospital-associated outbreaks, purely temporal clusters such as seasonal increases, or early warning signs of diseases with pandemic potential are described elsewhere [21-24].
Table 1. Summary of input file and parameter setting recommendations for designing a system to detect reportable communicable disease outbreaks using SaTScan.

<table>
<thead>
<tr>
<th>Section and subtopic</th>
<th>Recommendations</th>
</tr>
</thead>
</table>
| **Input data**       | 1. The case file for each reportable disease should contain the census tract of the patient home address and the best approximation of illness onset date and can contain additional patient information such as gender and age.  
2. Include all reported events in the case file, whether subsequently confirmed or not.  
3. The coordinates file should include 1 row per census tract with its latitude and longitude.  
4. A network file can be used to avoid clusters that span hard-to-cross boundaries, such as lakes, rivers, and mountain ranges. |
| **Parameter settings** | 5. Use prospective space-time analysis.  
6. Use the space-time permutation probability model.  
7. If case ascertainment over both space and time is highly affected by testing variability, the Bernoulli or Poisson model may be used instead. This additionally requires a control file of negative tests or a population file of all tests.  
8. The space-time permutation model automatically adjusts for purely spatial variations and for purely temporal patterns.  
9. Adjust for day-of-week by space interaction.  
10. To quickly detect outbreaks before many events accrue, use the default minimum number of events of 2.  
11. To support resource allocation to areas with high and increasing disease transmission rather than prioritize case investigations, the minimum number of events can be increased.  
12. For most jurisdictions, allow the cluster to expand in size to include up to 50% of all events during the study period.  
13. For geographically large jurisdictions, either use 10% as the maximum cluster size or add another maximum cluster size (eg, 200 km).  
14. For most communicable diseases, scan for clusters with a minimum of 2 days and a maximum of 30 days.  
15. The study period should preferably be at least 3 times as long as the maximum temporal cluster size. A study period of 1 year is usually reasonable.  
16. The end of the study period should be set to previous day or the most recent day with sufficiently complete data.  
17. Aggregate to daily resolution.  
18. Run analyses daily.  
19. Use the default P value method with the maximum number of Monte Carlo replications set to 999.  
20. Use a signaling threshold of RI ≥ 100 days.  
21. Consider an RI of 100 to <365 days as a weak cluster, an RI of 365 days to <5 years as a moderate cluster, an RI of 5 to <100 years as a strong cluster, and an RI of ≥100 years as a very strong cluster.  
22. Enable the Most Likely Clusters, Hierarchically option and choose No Cluster Centers in Other Clusters. |

aRI: recurrence interval.
Table 2. Summary of output files and built-in feature recommendations for cluster interpretation when using SaTScan to detect reportable communicable disease outbreaks.

<table>
<thead>
<tr>
<th>Section and subtopic</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output files and built-in features</strong></td>
<td></td>
</tr>
<tr>
<td>Text-based results file</td>
<td>1. Review information about the detected clusters and the analysis performed.</td>
</tr>
<tr>
<td>Line list of cluster events</td>
<td>2. Review line list of events in clusters exceeding the RI(^a) signaling threshold.</td>
</tr>
<tr>
<td></td>
<td>3. Distinguish events that are newly added to ongoing clusters by using an events cache input file.</td>
</tr>
<tr>
<td></td>
<td>4. Inspect the line list for data quality problems that may have been missed by other routine checks, such as duplicated events in the case file, misreported patient addresses, or addresses unrelated to possible exposure sites.</td>
</tr>
<tr>
<td></td>
<td>5. Inspect the line list to determine whether the cluster affects multiple households and suggests community transmission vs primarily affects persons in the same household or building.</td>
</tr>
<tr>
<td>Map</td>
<td>6. Visualize the spatial extent of clusters.</td>
</tr>
<tr>
<td>Temporal graphs</td>
<td>7. Observe patterns of event locations and group by any chosen variable included in the case file.</td>
</tr>
<tr>
<td>Drilldown analysis</td>
<td>8. Visualize observed and expected event counts inside and outside the geographical cluster area, marking the period before and during the cluster.</td>
</tr>
<tr>
<td></td>
<td>9. Conduct a drilldown analysis if you wish to determine whether events in a cluster are randomly distributed within that cluster.</td>
</tr>
<tr>
<td></td>
<td>10. In parallel, consider rerunning the analysis applying a maximum reported spatial cluster size as both approaches may detect significant clusters within larger clusters.</td>
</tr>
</tbody>
</table>

\(^a\)RI: recurrence interval.
### Table 3. Summary of recommendations for assessing and fine-tuning a system to detect reportable communicable disease outbreaks using SaTScan.

<table>
<thead>
<tr>
<th>Section and subtopic</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>System assessment</strong></td>
<td></td>
</tr>
<tr>
<td>Proof of concept</td>
<td>1. Identify at least 1 known historical acute, localized, community-acquired outbreak and mimic prospective surveillance around that period to assess the speed and precision with which the outbreak could have been detected.</td>
</tr>
<tr>
<td>System performance</td>
<td>2. If the system consistently produces clusters that are too large or too delayed to be useful, misses signals, or produces too many false or weak signals, first examine data quality by inspecting input files and line list output files for data errors before fine-tuning the system.</td>
</tr>
<tr>
<td><strong>System fine-tuning when outbreak detection is delayed or missed</strong></td>
<td></td>
</tr>
<tr>
<td>Difficulty detecting geographically small outbreaks or outbreaks at the boundaries of geographic units</td>
<td>3. Disaggregate data into smaller geographic units.</td>
</tr>
<tr>
<td>Difficulty detecting outbreaks affecting a particular age group</td>
<td>4. Conduct additional age-restricted subgroup analyses, such as for children aged &lt;5 years.</td>
</tr>
<tr>
<td>Difficulty detecting outbreaks in which people are exposed far from their home</td>
<td>5. Collect and analyze multiple addresses per patient, such as both home and work.</td>
</tr>
<tr>
<td>Difficulty detecting outbreaks because of missing data</td>
<td>6. Conduct quality assurance to detect and resolve drop-offs in laboratory reporting by laboratory and disease.</td>
</tr>
<tr>
<td>Difficulty detecting new outbreaks in areas with previous outbreaks</td>
<td>7. To avoid missing outbreaks affecting unsheltered patients experiencing homelessness and without a geocodable address, assign them to an artificial census tract.</td>
</tr>
<tr>
<td>Difficulty detecting outbreaks affecting people who reside in areas not connected in the network file</td>
<td>8. From the baseline period in the case file, exclude days during major previous outbreaks.</td>
</tr>
<tr>
<td>Difficulty detecting weak clusters that are potentially actionable outbreaks</td>
<td>9. For previous outbreaks limited to a specific building or institution, exclude those events from the baseline period in the case file.</td>
</tr>
<tr>
<td><strong>System fine-tuning when consistently finding large, uninteresting clusters</strong></td>
<td></td>
</tr>
<tr>
<td>Clusters span hard-to-cross boundaries, such as lakes, rivers, and mountain ranges</td>
<td>10. Add more connections to the network file.</td>
</tr>
<tr>
<td>Clusters with a relative risk near 1 are of limited public health interest</td>
<td>11. If using an RI threshold for signaling of ≥1 year, decrease the threshold.</td>
</tr>
<tr>
<td><strong>System fine-tuning when there are too many signals</strong></td>
<td></td>
</tr>
<tr>
<td>Clusters are driven by duplicate events or incorrect patient addresses</td>
<td>12. Use the network input file.</td>
</tr>
<tr>
<td>Clusters are driven by changing nature of input data, such as recent adoption of culture-independent diagnostic testing</td>
<td>13. Impose a minimum relative risk restriction.</td>
</tr>
<tr>
<td>Clusters are driven by reports that do not represent true illnesses, such as reports of patients with only negative laboratory test results</td>
<td>14. Proactively check for and correct data errors.</td>
</tr>
<tr>
<td>Clusters are driven by within-household transmission, so they do not represent community transmission</td>
<td>15. Shorten the baseline study period.</td>
</tr>
<tr>
<td>Clusters represent weak outbreaks that are not actionable given available resources</td>
<td>16. Use stricter disease definition inclusion criteria and impose a time lag to allow for investigating and ruling out cases.</td>
</tr>
<tr>
<td>Clusters represent weak outbreaks that are not actionable given available resources</td>
<td>17. Retain only 1 event per household.</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td></td>
</tr>
</tbody>
</table>

Providing context and examples to build intuition, we discuss system setup requirements, data quality assurance, input file preparation, parameter setting recommendations derived from public health and statistical principles, and new features in the 2022 SaTScan release that simplify routine analyses. Multimedia Appendix 1 [2,13,14,25-37] provides further details about parameter settings and advanced options, demonstrates cluster interpretation using artificial but realistic data, and offers strategies for fine-tuning analyses. In Multimedia Appendix 2, a video demonstrates how to set up a cluster detection system, and sample input and output files are provided.

*RI: recurrence interval.*
Analysis Design

Ethical Considerations
The BCD’s prospective cluster detection activity was deemed nonresearch, public health surveillance by the NYC Department of Health and Mental Hygiene Institutional Review Board. Figures in this tutorial are for illustrative purposes only and reveal no confidential patient information.

Requirements and System Setup
A strong informatics infrastructure and uninterrupted data streams are key for a successful prospective reportable disease outbreak detection system. Near–real-time electronic laboratory reporting with patient location data is required, as well as epidemiologists to interpret signals and resources to investigate outbreaks.

Local installation of SaTScan is needed, and confidential information is never transmitted elsewhere. Google Earth installation and mail server information are optional requirements for certain SaTScan features.

To ensure optimal system performance, we recommend proactive data quality assurance practices, including monitoring for data feed interruptions, deduplicating events reported more than once, and correcting patient information (eg, correcting typographical errors in addresses failing to geocode, correcting when laboratories misreport patients’ residential addresses, maintaining geocoding reference files to account for newly constructed buildings, and correcting implausible onset dates).

SAS or R can be used to prepare input data, run SaTScan in batch mode, create output summary files and visualizations, and send email alerts. Sample code for these systems is freely available [2,38-40], but the 2022 SaTScan release eliminates the need for much of this complicated, external code.

Defining the Aim
Most routine prospective SaTScan analyses of reportable communicable diseases are designed to quickly detect and monitor spatiotemporal outbreaks using patients’ home addresses [2]. We refer to this as our base analysis. In these disease-specific analyses, cluster summaries and patient line lists are routed to in-house disease experts for interpretation and potential action to stop transmission or to focus public health resources. We believe this system is relevant for all acute infectious diseases except for diseases of which a single case is a public health emergency, such as anthrax or botulism. For a few diseases, we modified the base analysis (Table S1, Figure S1, and the Analysis Design: Supplement section in Multimedia Appendix 1).

Input Data
Overview
SaTScan detects data aberrations, which can represent true clusters or data errors. Hence, input data require attention to detail. A case file contains all disease events with their epidemiologically relevant location (eg, census tract of home address) and date (eg, illness onset date), and a coordinates file includes all locations (eg, census tracts) together with the latitude and longitude of their center points. We refer throughout to events as all reported disease events and cases as the subset that meet the surveillance case definitions established by the US Centers for Disease Control and Prevention and the Council of State and Territorial Epidemiologists [41].

Case File
The case file must contain a location ID (eg, census tract number), a temporal element such as a date, and the number of events at that location and time. The case file may be aggregated by location and date. Alternatively, it can be formatted as having 1 row per event and may contain additional columns that are not used in the analysis but can be output in a line list or used to group events on a map (refer to the Built-In SaTScan Implementation and Output Interpretation Features section).

For our base analyses, the temporal element is the best approximation of the date when patients became ill, prioritizing symptom onset date, diagnosis date, or report date such that there is no missingness in the temporal data element. The spatial location is the census tract of patient residence. Some degree of spatial aggregation is necessary to scan along a network (refer to the Network File section), and geocoding to census tracts is sufficiently fine resolution. Even if the public health action occurs at a larger spatial unit, using smaller geographic units increases both the precision and statistical power to quickly detect emerging outbreaks, including those that are geographically small or that do not conform to the boundaries of larger administrative areas. Patients who have a missing, inaccurate, or undefined spatial location can either be purposely excluded from the analysis or included by assigning them a location (eg, assigning patients whose home address is a post office box to the census tract of the post office holding their mail and assigning unsheltered patients experiencing homelessness to a unique, artificial census tract).

For some analyses, we restrict the case input file. For amebiasis, cryptosporidiosis, giardiasis, and shigellosis, we analyze all ages combined as well as children aged <5 years separately to detect outbreaks that might affect patients attending childcare programs. For diseases for which household transmission is likely, for example, norovirus and shigellosis, we restrict the case input file to include only 1 event per household (refer to the Difficulty Detecting Outbreaks Affecting a Particular Age Group and Clusters Are Driven By Within-Household Transmission sections in Multimedia Appendix 1).

If there is a spatiotemporal outbreak in the baseline, it may become harder to prospectively detect a subsequent outbreak. If a purely spatial adjustment is applied, either explicitly or implicitly as in the space-time permutation model, then it will be harder to detect subsequent outbreaks around the same location. This problem can be resolved by removing historic outbreaks from the input files (refer to the Difficulty Detecting New Outbreaks in Areas With Prior Outbreaks section in Multimedia Appendix 1).

We include all reported events during the study period specified in the parameter settings regardless of whether they are subsequently confirmed according to Centers for Disease Control and Prevention and Council of State and Territorial
Epidemiologists surveillance case definitions [41] upon investigation. By analyzing all reported events, timeliness is preserved in that events can be analyzed as soon as they are reported rather than waiting until investigation and case classification processes are complete. Ensuring consistent event inclusion criteria across the study period supports a valid comparison of total reporting volume between current and historical periods [42].

Coordinates File

The coordinates file has 1 row per location with a location ID and its geographic coordinates. Most BCD analyses use the latitude and longitude of each census tract centroid per census 2010 definitions.

Alternatively, space need not be conceptualized as geographical coordinates. To identify temporal clusters of infections by particular Salmonella serotypes, we replace space with arbitrary coordinates for serotype and detect citywide clusters of particular serotypes that could not be explained by the overall seasonality of Salmonella infections [25].

Network File

Spatial scan statistics can be used with windows of different shapes, including circular, elliptical, and nonparametric [43-46]. SaTScan is typically used with a circular scanning window, which also has very good statistical power to detect clusters of other geographical shapes and sometimes better power to detect irregularly shaped clusters than flexibly shaped spatial scan statistics [47]. This is because flexibly shaped spatial scan statistics without some form of noncompactness penalty will often detect spindly, octopus-like clusters that cherry-pick areas with high rates and with long, thin connections between them [47].

An enhancement in SaTScan version 10.0 is the option to scan locations along a network, which may represent travel distance between locations or the amount of interaction between different communities. For example, census tracts or zip code tabulation areas may be geographically close but separated by lakes, rivers, or mountain ranges. We recommend constructing a network file connecting neighboring locations unless separated by those barriers so that clusters can form around rather than through them (Figure 1; refer to Figure S2 and the Network File: Supplement section in Multimedia Appendix 1).

Figure 1. Illustrative cluster detected using a circular scanning window (in yellow) versus scanning along a network (in red) using the same input data. Scanning along a network accounts for limited connectivity across water bodies, resulting in more precise spatial clusters, such as illustrative cluster 4 in the Rockaway peninsula in Queens, New York City.
Parameter Settings

Unlike machine learning methods, in which hyperparameters are tuned using real data during model development, SaTScan parameter settings are derived from underlying public health and statistical principles that should be generalizable across jurisdictions.

Type of Analysis and Probability Model

We use the prospective space-time permutation scan statistic for almost all routine SaTScan analyses conducted at the BCD, including the base analysis [2,17,25]. Prospective analyses detect currently active unusual clusters, evaluating only the subset of possible clusters that encompass the last day of the study period together with a flexible number of previous days. Cluster statistics are determined by the counts of observed and expected events within a cluster window, which is bounded by the user-specified minimum and maximum temporal cluster size. For the space-time permutation scan statistic, the expected number of events for any area is a function of the observed counts during the baseline period and the jurisdiction-wide trend during the temporal window [17]. For example, if a collection of 5 census tracts has 2.3% of cases during the baseline period and there are 100 total cases in the jurisdiction during the cluster period, then the expected number of cases during the cluster period in those 5 census tracts is 2.3.

SaTScan compares a statistical likelihood function calculated from the real data sets with likelihoods calculated from random data sets generated under the null hypothesis of no clustering. There is a cluster if a geographical area has a higher proportion of its events in the cluster period compared with the previous baseline period or, equivalently, if a period has a higher proportion of its events within versus outside the geographical cluster area. For example, if Manhattan’s Upper East Side has 10% of all NYC events during the last 7 days but only 3% of all events during the previous months, a prospective space-time permutation analysis will detect a cluster in the Upper East Side during the most recent week. Note that, while such a cluster is most likely due to an outbreak or increased testing in the Upper East Side, it can also be generated by a sudden decrease in disease occurrence or ascertainment in the rest of the city. The former is more likely for geographically small clusters, whereas larger clusters could have either explanation.

The permutation in the space-time permutation model refers to keeping all observed locations and dates fixed when simulating random data sets under the null hypothesis while randomly shuffling which observed event locations are connected to which observed dates. By conditioning on locations and dates with observed events, the space-time permutation model automatically adjusts nonparametrically for any purely temporal patterns, including seasonality, secular trends, and day-of-week effects, as well as any purely geographical variations in disease occurrence, diagnosis, and reporting, such as due to insurance coverage and access to care. An area’s larger population size, which also affects disease incidence, is accounted for by having more disease counts observed in the baseline period.

For count data such as disease events, Bernoulli and Poisson models are also available. However, the space-time permutation model is often preferred because it requires only event data and automatically adjusts nonparametrically for both purely temporal and spatial variations (refer to the Bernoulli and Poisson Probability Models section and Figure S1 in Multimedia Appendix 1).

Space and Time Adjustments

If there is an overall increasing temporal trend, SaTScan may identify an outbreak if there is no temporal adjustment. One may want to adjust for naturally occurring seasonal variation, for example, while still detecting local outbreaks not explained by jurisdiction-level seasonal trends. When applying a spatial adjustment, SaTScan will identify clusters in which case counts are increasing faster or decreasing slower than elsewhere in the jurisdiction even if the absolute level is lower than that in surrounding areas.

The space-time permutation model automatically adjusts for purely temporal patterns and purely spatial variations, unlike the Bernoulli and Poisson models, for which these adjustments must be explicitly requested. We additionally adjust for space by day-of-week interaction in our base analysis because the day-of-week pattern of health care-seeking behavior and diagnoses may vary geographically in NYC. With this adjustment, the dates in which events occurred are shuffled and assigned to the original event locations, restricting them to the same day of the week. This adjustment should only be applied when there are several weeks in the study period; with only a few weeks, the day-of-week–specific estimates become too unstable. We do not use covariate adjustment because this is unnecessary for prospective outbreak detection and doing so could adjust away emerging outbreaks affecting a particular demographic group.

Scanning for High or Low Rates

Most BCD analyses scan for areas with high rates as we are interested in detecting outbreaks of excess events. The one exception is data quality analyses, where we scan for low rates to quickly identify laboratories with unusually low reporting volume [26].

Minimum Number of Events

In the base analysis, we require only 2 events at minimum because we want to quickly detect outbreaks before many events accrue. For example, 2 salmonellosis events on the same day residing on the same block could be of great interest. For COVID-19, we set the minimum number of events in a cluster to 100 when volume was high because the purpose was to support resource allocation to areas with high and increasing disease transmission rather than to conduct case investigations and contact tracing to identify a common exposure source.

Maximum Spatial Cluster Size

The option that imposes the fewest assumptions is to allow the cluster to expand in size to include up to 50% of all events during the study period. This is reasonable for modestly sized areas such as NYC but might yield too large and uninformative clusters in national- or state-level jurisdictions. For geographically large study areas, we recommend either using 10% as the maximum geographical cluster size or adding another
maximum geographical cluster size. For example, a maximum of 50% may appropriately limit the cluster size around Anchorage, whereas Western Alaska may be better served with a 200-km maximum cluster radius. When using a network file and defining a geographical maximum spatial cluster size, a cluster will expand until reaching a network location at the specified distance from the cluster centroid, which allows for irregularly shaped clusters. We recommend erring on the side of searching for larger clusters, followed by using other principled approaches to interpret or prevent large clusters (refer to the System Fine-Tuning section in Multimedia Appendix 1).

The BCD has on occasion applied a maximum spatial cluster size of a 1-km radius. To search for clusters of locally acquired Zika virus infection, we assumed that a maximum 1-km radius was consistent with the typical flight distance of an *Aedes* mosquito [27] plus some additional distance to include residences of people moving around and coming into contact with the same infected mosquito. To search for clustering of legionellosis events around cooling towers to guide environmental sampling [13], we considered a 1-km radius to be consistent with the area of highest risk of exposure around a contaminated cooling tower.

Of note, setting the maximum spatial cluster size is different from applying a maximum reported spatial cluster size in the advanced output features, which limits the clusters reported in the output rather than limiting the clusters evaluated by the analysis. While it is OK to vary the maximum reported spatial cluster size, an analysis should never be repeated with a different maximum spatial cluster size as *P* values and recurrence intervals (RIs) are inaccurate if one manipulates and tries multiple such parameter settings [48].

**Minimum and Maximum Temporal Cluster Size**

The base analysis scans for clusters that are between 2 and 30 days long, which is enough to encompass the upslope of an epidemic curve for point-source outbreaks, typically spanning a few days or weeks depending on the pathogen. We extended the maximum temporal cluster size to 60 days for salmonellosis, Shiga toxin–producing *Escherichia coli*, paratyphoid fever, and typhoid fever and to 120 days for listeriosis to align with PulseNet USA definitions based on clustering in time and genetic relatedness [49]. Although campylobacteriosis, shigellosis, noncholera *Vibrio spp.* infection, and yersiniosis are also PulseNet organisms, we kept the maximum temporal cluster size at 30 days to slightly increase power to detect more focal clusters of these diseases. These diseases are higher volume and commonly detected using culture-independent diagnostic testing such that laboratory subtyping is not routinely available in NYC to support defining clusters. We similarly extended the maximum temporal cluster size to 60 days for hepatitis A because the incubation period can extend to 50 days [50] and to allow additional time for patient care seeking.

A maximum of 30 or 60 days may not be long enough for diseases with long incubation periods, extended propagated transmission, or intermittent common-source outbreaks. For example, exposure to point sources of listerioses can be intermittent when persistent environmental contamination, and the incubation period can be nearly 70 days [28], so for this disease, we scan for clusters between 7 and 365 days at weekly aggregation.

In the base analysis, we set the minimum temporal cluster size to 2 days because few true outbreaks of the diseases we monitor span only 1 day, and not evaluating clusters of only 1-day duration slightly improves power given less statistical adjustment for multiple comparisons. However, note that a 2-day cluster can have all cases on the same day. If there will be no public health action unless a cluster has persisted for some time, the minimum temporal cluster size can be lengthened. For COVID-19, for example, we scanned for clusters with a minimum temporal size of 14 days to support resource reallocation to areas with sustained high test positivity.

**Study Period**

It is important to balance the need for a period long enough to establish a stable local baseline for each spatial unit yet short enough to avoid variable secular trends due to, for example, geographical population shifts over time. As a rule of thumb, the study period should preferably be at least 3 times as long as the maximum temporal cluster size. This is analogous to a case-control study, which has diminishing returns in statistical power when the control-to-case ratio is increased beyond approximately 3 [51]. If there is a permanent change in surveillance so that recent and older data are no longer comparable, the study period should be shortened. For example, laboratories adopting a culture-independent diagnostic test can lead to increased case ascertainment, and temporarily shortening the study period to begin after the new test was adopted can restore consistency in the baseline [29] (refer to the Clusters Are Driven by the Changing Nature of Input Data section in Multimedia Appendix 1).

For timeliness, the end of the study period should be set to the previous day or the most recent day with sufficiently complete data. In the base analysis, the study period is 1 year with an end date set to the day before the analysis is run. Although 1 year is >3 times longer than the maximum temporal cluster size of 30 days, case ascertainment is generally consistent over a 1-year period for most diseases, and the longer study period provides a more stable baseline. For COVID-19, we shortened the study period to 63 days because volatility in testing access and outreach during the public health emergency made older data less comparable, and we set the end of the study period to 3 days before the run date because more recent testing data were largely missing [14].

**Time Aggregation**

With thousands of locations, long study periods, and multiple data streams, space-time SaTScan analyses can be computationally intensive. To reduce computing time, data may be aggregated into longer time intervals by, for example, aggregating daily data into weekly data. Another reason for aggregating to time intervals of 7 days is to remove day-of-week effects. For the base analysis, we use a time aggregation of 1 day to detect outbreaks as quickly as possible, increase the statistical power to detect outbreaks that do not conform to weekly or other prespecified time intervals, and obtain a more precise estimate of the cluster start date.
**Frequency of Analyses**

With near–real-time electronic laboratory reporting, we run analyses daily. If prospective analyses are conducted at a different frequency than the aggregated time units, such as weekly analyses with daily data, the analysis frequency must be specified. This may be needed if data are reported weekly but with daily resolution; if weekly analyses are sufficiently frequent given limited staffing capacity; or to support public health objectives other than acute outbreak investigations, such as guiding resource allocation. It is important to monitor for unusual clusters near continuously (eg, hourly analyses to detect early indications of a possible bioterrorist attack), consider inputting data using the generic time precision option.

**Secondary Clusters**

Any disease may have multiple active outbreaks at any moment, so secondary clusters should be reported by enabling the Most Likely Clusters, Hierarchically option. We use the No Cluster Centers in Other Clusters option rather than the default No Geographical Overlap. This allows for reports of slightly overlapping clusters so that clusters located close to each other can be detected and can also help define the extent of irregularly shaped outbreaks. At the same time, it avoids redundant clusters that are almost identical to each other as well as clusters almost completely driven by the primary cluster but with large outlying areas with no or modest excess risk. Epidemiological judgment must be used to determine whether events in overlapping clusters are attributable to a single or multiple outbreaks or whether the secondary cluster is of limited interest. It may not be worthwhile to investigate every event in a large secondary cluster with a much lower RI that has a large proportion of events that are also in the primary cluster.

**RIs and Inference**

Monte Carlo hypothesis testing compares the maximum likelihood value for the real data with the maximum likelihoods from each of the random replicas of the data set. For prospective analyses, SaTScan assigns an RI to each cluster. The RI is an alternative to the P value (RI=1/P value), and the greater the RI, the less likely the cluster is due to chance. If a cluster has an RI of 1 year, then under the null hypothesis and during any 1-year period, the expected number of false signals with the same and higher magnitude is 1. For rare diseases, this expected number is lower depending on cluster size restrictions. While P values are commonly used for determining statistical significance, the typical cutoff of P<.05 is not meaningful for prospective analyses. With daily analyses for 1 disease, applying a P<.05 threshold would generate 1 false signal every 20 days on average, or approximately 18 false signals per year, which is too frequent.

Increasing the number of Monte Carlo replications generating random data sets under the null hypothesis slightly increases the statistical power of the scan statistic but also increases the run time. Our base analysis uses 999 replications, with the default P value option to end early if the P value is large. The number of replications must be at least 999 to avoid unnecessary loss of power.

**Signaling Threshold**

In the base analysis, we set the signaling threshold to RI≥100 days. We think of a cluster with an RI of 100 to <365 days as a weak cluster, a cluster with an RI of 365 days to <5 years as a moderate cluster, a cluster with an RI of 5 to <100 years as a strong cluster, and a cluster with an RI of ≥100 years as a very strong cluster. P values or RI thresholds determine which clusters trigger an alert, but they should be considered alongside other factors to establish whether a cluster is of public health importance [52]. While we have suggested rules of thumb for RI interpretation, investigators should holistically interpret other cluster characteristics and apply epidemiological judgment considering the disease severity, relative risk, location, period, and patient characteristics in the line list (refer to the Cluster Investigation and Response section).

**Assessment of Analysis Design**

Our SaTScan parameter settings are derived from underlying public health and statistical principles. While we believe that our recommendations are generalizable, parameter settings may require adjustment for other diseases and jurisdictions. When using real-world rather than simulated data, there is generally no gold standard that can be used to evaluate results. For local proof of concept, one way to determine whether an analysis needs adjustments is to identify at least one known historical acute, localized, community-acquired outbreak and mimic prospective surveillance around that period to assess the speed and precision with which the outbreak would have been detected.

If an analysis consistently produces results that are unsatisfactory and data quality issues have been ruled out as a cause, then the input files or parameter settings may need fine-tuning. Issues we have encountered include missed or delayed outbreak detection, clusters too large to be useful, and too many signals to be actionable given available resources (refer to the System Fine-Tuning section in Multimedia Appendix 1).

**Built-In SaTScan Implementation and Output Interpretation Features**

SaTScan version 10.1 introduced several built-in features to simplify adopting and implementing routine prospective analyses and streamline output interpretation. These include a tool for automating multiple prospective analyses, temporal and geographical visualizations of clusters and events, line lists with information about cluster events, and a feature that allows users to send automated email alerts summarizing analysis results (refer to the Cluster Output Interpretation section in Multimedia Appendix 1 and video demonstration in Multimedia Appendix 2).

**Multiple Analyses**

When independently monitoring multiple diseases, running multiple analyses is necessary. This is easily managed using the multiple analyses feature. The only preanalysis coding required is to generate the input files in the appropriate format (refer to the video demonstration in Multimedia Appendix 2).
Temporal Graphs
Temporal graphs are useful for visualizing epidemic curves. SaTScan can produce temporal graphs that display observed and expected event counts inside and outside the geographical cluster area, marking the period before and during the cluster (Figure 2; refer to the Temporal Graphs: Supplement section in Multimedia Appendix 1).

Figure 2. SaTScan-generated temporal graph depicting the observed and expected event counts in the geographical area of the cluster during (band shaded in gray) and before the cluster temporal window.

Visualizing Clusters and Cases on a Map
SaTScan can produce maps for visualizing the spatial extent of clusters in HTML, KML, and shapefile formats. Locations of events during the study period can be added to the map and can be grouped by any variable (eg, case status, age group, and gender), distinguished using different icons. A legend that distinguishes events as being Inside Cluster, new entry; Inside Cluster, not new entry; or Outside Clusters exceeding the RI signaling threshold can be displayed. Events are also distinguished as Recent versus Prior to a date specified on the Recent Events time slider in the HTML file, by default the start date of the most likely cluster (Figure 3; refer to the Visualizing Clusters and Cases on a Map: Supplement section in Multimedia Appendix 1).

Figure 3. HTML visualization produced by SaTScan depicting 4 clusters exceeding the signaling threshold with (artificially generated) events mapped to their exact locations. Events are distinguished by disease status and whether they are newly identified in a cluster exceeding the signaling threshold. Disease status may be replaced with any categorical variable included in the case input file.

Drilldown Analysis
Geographically large clusters can be further investigated using the SaTScan drilldown tool to determine whether events are randomly distributed or clustered within a cluster (Figure 4; refer to the Drilldown Analysis: Supplement section in Multimedia Appendix 1).
**Line List of Cluster Events**

When a cluster is detected, SaTScan can produce a line list of all events in clusters exceeding the RI signaling threshold with information such as age, gender, and location of each event. This may help with determining whether an outbreak investigation is warranted, as well as with the investigation itself (refer to the Line List of Cluster Events: Supplement section in Multimedia Appendix 1).

Examining the line list is also an opportunity to check for data quality issues and other unintended drivers of signals, such as duplicate individuals, erroneous addresses, multiple events in the same household, events that do not ultimately meet the surveillance case definition, or events for which it can be determined that an available address is unrelated to an exposure site (eg, a person experiencing homelessness whose address is a hospital). Because the purpose of the system is to detect ongoing community clusters of disease, an analysis may be rerun excluding these events to determine whether the community cluster persists. At the BCD, we proactively check for and correct data quality issues (refer to the Requirements and System Setup section), but inspecting the line list can detect problems missed by other routine checks.

In addition, the line list may be useful for filtering detected clusters by the number of events with certain statuses, for example, requiring a minimum number of confirmed or probable cases to investigate. At the BCD, although we scan for clusters with as few as 2 events, we normally require ≥3 confirmed, probable, suspected, or pending events to warrant further investigation. Thus, producing a line list that includes a case status field can support cluster prioritization.

**Automated Email Alerts**

The email alert feature can be used to automatically notify up to 2 groups of recipients of analysis results. One email can notify that the analysis has been completed, irrespective of the results, and another email can notify the same or a different group of recipients when a cluster exceeds the RI signaling threshold (refer to the Email Alerts section in Multimedia Appendix 1).

**Cluster Investigation and Response**

A well-developed cluster detection system focuses staff attention on emerging outbreaks to catalyze field investigations [53]. At the BCD, disease experts interpret cluster output and determine whether to launch an investigation and response, which, depending on the disease, might involve prioritizing patient
interviews, conducting environmental investigations (eg, inspecting cooling towers or food service establishments), and conducting community outreach and education to rapidly identify outbreak sources and interrupt ongoing transmission.

**Discussion**

**Overview**

This tutorial is intended to support public health officials in understanding the details of the successful communicable disease outbreak detection system in NYC [2,12-14] and how to adapt and fine-tune a similar system for their own jurisdictions. For early outbreak detection, spatiotemporal methods are preferable to purely temporal methods (eg, the refined historical limits method [42] and other time-series analyses) because, if data are analyzed at one resolution (eg, by county and week) while an outbreak is emerging at a different resolution (eg, a collection of a few census tracts over a few days), the outbreak may be difficult to detect. Unlike other software for spatiotemporal cluster detection (eg, the ArcGIS Emerging Hot Spot Analysis tool [54]), SaTScan analyzes data using scan statistics to search flexibly over time and space rather than within arbitrary administrative boundaries while accounting for multiple comparisons. SaTScan also avoids the modifiable areal unit problem [55] when using input data at the finest geographic resolution available, allowing for more precise identification of areas with elevated rates, and has options for spatial adjustments and nonparametric temporal adjustments to account for jurisdiction-wide trends. SaTScan version 10.1 includes multiple enhancements allowing for increased customizability, automation, and output visualization, which have not previously been described in real-world applications. Assuming input data are timely, complete, and accurate, the primary limitation of the system is diminished power to detect spatiotemporal outbreaks involving very few patients or involving patients located in a long and narrow area, such as along a river [56].

**Health Equity**

Our cluster detection system supports the NYC Department of Health and Mental Hygiene’s mission to protect and promote the health of all New Yorkers. Many communicable diseases disproportionately affect residents of high-poverty areas [57]. The faster that public health officials can detect outbreaks anywhere, the sooner disease transmission can be interrupted to support health equity and harmful environmental exposures can be remediated. To minimize the number of patients excluded from spatiotemporal analyses, we maintain data quality by proactively correcting typographical errors in addresses failing to geocode and assigning persons who are unsheltered and without a geocodable address to a separate category. By selecting the space-time permutation probability model for our base analyses, we account for purely spatial patterns, such as areas with comparatively lower access to care and testing for reportable diseases. Nevertheless, if very few patients in an outbreak access care (eg, as a consequence of systemic racism), then outbreak detection may be delayed or missed entirely. Hence, better and wider access to health care will not only benefit those patients but also strengthen public health surveillance and response efforts that benefit all residents.

**Future Directions**

Going forward, we advocate for more centralized analytic technical support for outbreak detection for state, tribal, local, and territorial public health departments in keeping with the Data Modernization Initiative [58] and other public health system investments. We hope to provide video tutorials to orient new users and foster a community of users to share knowledge and best practices. While we believe our recommendations should be generalizable to other jurisdictions in the United States and internationally, our practical experience is limited to monitoring certain reportable communicable diseases in one dense, urban area. To better understand generalizability, additional jurisdictions will similarly need to add their voices to this conversation and contribute their experiences using this system. We are particularly interested in the experiences of jurisdictions covering larger geographic and rural areas; experiences using the network locations file with user-specified, non-Euclidean distances; and the adaptation of this system for additional jurisdictions will similarly need to add their voices to this conversation and contribute their experiences using this system. We are particularly interested in the experiences of jurisdictions covering larger geographic and rural areas; experiences using the network locations file with user-specified, non-Euclidean distances; and the adaptation of this system for other data streams for infectious diseases, including syndromic surveillance, social media, wastewater, antimicrobial susceptibility testing, and veterinary data.

**Acknowledgments**

ALR was supported by the Epidemiology and Laboratory Capacity for Prevention and Control of Emerging Infectious Diseases Cooperative Agreement (ELC) Coronavirus Aid, Relief, and Economic Security Act (CARES; grant NU50CK000517-01-09) funded by the US Centers for Disease Control and Prevention (CDC). ERP and SKG were supported by the Public Health Emergency Preparedness cooperative agreement (grant NU90TP922035-03-03) funded by the CDC. MK and SH were supported by the SaTScan Enhancements Project managed by the Fund for Public Health in New York City and funded by CDC ELC CARES, the CDC Foundation, and the Alfred P. Sloan Foundation. The authors thank the many colleagues at the New York City Department of Health and Mental Hygiene who have provided feedback and suggestions on the BCD’s communicable disease cluster detection system, particularly Julia Latash for reviewing this tutorial and Gretchen Culp for contributing with geographical expertise. The authors did not use generative artificial intelligence for any portion of manuscript writing.

**Authors’ Contributions**

ALR is a cluster detection analyst at the Bureau of Communicable Disease (BCD) and the lead analyst of the SaTScan Enhancements Project. MK provided statistical and analytical support. ERP developed disease-specific analytic customizations and features to support cluster interpretation and investigation at the BCD, including line lists and dynamic maps, which formed the basis for
outputs now directly produced by SaTScan as of version 10.1. MK and SH provided technical software support to the project. SG has overseen the BCD’s reportable disease spatiotemporal cluster detection analyses since 2012 and is principal investigator of the SaTScan Enhancements Project, which implemented several of the analytic enhancements described in this tutorial. ALR, MK, and SG drafted the tutorial. All authors critically reviewed and revised the tutorial for important intellectual content and gave final approval for the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Further details about parameter settings and guidance for advanced options, output interpretation, and strategies for fine-tuning an analysis.

[PDF File (Adobe PDF File), 411 KB - publichealth_v10i1e50653_app1.pdf ]

Multimedia Appendix 2

Video that demonstrates how to set up a cluster detection system and sample input and output files.

[ZIP File (Zip Archive), 108502 KB - publichealth_v10i1e50653_app2.zip ]

References


3. Kulldorff M. SaTScan v10.1: software for the spatial and space-time scan statistics. SaTScan. [accessed 2022-10-04]


40. Bemis K. RSaTScan cluster detection. GitHub. URL: https://github.com/kb230557/rsatscan-cluster-detection [accessed 2023-04-12]

Abbreviations

BCD: Bureau of Communicable Disease
NYC: New York City
RI: recurrence interval
Investigating the Interrelationships Among Mental Health, Substance Use Disorders, and Suicidal Ideation Among Lesbian, Gay, and Bisexual Adults in the United States: Population-Based Statewide Survey Study

Alex Siu Wing Chan¹*, PhD; Hon Lon Tam², PhD; Florence Kwai Ching Wong³, PhD; Gordon Wong⁴, MD; Lok Man Leung¹*, MSc; Jacqueline Mei Chi Ho⁵, MSc; Patrick Ming Kuen Tang⁶, PhD; Elsie Yan¹*, PhD

¹Department of Applied Social Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China
²Nethersole School of Nursing, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China
³Factor Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada
⁴School of Clinical Medicine, Department of Psychiatry, The University of Hong Kong, Hong Kong, China
⁵School of Nursing, Faculty of Health and Social Sciences, Hong Kong Polytechnic University, Hong Kong, China
⁶Department of Anatomical and Cellular Pathology, Prince of Wales Hospital, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

* these authors contributed equally

Corresponding Author:
Hon Lon Tam, PhD
Nethersole School of Nursing
Faculty of Medicine
The Chinese University of Hong Kong
ELB704B
Hong Kong, 999077
China
Phone: 852 39439306
Email: hltam@cuhk.edu.hk

Abstract

Background: Mental health disparities have been documented among lesbian, gay, and bisexual (LGB) adults in the United States. Substance use disorders and suicidal ideation have been identified as important health concerns for this population. However, the interrelationships among these factors are not well understood.

Objective: This study aims to investigate the interrelationships among mental health, substance use disorders, and suicidal ideation among LGB adults in the United States using a population-based statewide survey.

Methods: Our study was an observational cross-sectional analysis, and the data for this study were collected from a sample of LGB adults who participated in the statewide survey. The survey collected information on mental health, substance use disorders, and suicidal ideation using validated measures. Descriptive statistics and inferential data analysis were conducted to explore the interrelationships among these factors.

Results: The results showed that LGB adults who reported higher levels of depression and drug abuse and dependence also reported higher levels of suicidal tendency and mental illness. Inferential data analysis using $\chi^2$ tests revealed significant differences in depression score ($\chi^2=458.241; P<.001$), drug abuse and dependence score ($\chi^2=226.946; P<.001$), suicidal tendency score ($\chi^2=67.795; P<.001$), and mental illness score ($\chi^2=363.722; P<.001$) among the 3 sexual identity groups. Inferential data analysis showed significant associations between sexual identity and mental health outcomes, with bisexual individuals reporting the highest levels of depression, drug abuse and dependence, suicidal tendency, and mental illness.

Conclusions: This study provides important insights into the interrelationships among mental health, substance use disorders, and suicidal ideation among LGB adults in the United States. The findings underscore the need for targeted interventions and research aimed at addressing the mental health needs of sexual minority populations. Future research should aim to better understand the underlying mechanisms driving these disparities and develop culturally sensitive and tailored interventions that
Introduction

Lesbian, gay, and bisexual (LGB) individuals encounter a multitude of health disparities stemming from their marginalized status and the strain of minority stress [1-3]. The focal points of concern encompass mental health, substance use disorders, and suicidal ideation within the LGB community [4-6]. Despite the recognition of these issues, there remains a discernible void in the comprehensive exploration of the intricate interplay among these concerns among LGB individuals within the United States.

Within the context of the LGB community, the nexus of mental health, substance use disorders, and suicidal ideation presents a complex landscape. Existing research underscores that LGB individuals are confronted with an elevated risk of grappling with mental health challenges—ranging from depression and anxiety to posttraumatic stress disorder—a consequence of societal stigma, discrimination, and violence [6-8]. The resultant minority stress often propels substance use disorders as a coping mechanism, manifesting in a higher prevalence of substance abuse and dependency among LGB individuals in contrast to the general population [9-11].

Furthermore, the vulnerability of LGB individuals to suicidal ideation and attempts surpasses that of the broader populace. The multifaceted risk factors underpinning suicidal ideation within this demographic encompass discrimination, victimization, familial or social rejection, internalized homophobia, and compromised mental health [12,13]. Remarkably, substance use disorders have been shown to further amplify the propensity for suicidal ideation and attempts among LGB individuals [5,14,15].

The intricate interlinking of mental health, substance use disorders, and suicidal ideation among LGB individuals is underscored by a nuanced interdependence [4]. For instance, compromised mental health may act as a precursor to substance use disorders, in turn fostering a milieu for the emergence of suicidal ideation. In a different trajectory, the direct influence of discrimination and victimization can accentuate both mental health adversities and heightened substance use [16-19]. In light of these multifaceted dynamics, a holistic approach to address mental health, substance use disorders, and suicidal ideation becomes imperative within the LGB community.

To this end, comprehensive interventions and policies targeting the underlying causative factors—discrimination and minority stress—are paramount to curtailting the prevalence of mental health issues, substance use disorders, and suicidal ideation among LGB individuals. Bridging this research gap, this study is dedicated to an in-depth exploration of the interrelationships among mental health, substance use disorders, and suicidal ideation among LGB adults within the United States. Leveraging a population-based statewide survey conducted in 2019 through an extensive web-based questionnaire, our study seeks to illuminate both the prevalence of these issues and the intricate associations binding them. Through these insights, we aspire to foster the development of efficacious interventions and policies designed to enhance the health and well-being of LGB individuals across the United States. The outcomes of this investigation are poised to advance the comprehension of the intricate interplay among mental health, substance use disorders, and suicidal ideation within the LGB community, thereby serving as a guiding compass for the implementation of evidence-based strategies to alleviate the burden of these issues within this population.

Methods

Setting

The study used data from the 2019 National Survey on Drug Use and Health (NSDUH), a nationally conducted population-based survey covering noninstitutionalized individuals across all 50 states within the United States.

Population

The focus was on adults aged 35 years and above, resulting in a final sample size of 59,864 individuals.

Sample

The sample consisted of individuals who participated in the 2019 NSDUH, a data set accessible through the web-based repository provided [20].

Sampling

The NSDUH used a multistage probability sampling approach for each state to ensure a representative sample of the diverse population [21].

Recruitment Process

Participants were selected through computer-assisted personal interviews and audio self-interviews, aiming to enhance the credibility of responses, especially on sensitive aspects of the survey. A remuneration was provided upon successful completion of the interview.

Data Collection

The study used a comprehensive analysis of the NSDUH data set, covering domains such as mental health, substance use, and suicidal ideation.

Data Analysis

The analysis involved the use of SPSS software (version 27.0; IBM). An observational cross-sectional analysis was conducted,
using 4 derived variables: depression score, drug abuse and dependence score, suicidal tendency score, and mental illness score. These variables were calculated from specific questions within the survey, allowing a nuanced exploration of mental health within the LGB community [22]. Descriptive statistics in this study illustrated the distribution of depression scores, drug abuse and dependence scores, suicidal tendency scores, and mental illness scores across different sexual identities. The chi-square test of independence demonstrated the association between sexual identity and socioeconomic and health characteristics, revealing significant differences in total family income, poverty level, and overall health across sexual identity groups. Inferential data analysis indicated significant differences in mean ranks of depression score, drug abuse and dependence score, suicidal tendency score, and mental illness score across heterosexual, lesbian or gay, and bisexual groups.

Variables
The meticulous assessment of mental health within the context of heterosexual, gay, and bisexual men in the United States engenders the use of 4 distinct derived variables [23].

Depression Score
Calculated through an intricate amalgamation of 8 questions sourced from the adult depression module of the survey. This module is broadly acknowledged in research circles for its efficacy in quantifying the severity of depressive symptoms and their reverberations upon an individual’s quality of life.

Drug Abuse and Dependence Score
This metric is derived through an intricate synthesis of 13 carefully curated questions, explicitly tailored to assess alcohol and drug abuse and dependence within the adult demographic. The strategic inclusion of this variable attests to the complex interrelationships between substance abuse and compromised mental health outcomes.

Suicidal Tendency Score
Forged from a set of 3 core questions, this variable delves into respondents’ perceptions and emotions surrounding suicide.

Mental Illness Score
Given the profound gravity of this issue within the realm of mental health, this variable emerges as a pivotal yardstick in evaluating the prevalence of suicidal tendencies within the study cohort.

Descriptive Analysis
The descriptive analysis results (Table 1) show the distribution of depression score, drug abuse and dependence score, suicidal tendency score, and mental illness score across different sexual identities. The sample consisted of 59,864 individuals, of which 96.2% (n=57,575) identified as heterosexual, 1.7% (n=1017) identified as lesbian or gay, and 2.1% (n=1272) identified as bisexual.

Table 1. Descriptive statistics of depression, drug abuse and dependence, suicidal tendency, and mental illness scores by sexual identity.

<table>
<thead>
<tr>
<th>Sexual identities</th>
<th>Respondents, n (%</th>
<th>Depression score (range 0-1), mean (SD)</th>
<th>Drug abuse and dependence score (range 0-0.85), mean (SD)</th>
<th>Suicidal tendency score (range 0-2), mean (SD)</th>
<th>Mental illness score (range 0-5), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterosexual</td>
<td>57,573 (96.2)</td>
<td>0.5549 (0.22978)</td>
<td>0.0067 (0.03306)</td>
<td>0.4763 (0.24216)</td>
<td>0.5932 (1.14332)</td>
</tr>
<tr>
<td>Lesbian or gay</td>
<td>1038 (1.7)</td>
<td>0.6387 (0.25316)</td>
<td>0.0137 (0.05186)</td>
<td>0.5119 (0.30884)</td>
<td>0.9518 (1.44794)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>1253 (2.1)</td>
<td>0.6972 (0.25646)</td>
<td>0.0208 (0.06673)</td>
<td>0.5890 (0.40123)</td>
<td>1.3113 (1.68115)</td>
</tr>
<tr>
<td>Total</td>
<td>59,864 (100)</td>
<td>0.5594 (0.23192)</td>
<td>0.0071 (0.03458)</td>
<td>0.4792 (0.24839)</td>
<td>0.6144 (1.16840)</td>
</tr>
</tbody>
</table>

The mean depression score for heterosexual individuals was 0.5549 (SD 0.22978), while the mean depression score for lesbian or gay individuals was 0.6387 (SD 0.25316), and for bisexual individuals it was 0.6972 (SD 0.25646). The mean drug abuse and dependence score for heterosexual individuals was 0.0067 (SD 0.03306), while the mean score for lesbian or gay individuals was 0.0137 (SD 0.05186), and for bisexual individuals it was 0.0208 (SD 0.06673). The mean suicidal tendency score for heterosexual individuals was 0.4763 (SD 0.24216), while the mean score for lesbian or gay individuals was 0.5119 (SD 0.30884), and for bisexual individuals it was 0.5890 (SD 0.40123). The mean mental illness score for heterosexual individuals was 0.5932 (SD 1.14332), while the mean score for lesbian or gay individuals was 0.9518 (SD 1.44794), and for bisexual individuals it was 1.3113 (SD 1.68115).
The results indicate that individuals who identified as lesbian, gay, or bisexual reported higher scores on all 4 variables compared to those who identified as heterosexual. The percentage of the total number for each sexual identity group was consistent across all 4 variables. The minimum and maximum scores for depression, drug abuse and dependence, and suicidal tendencies were 0 and 1, while the maximum score for mental illness was 5. The SD for all 4 variables was highest for bisexual individuals, indicating greater variability in scores for this group. The mean and SD of all 4 variables are presented for the overall sample, which can serve as a reference for future comparisons with other populations.

**Chi-Square Test of Independence**

The chi-square test of independence was conducted to examine the association between sexual identity and 3 characteristics: total family income, poverty level, and overall health. The sample consisted of 59,864 individuals, of whom 96.2% (n=57,573) identified as heterosexual, 1.7% (n=1038) identified as lesbian or gay, and 2.1% (n=1253) identified as bisexual.

The results of the chi-square test of independence (Table 2) revealed that sexual identity was significantly associated with total family income ($\chi^2_{6}=97.704; P<.001$), poverty level ($\chi^2_{4}=117.163; P<.001$), and overall health ($\chi^2_{8}=32.945; P<.001$). Specifically, higher proportions of LGB individuals reported lower total family income and higher poverty levels compared to heterosexual individuals. For example, 13.2% (n=7892) of individuals reporting a total family income of less than US $20,000 were heterosexual, while 0.3% (n=180) of individuals reporting the same income level were lesbian or gay and 0.4% (n=263) were bisexual. Similarly, 11.3% (n=6767) of individuals reporting living in poverty were heterosexual, while 0.2% (n=141) were lesbian or gay and 0.4% (n=244) were bisexual (Table 2). The associations between sexual identity and total family income and poverty level were statistically significant at the $P<.001$ level.

Regarding overall health in Table 2, there were statistically significant differences in the percentage of individuals reporting excellent, very good, good, fair, and poor health across sexual identity groups. However, the differences were relatively small, and the associations were statistically significant at the $P<.05$ level. Overall, these findings suggest that sexual identity is associated with socioeconomic status and health, with LGB individuals experiencing greater social and economic disadvantage compared to heterosexual individuals.

These additional results provide more detailed information about the distribution of each characteristic within and across sexual identity groups, which may be useful for understanding patterns of inequality and designing interventions aimed at reducing disparities.
Table 2. Association between sexual identity and socioeconomic and health characteristics ($\chi^2$ test results; N=59,864). Percentages represent comparisons within different groups and sexual identity categories in the data.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Heterosexual (n=57,573, 96.2%)</th>
<th>Lesbian or gay (n=1038, 1.7%)</th>
<th>Bisexual (n=1253, 2.1%)</th>
<th>Chi-square (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total family income</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than US $20,000 (n=8335)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>7892 (13.2)</td>
<td>180 (0.3)</td>
<td>263 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Percentage within income group</td>
<td>94.7</td>
<td>2.2</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>13.7</td>
<td>17.3</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>US $20,000-US $49,999 (n=16,499)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>15,838 (26.5)</td>
<td>266 (0.4)</td>
<td>395 (0.7)</td>
<td></td>
</tr>
<tr>
<td>Percentage within income group</td>
<td>96</td>
<td>1.6</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>27.5</td>
<td>25.6</td>
<td>31.5</td>
<td></td>
</tr>
<tr>
<td>US $50,000-US $74,999 (n=9621)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>9265 (15.5)</td>
<td>165 (0.3)</td>
<td>191 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Percentage within income group</td>
<td>96.3</td>
<td>1.7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>16.1</td>
<td>15.9</td>
<td>15.2</td>
<td></td>
</tr>
<tr>
<td>US $75,000 or more (n=25,409)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>24,578 (41.1)</td>
<td>427 (0.7)</td>
<td>404 (0.7)</td>
<td></td>
</tr>
<tr>
<td>Percentage within income group</td>
<td>96.7</td>
<td>1.7</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>42.7</td>
<td>41.1</td>
<td>32.2</td>
<td></td>
</tr>
<tr>
<td><strong>Poverty level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living in poverty (n=7152)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>6767 (11.3)</td>
<td>141 (0.2)</td>
<td>244 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Percentage within income group</td>
<td>94.6</td>
<td>2</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>11.8</td>
<td>13.6</td>
<td>19.5</td>
<td></td>
</tr>
<tr>
<td>Income up to twice poverty line (n=11,067)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>10,625 (17.7)</td>
<td>150 (0.3)</td>
<td>292 (0.5)</td>
<td></td>
</tr>
<tr>
<td>Percentage within income group</td>
<td>96</td>
<td>1.4</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>18.5</td>
<td>14.5</td>
<td>23.3</td>
<td></td>
</tr>
<tr>
<td>Income more than twice poverty line (n=41,645)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>40,181 (67.1)</td>
<td>747 (1.2)</td>
<td>717 (1.2)</td>
<td></td>
</tr>
<tr>
<td>Percentage within income group</td>
<td>96.5</td>
<td>1.8</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>69.8</td>
<td>72</td>
<td>57.2</td>
<td></td>
</tr>
<tr>
<td><strong>Overall health</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent (n=11,473)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>11,094 (18.5)</td>
<td>194 (0.3)</td>
<td>185 (0.1)</td>
<td></td>
</tr>
<tr>
<td>Percentage within health condition</td>
<td>96.7</td>
<td>1.7</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>19.3</td>
<td>18.7</td>
<td>14.8</td>
<td></td>
</tr>
<tr>
<td>Very good (n=21,091)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total, n (%)</td>
<td>20,297 (33.9)</td>
<td>370 (0.6)</td>
<td>424 (0.7)</td>
<td></td>
</tr>
<tr>
<td>Percentage within health condition</td>
<td>96.2</td>
<td>1.8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Percentage within sexual identity</td>
<td>35.3</td>
<td>35.6</td>
<td>33.9</td>
<td></td>
</tr>
<tr>
<td>Good (n=18,235)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Inferential Data Analysis

Inferential data analysis (Table 3) was conducted to determine whether there were significant differences in mean ranks of depression score, drug abuse and dependence score, suicidal tendency score, and mental illness score across 3 sexual identity groups: heterosexual, lesbian or gay, and bisexual. The sample consisted of 59,864 individuals, with 96.2% (n=57,599) identifying as heterosexual, 1.7% (n=1018) identifying as lesbian or gay, and 2.1% (n=1257) identifying as bisexual.

The results of the inferential data analysis revealed statistically significant differences in the mean ranks of all 4 variables across the 3 sexual identity groups. Specifically, heterosexual individuals had the lowest mean ranks for all 4 variables, while bisexual individuals had the highest mean ranks for all 4 variables. Lesbian or gay individuals had intermediate mean ranks for all 4 variables. The chi-square test of independence was used to determine the statistical significance of the differences in mean ranks. The results of the chi-square tests were statistically significant for all 4 variables (depression score: \( \chi^2 = 458.241; P < .001 \); drug abuse and dependence score: \( \chi^2 = 226.946; P < .001 \); suicidal tendency score: \( \chi^2 = 67.795; P < .001 \); and mental illness score: \( \chi^2 = 363.722; P < .001 \)). The results suggest that there are significant differences in the distribution of depression score, drug abuse and dependence score, suicidal tendency score, and mental illness score across sexual identity groups.

These findings suggest that sexual identity is an important factor to consider when examining depression, drug abuse and dependence, suicidal tendency, and mental illness. The findings also underscore the importance of addressing the specific mental health needs of sexual minority populations.

Table 3. Mean ranks of depression score, drug abuse and dependence score, suicidal tendency score, and mental illness score by sexual identity.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean ranks</th>
<th>Depression score</th>
<th>Drug abuse and dependence score</th>
<th>Suicidal tendency score</th>
<th>Mental illness score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sexual identity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual (straight)</td>
<td>29,650.64</td>
<td>29,845.84</td>
<td>29,846.33</td>
<td>29,731.5</td>
<td></td>
</tr>
<tr>
<td>Lesbian or gay</td>
<td>35,164.76</td>
<td>31,287.05</td>
<td>30,820.18</td>
<td>33,393.78</td>
<td></td>
</tr>
<tr>
<td>Bisexual</td>
<td>38,549.02</td>
<td>32,792.46</td>
<td>33,156.47</td>
<td>36,300.86</td>
<td></td>
</tr>
<tr>
<td>Chi-square (df)</td>
<td>458.241 (2)(^a)</td>
<td>226.946 (2)(^b)</td>
<td>67.795 (2)(^c)</td>
<td>363.722 (2)(^c)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Significant result: \( P < .01 \).
\(^b\)Significant result: \( .001 \leq P < .01 \).
\(^c\)Significant result: \( .01 \leq P < .05 \).
Discussion

Principal Results

In recent years, there has been growing recognition of the unique mental health needs of LGB individuals. Studies have consistently shown that sexual minority populations are at increased risk for mental health problems including depression, anxiety, substance use disorders, and suicidal ideation. These disparities have been attributed to a range of factors, including discrimination, stigma, and marginalization, as well as social and economic disadvantage [24-26].

The results of this study support these findings, demonstrating that sexual minority individuals in the United States are at greater risk for experiencing social and economic disadvantage, as well as mental health problems, compared with heterosexual individuals [27]. The descriptive statistics analysis revealed that LGB individuals were more likely to report lower total family income and higher poverty levels compared with heterosexual individuals. These differences were statistically significant, indicating a robust association between sexual identity and socioeconomic status. Furthermore, the chi-square test of independence demonstrated that a higher percentage of LGB individuals reported lower total family income and higher poverty levels compared with heterosexual individuals. These findings suggest that sexual minority populations experience greater social and economic disadvantage compared with heterosexual individuals, which may contribute to their increased risk for mental health problems.

Moreover, the inferential data analysis demonstrated that sexual minority individuals are at increased risk for mental health problems compared with heterosexual individuals. Bisexual individuals, in particular, had the highest mean ranks for depression score, drug abuse and dependence score, suicidal tendency score, and mental illness score. These findings underscore the need for targeted interventions to address the mental health needs of sexual minority populations, particularly those who identify as bisexual.

The mental health disparities experienced by sexual minority populations have important implications for health care providers, policy makers, and other stakeholders. These disparities must be recognized and addressed through the development of targeted interventions aimed at reducing stigma and discrimination, increasing access to mental health services, and addressing the social and economic factors that contribute to poor mental health outcomes among sexual minority individuals [13,28-30]. Health care providers can play a critical role in addressing the mental health needs of sexual minority populations by creating safe and supportive environments that are free from discrimination and stigma [31-34]. This may involve providing culturally sensitive care, offering support groups or counseling services that are specifically tailored to the needs of sexual minority individuals, and increasing access to mental health services through community-based organizations and other resources.

Policy makers can also play an important role in addressing the mental health needs of sexual minority populations. This may involve passing laws and policies that protect the rights of sexual minority individuals such as antidiscrimination laws and laws that ensure access to health care and other resources [35,36]. Additionally, policy makers can work to increase funding for research and intervention programs aimed at reducing mental health disparities among sexual minority populations [37]. The results of this study highlight the importance of addressing the unique mental health needs of sexual minority populations. The findings suggest that sexual minority individuals are at increased risk for experiencing social and economic disadvantage, as well as mental health problems, compared with heterosexual individuals [38]. Health care providers, policy makers, and other stakeholders must work together to develop targeted interventions aimed at reducing these disparities and ensuring that all individuals have access to the support and care they need to lead healthy and fulfilling lives.

Future Implications

The findings of this study have important implications for future research and intervention efforts aimed at addressing the mental health needs of sexual minority populations. The following section outlines several potential future directions for research and intervention in this area.

First, future research should aim to better understand the underlying mechanisms driving the mental health disparities experienced by sexual minority populations. This may involve investigating the role of discrimination, stigma, and marginalization in contributing to poor mental health outcomes among sexual minority individuals. Additionally, future research could explore the impact of other social and environmental factors, such as social support and access to health care, on mental health outcomes among sexual minority populations.

Second, future intervention efforts should aim to address the specific mental health needs of sexual minority populations. This may involve developing culturally sensitive and tailored interventions that are designed to meet the unique needs of LGB individuals. For example, interventions could include support groups or counseling services that are specifically tailored to the needs of sexual minority individuals, as well as community-based programs that provide access to mental health services and other resources.

Third, future interventions should also focus on reducing stigma and discrimination against sexual minority populations. This may involve working with health care providers and policy makers to develop and implement policies and programs that protect the rights of sexual minority individuals, increase access to health care and other resources, and reduce discrimination and stigma in health care settings.

Finally, future research and intervention efforts should also focus on the mental health needs of bisexual individuals. The results of this study suggest that bisexual individuals may be at particularly high risk for poor mental health outcomes, compared with other sexual minority populations. Therefore, interventions aimed at addressing mental health disparities among sexual minority populations should include a specific focus on the needs of bisexual individuals.
The findings of this study underscore the need for targeted interventions and research aimed at addressing the mental health needs of sexual minority populations. By developing culturally sensitive and tailored interventions, reducing stigma and discrimination, and addressing the specific needs of bisexual individuals, we can work to reduce mental health disparities and ensure that all individuals have access to the support and care they need to lead healthy and fulfilling lives.

Limitations

The primary focus of this study revolves around examining the intricate relationships between various variables within the context of mental health, substance use disorders, and suicidal ideation in LGB adults. However, it is important to acknowledge that our investigation does not extend to the assessment of specific interventions targeting these issues. Consequently, although we have successfully discerned associations between the examined variables, our study is unable to furnish insights into the effectiveness of interventions aimed at ameliorating mental health concerns, substance use disorders, or suicidal ideation among LGB adults.

This limitation not only emphasizes the scope of our study but also underscores the necessity for future research endeavors to venture into the domain of interventions. To comprehensively address the challenges faced by the LGB community in terms of mental health and well-being, further investigations should explore and evaluate the potential impact and efficacy of intervention strategies. By delving into this unexplored territory, future studies can contribute to a more comprehensive understanding of effective approaches for reducing mental health disparities, substance use issues, and suicidal tendencies within the LGB population.

Conclusion

This study highlights the need for future research and interventions that address the mental health needs of sexual minority populations. Future research should investigate the underlying mechanisms driving mental health disparities and explore the impact of social and environmental factors on mental health outcomes. Interventions should be developed that address the specific needs of minority populations including tailored support groups and community-based programs. Efforts to reduce stigma and discrimination against sexual minority individuals are also crucial. Finally, interventions should focus on the mental health needs of bisexual individuals, who may be at particularly high risk for poor mental health outcomes. By addressing these issues, we can work toward reducing mental health disparities and ensuring that all individuals have access to the support and care they need.

Data Availability

The data sets generated or analyzed during this study are available in the 2019 National Survey on Drug Use and Health (NSDUH) repository [20].

Authors’ Contributions

ASWC, HLT, and FKCW were primarily responsible for writing this manuscript. JMCH, PMKT, GW, and LML provided expert advice and recommendations throughout the discussion process. EY supervised the overall development of the manuscript. All authors made significant contributions to the article and approved the submitted version. Additionally, all authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

LGB: lesbian, gay, and bisexual

NSDUH: National Survey on Drug Use and Health

©Alex Siu Wing Chan, Hon Lon Tam, Florence Kwai Ching Wong, Gordon Wong, Lok Man Leung, Jacqueline Mei Chi Ho, Patrick Ming Kuen Tang, Elsie Yan. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 25.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic
information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Original Paper

Improving the Efficiency of Inferences From Hybrid Samples for Effective Health Surveillance Surveys: Comprehensive Review of Quantitative Methods

Mansour Fahimi¹, PhD; Elizabeth C Hair², PhD; Elizabeth K Do², MPH, PhD; Jennifer M Kreslake², MPH, PhD; Xiaolu Yan³, MS; Elisa Chan³, MS; Frances M Barlas³, PhD; Abigail Giles³, MS; Larry Osborn³, MPH

¹Marketing Systems Group, Horsham, PA, United States
²Truth Initiative Schroeder Institute, Washington, DC, United States
³Ipsos USA, San Francisco, CA, United States

Corresponding Author:
Mansour Fahimi, PhD
Marketing Systems Group
755 Business Center Drive
Suite 200
Horsham, PA, 19044
United States
Phone: 1 2156202880
Email: mfahimi@m-s-g.com

Abstract

Background: Increasingly, survey researchers rely on hybrid samples to improve coverage and increase the number of respondents by combining independent samples. For instance, it is possible to combine 2 probability samples with one relying on telephone and another on mail. More commonly, however, researchers are now supplementing probability samples with those from online panels that are less costly. Setting aside ad hoc approaches that are void of rigor, traditionally, the method of composite estimation has been used to blend results from different sample surveys. This means individual point estimates from different surveys are pooled together, 1 estimate at a time. Given that for a typical study many estimates must be produced, this piecemeal approach is computationally burdensome and subject to the inferential limitations of the individual surveys that are used in this process.

Objective: In this paper, we will provide a comprehensive review of the traditional method of composite estimation. Subsequently, the method of composite weighting is introduced, which is significantly more efficient, both computationally and inferentially when pooling data from multiple surveys. With the growing interest in hybrid sampling alternatives, we hope to offer an accessible methodology for improving the efficiency of inferences from such sample surveys without sacrificing rigor.

Methods: Specifically, we will illustrate why the many ad hoc procedures for blending survey data from multiple surveys are void of scientific integrity and subject to misleading inferences. Moreover, we will demonstrate how the traditional approach of composite estimation fails to offer a pragmatic and scalable solution in practice. By relying on theoretical and empirical justifications, in contrast, we will show how our proposed methodology of composite weighting is both scientifically sound and inferentially and computationally superior to the old method of composite estimation.

Results: Using data from 3 large surveys that have relied on hybrid samples composed of probability-based and supplemental sample components from online panels, we illustrate that our proposed method of composite weighting is superior to the traditional method of composite estimation in 2 distinct ways. Computationally, it is vastly less demanding and hence more accessible for practitioners. Inferentially, it produces more efficient estimates with higher levels of external validity when pooling data from multiple surveys.

Conclusions: The new realities of the digital age have brought about a number of resilient challenges for survey researchers, which in turn have exposed some of the inefficiencies associated with the traditional methods this community has relied upon for decades. The resilience of such challenges suggests that piecemeal approaches that may have limited applicability or restricted accessibility will prove to be inadequate and transient. It is from this perspective that our proposed method of composite weighting has aimed to introduce a durable and accessible solution for hybrid sample surveys.

(JMIR Public Health Surveill 2024;10:e48186) doi:10.2196/48186

https://publichealth.jmir.org/2024/1/e48186
KEYWORDS
hybrid samples; composite estimation; optimal composition factor; unequal weighting effect; composite weighting; weighting; surveillance; sample survey; data collection; risk factor

Introduction
The survey sampling landscape is rapidly evolving. In an era of diminishing response rates and escalating costs, more effective survey sampling alternatives are no longer academic curiosities [1]. While the new realities suggest that departures from traditional methods are becoming inevitable, they also beckon an immediate question as survey researchers continue to experiment with hybrid sampling techniques. That is,

Are such sampling alternatives conducive to the inferential integrity of scientific surveys by reaching a representative subset of the target population in a pragmatic and cost-effective manner?

In addition to adopting multiple modes of data collection [2] it has become a customary practice to use less expensive samples selected from online panels to supplement costly alternatives from address or telephone frames [3]. The so-called opt-in panels are compiled using a potpourri of recruitment techniques, mostly relying on social media to fish for individuals willing to partake in surveys—hence the term river sampling [4]. The resulting convenience of these recruitment methods, however, is often achieved at the expense of compromising the organic representation that has been a natural byproduct of probability-based samples. This trade-off becomes of elevated concern since with samples obtained from opt-in panels, typical geodemographic weighting adjustments may no longer be adequate for ensuring their representativity [5].

It has been suggested that with such samples more granular weighting and calibration adjustments become necessary to ameliorate their compromised representations [6]. Specifically, calibration adjustments result in behavioral and attitudinal benchmarks that go beyond geographic corrections might be needed to improve the representation of survey respondents from less representative samples [7,8]. Moreover, when mixing samples secured from different sampling frames, special procedures must be used to combine the various sample components in an optimal fashion [9]. If conducted effectively, the resulting hybrid samples may address both cost and coverage challenges of traditional single-frame sample surveys—especially when surveying rare or hard-to-get cohorts.

While the literature on how to improve the external validity of survey estimates from nonprobability samples is maturing, existing studies have focused solely on surveys of adults [10-12]. Moreover, proposed methods are often theoretical in nature or certain to ad hoc techniques with limited scalability. Given the increasing inefficiencies of traditional survey sampling methodologies on the one hand, and the growing possibilities of emerging alternatives on the other, it is incumbent upon data scientists to explore innovative options that can address the evolving challenges facing the survey research community. This includes being able to produce useful inferences even when working with less-than-perfect data [13].

In this paper, we present practical weighting and calibration techniques that can be used to address the unique nuances of hybrid samples with varying representational properties, including those for the surveys of hard-to-get cohorts such as young adults. We will start with a review of the classical statistical technique known as composite estimation for combining survey estimates from 2 samples [14,15]. Next, as an extension of the methodology developed by Fahimi [16], we will introduce the method of composite weighting that is significantly more efficient, both computationally and inferentially, when pooling data from multiple surveys. For empirical illustrations, we will demonstrate results using data from 3 surveys with hybrid samples composed of probability-based components from the United States Postal Service address database and supplemental samples from online panels.

Methods

Ethical Considerations
All study procedures were conducted in accordance with the Declaration of Helsinki and its amendments. All study participants provided informed consent prior to being included in the study, which was approved by Advarra Institutional Review Board (IRB protocol numbers: Pro00010120 and Pro00009087).

Mathematics of Survey Data Integration
As mentioned earlier, there is a growing interest in hybrid methodologies that combine 2 or more independent samples with varying representational properties to reduce cost. This includes combining probability samples that could be based on random digit dialing or address-based sampling, as well as the many instances where probability and nonprobability samples are combined. Integrating survey data from independent samples provides a larger analytical database with enhanced inferential possibilities.

Data pooling is also relevant to regional surveys that are conducted independently of their national counterparts, but in which both surveys collect similar data from a common cohort. In these situations, one might be interested in combining data from a regional survey with those obtained from the corresponding subset of the national survey. For example, the National Health Interview Survey [17] and the Behavioral Risk Factor Surveillance System [18] both have national as well as local components that can be combined to produce more robust estimates at overlapping domains.

While the various algebraic building blocks of composite estimation methodology have been referenced in different textbooks and papers, to the best of our knowledge, the entire inferential machinery underpinning this cumbersome process has never been furnished in full detail in 1 place. As such, in the following section, we will provide a comprehensive
description of the mathematics of this classical methodology under varying scenarios.

**Composite Estimation Methodology**

Traditionally, the method of composite estimation has been used to blend results from different surveys to improve the robustness of the resulting estimates [19]. That is, individual point estimates from different surveys are produced and then blended together into 1 estimate at a time. In this section, we will furnish the mathematical foundation for this arduous approach before a more efficient alternative is introduced that can produce more stable estimates while significantly reducing the computational burden.

Consider a population of $N$ units from which 2 independent samples of size $n_1$ and $n_2$ have been selected. Under the conventional composition methodology, individual estimates from the 2 samples are produced separately and then combined to produce composite estimates that might be more robust. When the parameter of interest is, say population mean $\mu$, the general composite estimator will have the following form [20]:

$$\hat{\mu} = \alpha \hat{\mu}_1 + (1-\alpha) \hat{\mu}_2$$

(1)

In the above equation, $\hat{\mu}_1$ and $\hat{\mu}_2$ represent independent estimates of $\mu$ from the first and second samples, respectively. Using this decomposition, the optimal value for the blending or composition factor $\alpha$ can be obtained by minimizing the mean square error of $\hat{\mu}$, which is a function of the variance and bias of this composite estimate [21]:

$$\text{MSE}(\hat{\mu}) = \text{Var}(\hat{\mu}) + \text{Bias}(\hat{\mu})^2$$

(2)

Under this general scenario when neither of the 2 estimates can be considered unbiased, the optimal value of $\alpha$ can be obtained by [22]:

$$\alpha^* = \frac{n_1}{n_1 + n_2}$$

(3)

As mentioned earlier, a growing number of surveys complement their main probability samples with less expensive supplements secured from online panels from which the resulting estimates may not be unbiased. When probability and nonprobability samples are to be combined whereby only 1 of the 2 samples can provide unbiased estimates, say when $\hat{\mu}_1$ the optimal value of $\alpha$ becomes:

$$\alpha^* = 1$$

(4)

However, when the 2 independent estimates $\hat{\mu}_1$ and $\hat{\mu}_2$ could be assumed to have a negligible bias due to the application of survey weights, then the optimal value of $\alpha$ can be obtained by simply minimizing the variance of $\hat{\mu}$:

$$\text{Var}(\hat{\mu}) = \frac{1}{n_1} \text{Var}(\hat{\mu}_1) + \frac{1}{n_2} \text{Var}(\hat{\mu}_2)$$

(5)

The minimum of the above quadratic function of $\alpha$ is that value for which its derivative is equal to 0, that is:

$$\alpha^* = \frac{n_2}{n_1 + n_2}$$

(6)

Consequently, in this case, the optimal value of the composition factor can be obtained by [16]:

$$\alpha^* = \frac{n_2}{n_1 + n_2}$$

(7)

Furthermore, when estimates from the 2 surveys are expected to exhibit comparable variabilities as well, the above becomes a simpler function in terms of the sample sizes $n_1$ and $n_2$ and their associated unequal weighting effects (UWE) and $\hat{\mu}_2$. In this case, the optimal value of the composition factor can be obtained by [16]:

$$\alpha^* = \frac{n_2}{n_1 + n_2}$$

(8)

Finally, there are situations where it would be justifiable to assume that $\hat{\mu}_1$ and $\hat{\mu}_2$ ratio to near unity. In such instances, the optimal value of $\alpha$ reduces to a basic function of the respective sample sizes of the 2 surveys:

$$\alpha^* = \frac{n_1}{n_1 + n_2}$$

(9)

Whether any of the above simplifying assumptions can be justified or not, the fact remains that the composite estimation methodology entails operational complexities and inferential inefficiencies. First, this burdensome approach requires that composite estimates be produced 1 estimate at a time. Given that for a typical survey one must produce dozens of estimates for key outcome measures, this computationally intensive methodology requires serious time and resources.

Second, and more importantly, this piecemeal process produces estimates that are based on individual samples of size $n_1$ and $n_2$, and not the larger combined sample of size $n = n_1 + n_2$. This means relying on 2 estimates that could have been created using different methodologies with weighting adjustment granularities that will be coarser than what would be possible with a larger combined sample. In the next section, we will introduce an alternative methodology that can bypass these inefficiencies and complexities.

**Composite Weighting Methodology**

As described above, the classical composite estimation methodology is both cumbersome and inferentially inefficient. Our proposed methodology detailed here eliminates the above inefficiencies and complexities by allowing the 2 samples to be integrated first so that a single set of composite weights could be generated for estimation purposes. Specifically, instead of producing composite estimates from the unintegrated survey data 1 at a time, under this alternative, a single set of weights...
will be generated so that estimates could be produced from the combined sample. This huge convenience also benefits from all the inferential dividends the larger combined sample can offer. While the following derivation is for integrating data from 2 surveys, as demonstrated later, this scalable approach can easily apply when more than 2 surveys are involved.

For ease of illustration and without the loss of generality, we can assume there is only 1 weighting cell for poststratification purposes, and let

\[ B_{1i}: \text{Sampling base weights from sample 1, } i = 1, \ldots, n_1 \]
\[ B_{2j}: \text{Sampling base weights from sample 2, } j = 1, \ldots, n_2 \]

Based on the conventional composition method, separately poststratified weights for the 2 samples will have the following form:

\[ (10) \]

But rather than computing separate point estimates using the above 2 sets of poststratified weights and then compositing them 1 at a time, if the condition in equation 9 holds, one can create component weights that could be combined to aggregate to the same population total \( N \). The resulting weights can then be used to produce point estimates directly from the combined data without any need for piecemeal compositions. These weights could be produced by:

\[ (11) \]

Given that the above approach still requires separate poststratification of individual samples, it would be desirable if the 2 sets of base weights could be combined first and then poststratified jointly. This is a vastly superior option because it can accommodate more consistent and granular weighting adjustments, courtesy of the larger combined sample. Mathematically, this can be accomplished by a simple rescaling of individual base weights first and then combining them for a join poststratification or raking by:

\[ (12) \]

The above, however, magnifies the respondents’ base weights across the 2 samples with the same poststratification factor irrespective of any differential precision that would be associated with the sample that has better representation. The procedure described next introduces a simple calibration adjustment that could be used to remove this inequity prior to a joint poststratification.

**Calibration of Base Weights for Joint Poststratification**

In order for the alternative weighting procedure to produce final weights that are identical to the composite weights, the following must hold:

\[ (13) \]

Specifically, the above conditions would hold if the following were satisfied:

\[ (14) \]

This means that the alternative method can produce the same composite weights, provided that base weights from the 2 samples are calibrated prior to poststratification. That is when base weights from the 2 samples are first scaled to their respective sample sizes. Having done this, instead of separately poststratifying base weights from the 2 samples and then producing composite weights, one can use the proposed calibrated base weights from the 2 samples so that the 2 can be combined and poststratified concurrently.

It should be noted that the above calibration correction easily carries over to more realistic situations with more than 1 poststrata, where the underlying assumption in equation 9 is easier to satisfy. Also, one can apply the above procedure under the less restrictive condition in equation 8 when the UWE do not ratio to unity. In this more realistic situation, the corresponding base weights must be calibrated to their respective effective sample sizes as shown:

\[ (15) \]

Estimates of survey-specific UWE are often readily available, or they can be quickly approximated as a function of poststratified (or base) weights by the following formula when finite population corrections can be ignored [23].

\[ (16) \]

With the above correction applied, it would then be possible to use the resulting calibrated base weights as input for a final poststratification or raking of the combined sample using an expanded set of benchmarks. Alternatively, the same benchmarks could be used but with more granularity to improve the representation of the combined sample with respect to finer categories of the weighting variables.

Of note, the above approximation for UWE in equation 16 can also be used to assess the variability of estimates across surveys. Recognizing that variance is an estimate-specific statistic and mostly influenced by sample size and dispersion of the moments of the given estimate, UWE can be used as a relative measure of variability at the survey level as well. Typically, larger values of UWE are indicative of less representative sample surveys for which more variable weights have been needed to realign respondents vis-à-vis their population benchmarks.

**Extension and Generalization**

An extension of the above applies to instances where more than 2 surveys are to be integrated. In such situations, the existing weights for each survey can be calibrated by the following optimal composition factors to produce the final blended weights for combining 5 different samples:

\[ (17) \]
Results

To illustrate the applications of our proposed methodology, data from 3 pairs of surveys were used. Each pair was comprised of data from the Truth Longitudinal Cohort (TLC) survey and the Truth’s Continuous Tracker Online (CTO) survey. The TLC is a national representative panel of youths and young adults (aged 15-24 years) recruited via address-based sampling with online data collection. The CTO is a weekly, cross-sectional survey of participants aged 15-24 years. The CTO surveys are conducted using a sample of approximately 300 respondents per week from the national Dynata online panel. Participants of the TLC and CTO are invited to complete short online surveys about attitudes, beliefs, experimentation, and frequency of use regarding tobacco and other substances.

For the first pair, the CTO component included weekly surveys from July 20, 2020, to November 17, 2020. For the second pair, these weekly samples span from July 18, 2020, to February 24, 2021. For the third pair, the CTO component was comprised of weekly samples from July 14, 2021, to September 15, 2021. For each pair, the TLC component included surveys conducted during contemporaneous time periods. Table 1 provides a summary of the 3 pairs of TLC and CTO surveys used for this research.

For this investigation, we focused on 2 key outcome parameters, prevalence of current use (eg, at least 1 day of the past 30 days) of cigarettes and e-cigarettes. Prevalence estimates were produced separately from each sample component of each survey for respondents 15 to 17, 18 to 24, and 15 to 24 years of age after each component was weighted to basic geodemographic benchmarks of the given cohort.

Table 2 provides a summary of weighted point estimates for each survey and cohort, while Table 3 shows the corresponding overall estimates for 15- to 24-year-olds. All estimates are accompanied by their associated lower confidence limit and upper confidence limit at 95% CI, for which SEs were estimated using the method of Taylor Series Linearization in SAS [24].

To compare survey estimates produced under our proposed method of composite weighting against those using composite estimation, first the above point estimates had to be combined using the latter method. That is, point estimates from individual surveys were blended one by one using the coarse weighting methodology each sample survey could tolerate. Table 4 provides a summary of the resulting composite estimates by survey pair, cohort, and outcome measure.

Subsequently, composite weights were computed for each sample pair using our proposed methodology outlined above. These weights were computed for the combined TLC and CTO samples, where more granular weighting adjustments were possible due to larger sample sizes. Moreover, additional calibration adjustments were applied to the combined sample for which the needed benchmarks were generated from the TLC sample component for each survey pair. These adjustments included corrections with respect to the following behavioral attributes, which were shown to differentiate between young adult respondents from online panels and their cohorts. (1) Length of residence: about how long have you lived at your current address? (2) Household (dwelling) type: which of the following best describes your home? (3) Financial comfort: how would you describe your family’s overall financial situation? (4) Living with parents: do you currently live in a household with at least 1 of your parents? (5) Social media influencer: do you like to be a social media influencer?

While the rationale for the above calibration adjustments for general population surveys is detailed in [6,10], for the CTO surveys further investigations were carried out to identify differentiating attributes unique to teens and young adults. It is worth noting that for instances when a probability sample component like the TLC is not available in parallel, it is possible to use government sources, such as the monthly Current Population Survey or the American Community Survey, to secure relevant benchmarks for calibration adjustments.

It is of particular importance to note that the above calibration adjustments would not have been possible under the traditional composition methodology, whereby estimates are generated from individual surveys and then combined. Table 5 provides a summary of the resulting estimates produced by integrating the TLC and CTO sample components first and then weighting them to a more granular set of benchmarks and calibration adjustments listed above.

While independent estimates of tobacco use behaviors among teens and young adults can vary greatly due to methodological differences between their corresponding surveys, the following statistics from the National Youth Tobacco Survey [25] conducted by the Centers for Disease Control and Prevention were used to assess the external validity of estimates produced using our proposed methodology. Confounding differences among surveys could be due to study design and mode of administration, as well as other differences in periodicity and questionnaire wording. Cognizant of these differences, nonetheless, our estimates using composite weighting methodology produce comparable estimates to those from the National Youth Tobacco Survey 2020, as summarized in Table 6.
Table 1. Sample size summary for the TLC\textsuperscript{a} and CTO\textsuperscript{b} sample components by survey pair.

<table>
<thead>
<tr>
<th>Survey and cohort</th>
<th>Pair 1, n</th>
<th>Pair 2, n</th>
<th>Pair 3, n</th>
<th>Total, N</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLC</td>
<td>15-17</td>
<td>998</td>
<td>422</td>
<td>555</td>
</tr>
<tr>
<td></td>
<td>18-24</td>
<td>2363</td>
<td>1019</td>
<td>1229</td>
</tr>
<tr>
<td></td>
<td>15-24</td>
<td>3361</td>
<td>1441</td>
<td>1784</td>
</tr>
<tr>
<td>CTO</td>
<td>15-17</td>
<td>1041</td>
<td>1112</td>
<td>857</td>
</tr>
<tr>
<td></td>
<td>18-24</td>
<td>2747</td>
<td>2876</td>
<td>2113</td>
</tr>
<tr>
<td></td>
<td>15-24</td>
<td>3788</td>
<td>3988</td>
<td>2970</td>
</tr>
<tr>
<td>TLC+CTO</td>
<td>15-17</td>
<td>2039</td>
<td>1534</td>
<td>1412</td>
</tr>
<tr>
<td></td>
<td>18-24</td>
<td>5110</td>
<td>3895</td>
<td>3342</td>
</tr>
<tr>
<td></td>
<td>15-24</td>
<td>7149</td>
<td>5429</td>
<td>4754</td>
</tr>
</tbody>
</table>

\textsuperscript{a}TLC: Truth Longitudinal Cohort.
\textsuperscript{b}CTO: Continuous Tracker Online.

Table 2. Point estimates and confidence limits by outcome measures, survey pair, and cohort.

<table>
<thead>
<tr>
<th>Survey pair and outcome</th>
<th>15 to 17-year-old</th>
<th>18 to 24-year-old</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample, n (%)</td>
<td>Estimate (95% CI)</td>
</tr>
<tr>
<td><strong>Pair 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLC\textsuperscript{a}</td>
<td>998</td>
<td>2.2 (1.3-3.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.2 (6.2-10.2)</td>
</tr>
<tr>
<td>CTO\textsuperscript{b}</td>
<td>1041</td>
<td>5.5 (3.7-7.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.4 (7.9-12.9)</td>
</tr>
<tr>
<td><strong>Pair 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLC</td>
<td>422</td>
<td>1.2 (0.1-2.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.0 (4.5-9.5)</td>
</tr>
<tr>
<td>CTO</td>
<td>1112</td>
<td>6.1 (4.4-7.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.0 (11.4-16.6)</td>
</tr>
<tr>
<td><strong>Pair 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLC</td>
<td>555</td>
<td>3.0 (1.4-4.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.7 (3.6-7.8)</td>
</tr>
<tr>
<td>CTO</td>
<td>857</td>
<td>5.7 (3.9-7.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12.4 (9.5-15.4)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}TLC: Truth Longitudinal Cohort.
\textsuperscript{b}CTO: Continuous Tracker Online.
<table>
<thead>
<tr>
<th>Survey pair and outcome</th>
<th>Sample size, n</th>
<th>(%) estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pair 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TLC</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3361</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td></td>
<td>5.3 (4.4-6.2)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td></td>
<td>12.9 (11.5-14.2)</td>
</tr>
<tr>
<td><strong>CTO</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3788</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td></td>
<td>16.9 (15.3-18.5)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td></td>
<td>20.8 (18.9-22.6)</td>
</tr>
<tr>
<td><strong>Pair 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TLC</strong></td>
<td>1441</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td></td>
<td>5.0 (3.6-6.3)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td></td>
<td>14.2 (12.0-16.3)</td>
</tr>
<tr>
<td><strong>CTO</strong></td>
<td>3988</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td></td>
<td>17.2 (15.6-18.8)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td></td>
<td>23.2 (21.4-25.1)</td>
</tr>
<tr>
<td><strong>Pair 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TLC</strong></td>
<td>1784</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td></td>
<td>5.7 (4.4-7.0)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td></td>
<td>14.2 (12.3-16.2)</td>
</tr>
<tr>
<td><strong>CTO</strong></td>
<td>2970</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td></td>
<td>18.9 (16.9-20.9)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td></td>
<td>23.6 (21.4-25.8)</td>
</tr>
</tbody>
</table>

<sup>a</sup>TLC: Truth Longitudinal Cohort.

<sup>b</sup>CTO: Continuous Tracker Online.
<table>
<thead>
<tr>
<th>Survey pair, cohort, and outcome</th>
<th>Composite Estimate, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor (α)</td>
</tr>
<tr>
<td><strong>Pair 1</strong></td>
<td></td>
</tr>
<tr>
<td>15-17 (n=2039)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.80</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.61</td>
</tr>
<tr>
<td>18-24 (n=5110)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.75</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.64</td>
</tr>
<tr>
<td>15-24 (n=7149)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.76</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.64</td>
</tr>
<tr>
<td><strong>Pair 2</strong></td>
<td></td>
</tr>
<tr>
<td>15-17 (n=1534)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.74</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.52</td>
</tr>
<tr>
<td>18-24 (n=3895)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.56</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.41</td>
</tr>
<tr>
<td>15-24 (n=5429)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.58</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.43</td>
</tr>
<tr>
<td><strong>Pair 3</strong></td>
<td></td>
</tr>
<tr>
<td>15-17 (1412)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.56</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.66</td>
</tr>
<tr>
<td>18-24 (3342)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.55</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.00</td>
</tr>
<tr>
<td>15-24 (4754)</td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>.70</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>.57</td>
</tr>
</tbody>
</table>
Table 5. Survey estimates using composite weights by survey pair, cohort, and outcome measure.

<table>
<thead>
<tr>
<th>Survey pair, cohort, and smoker</th>
<th>(%)</th>
<th>estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pair 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-17 (n=2039)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>3.6</td>
<td>(2.7-4.6)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>9.1</td>
<td>(7.6-10.6)</td>
</tr>
<tr>
<td>18-24 (n=5110)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>13.6</td>
<td>(12.5-14.7)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>19.5</td>
<td>(18.1-20.9)</td>
</tr>
<tr>
<td>15-24 (n=7149)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>10.5</td>
<td>(9.6-11.3)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>16.3</td>
<td>(15.2-17.3)</td>
</tr>
<tr>
<td><strong>Pair 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-17 (n=1534)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>4.4</td>
<td>(3.2-5.6)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>11.5</td>
<td>(9.6-13.4)</td>
</tr>
<tr>
<td>18-24 (n=3895)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>17.0</td>
<td>(15.5-18.5)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>23.9</td>
<td>(22.1-25.7)</td>
</tr>
<tr>
<td>15-24 (n=5429)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>13.0</td>
<td>(11.9-14.1)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>20.0</td>
<td>(18.6-21.4)</td>
</tr>
<tr>
<td><strong>Pair 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-17 (n=1412)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>4.4</td>
<td>(3.2-5.6)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>9.2</td>
<td>(7.4-10.9)</td>
</tr>
<tr>
<td>18-24 (n=3342)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>15.9</td>
<td>(14.4-17.5)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>23.4</td>
<td>(21.5-25.2)</td>
</tr>
<tr>
<td>15-24 (n=4754)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarette</td>
<td>12.3</td>
<td>(11.2-13.5)</td>
</tr>
<tr>
<td>e-Cigarette</td>
<td>18.9</td>
<td>(17.5-20.3)</td>
</tr>
</tbody>
</table>

Table 6. Prevalence of cigarette and e-cigarette use among middle and high school students.

<table>
<thead>
<tr>
<th>School cohort</th>
<th>Past 30-day use</th>
<th>e-Cigarette, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cigarette, %</td>
<td></td>
</tr>
<tr>
<td>Middle school (n=7042)</td>
<td>1.2-2.2</td>
<td>3.6-6.0</td>
</tr>
<tr>
<td>High school (n=7453)</td>
<td>3.6-6.0</td>
<td>17.2-22.2</td>
</tr>
<tr>
<td>Middle and high school (n=14,531)</td>
<td>2.6-4.2</td>
<td>11.3-15.0</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

Conducting credible survey research in the 21st century is an endeavor subject to evolving challenges that require thinking outside of the traditional survey sampling toolbox. The proliferation of such challenges has had 2 distinct impacts on survey sampling. On the one hand, there are emerging improvisational methods of sampling and weighting that, while expedient, are void of scientific underpinnings. On the other, there is a growing disenchantment with traditional methods that
Despite their complex and computationally intensive nature, struggle with the new realities of the digital age. Our proposed method of composite weighting addresses both deficiencies, by offering an accessible approach that is soundly grounded in inferential sciences.

The statistical machinery, survey researchers have relied upon for decades, made it possible to make measurable inferences about target populations when samples of modest size are selected from complete sampling frames; all sampling units carry known and nonzero selection probabilities; and surveys achieve near-perfect rates of response [26]. For various reasons, but most notably the growing rates of nonresponse and survey costs, many of the surveys conducted these days struggle to fulfill the fundamental tenets of this sampling paradigm [27-29]. While such violations have been commonplace and routinely discounted by market researchers for whom theoretical considerations are often trumped by cost and time constraints, arguably, even large-scale government surveys are no longer exempt from such challenges [30].

A strategy that is gaining popularity for dealing with the rising costs and coverage challenges of surveys is to combine 2 or more independent samples selected from separate sampling frames with varying representations of the target population. In particular, such hybrid alternatives can pay considerable dividends when survey data secured from certain sample components are significantly less costly. There are also instances when multiple samples are required for the design of a study, while in other situations existing data from different surveys are pooled to address the size and analytical needs of a given research.

In comparison to the traditional method of composite estimation whereby separate estimates are combined from different surveys 1 at a time, our proposed composite weighting methodology for integrating survey data offers at least 5 distinct advantages.

First, the method of composite weighting is less cumbersome than that of composite estimation because it enables researchers to work with a single data file and not multiple sets of data and weights from unintegrated surveys.

Second, an integrated database that is larger than any of the individual sample components accommodates more nuanced weighting adjustments than what might be possible with individual surveys. This becomes especially appealing when one of the surveys is based on a small sample size, whereby coarse weighting can fail to improve the representation of its respondents.

Third, integrated survey data allows more in-depth analyses, particularly when comparisons of smaller analytical subgroups are of interest. Such deep-dive multivariate analyses are not feasible when producing separate estimates from individual surveys, some of which could be of modest size.

Fourth, related to the above, survey estimates from the resulting integrated data will be subject to smaller and consistently calculated SEs courtesy of the larger sample size and a single data set to make inferences from. Composite weighting eliminates extraneous variabilities that are inevitable under composite estimation due to the application of inconsistent weighting procedures for individual surveys, such as the use of different benchmarks, raking algorithms, and weight-trimming rules.

Finally, and perhaps most importantly, composite weighting offers a unique advantage that is of particular importance as interest in combining data from probability and nonprobability samples continues to grow. Specifically, this methodology offers the possibility of additional calibration adjustments of respondents from nonprobability samples to benchmarks that are not externally available but can be generated from the “more representative” subset of respondents.

The above advantages are particularly relevant to public health research initiatives, as they often require faster data collection turnarounds from large pools of respondents. Unlike opinion polling and commercial survey applications that typically aim for “good-enough” estimates of trends, health studies often require reliable assessments that can guide public policies. It is in this context that unreliable inferences, or those that are reliable but slow to produce, can have dire consequences or be obsolete.

Yet just like any new methodology, what we have proposed is not a panacea or free from limitations. For instance, neither this approach nor any other can produce measurable inferences from data secured from sample surveys that fail to represent their target cohort due to systematic exclusions. Another potential limitation worth mentioning is the technical sophistication the application of this methodology may require. While substantially less complicated and resource-intensive as compared to the method of composite estimation, our proposed method still requires some level of inferential acumen and know-how. Then again, working with complex surveys that entail mixing data from multiple sources is an undertaking that requires a decent level of survey sampling and inferential familiarity to begin with. If tasked with the challenge of mixing data from multiple surveys, our hope is that researchers will find the option of composite weighting more accessible and efficient.

Concluding Remarks

Despite the growing challenges facing the survey research community, practitioners should not succumb to suboptimal practices for cost-saving purposes alone. Such unilateral guidelines have contributed to the stigma that commercial surveys have inadequate concern for rigor. On the other hand, undue allegiance to traditional methods of survey sampling can also confine researchers to inefficient practices that are losing their pragmatism. This position becomes particularly untenable when such adherences are simply for the sake of preserving the optics.

Traditional methods of survey research are becoming inefficient, both with respect to data quality and cost, begging for novel and pragmatic alternatives that do not forego rigor. It is from this perspective that we hope the methodology we have furnished in this paper could address some of such inefficiencies, enabling survey researchers to take fuller advantage of the data resources they have at their disposal. Results from the detailed comparisons we have exhibited in this paper show that composite weighting methodology is vastly less cumbersome

https://publichealth.jmir.org/2024/1/e48186
and produces estimates that are at least as reliable as what can be produced using composite estimation.

As a final parting note we would like to reiterate that with declining response rates, surveys require progressively more comprehensive weighting adjustments to restore the representation of their respondents. As such, it is not advisable to shy away from more aggressive weighting and calibration adjustments only to keep the resulting UWE at bay [31]. Of course, this is fully cognizant of the proverbial seesaw occupied on one side by bias and variance on the other. Above all, it is imperative to retain full transparency about adopted methodologies and their potential shortfalls as we explore new possibilities for survey sampling in the digital age.

Acknowledgments
This research would not have been possible without the invaluable contributions of an impressive cadre of researchers from the Truth Initiative: ECH, EKD, and JMK, as well as Donna Vallone of Schroeder Institute. Special thanks are offered to XY and EC of Ipsos USA for their impeccable statistical support, as well as AG and LO for their project management. A data sharing agreement is required for the use of all data used in this research. Approved investigators may access data sets via an analytic portal owned and administered by Truth Initiative. Finally, we acknowledge that generative artificial intelligence has not been used in any shape or form in the creation of this paper.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to their proprietary nature and privacy concerns; however, they can be made available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

References


Abbreviations

CTO: Continuous Tracker Online
TLC: Truth Longitudinal Cohort
UWE: unequal weighting effect

©Mansour Fahimi, Elizabeth C Hair, Elizabeth K Do, Jennifer M Kreslake, Xiaolu Yan, Elisa Chan, Frances M Barlas, Abigail Giles, Larry Osborn. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 07.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Projected Time for the Elimination of Cervical Cancer Under Various Intervention Scenarios: Age-Period-Cohort Macrosimulation Study

Yi-Chu Chen¹, PhD; Yun-Yuan Chen², PhD; Shih-Yung Su³, PhD; Jing-Rong Jhuang⁴, PhD; Chun-Ju Chiang¹,⁵, PhD; Ya-Wen Yang⁵, MSc; Li-Ju Lin⁶, PhD; Chao-Chun Wu⁶, MD; Wen-Chung Lee¹,⁵,⁷, MD, PhD

¹Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan
²Head Office, Taiwan Blood Services Foundation, Taipei, Taiwan
³Master Program in Statistics, National Taiwan University, Taipei, Taiwan
⁴Institute of Statistical Science, Academia Sinica, Taipei, Taiwan
⁵Taiwan Cancer Registry, Taipei city, Taiwan
⁶Health Promotion Administration, Ministry of Health and Welfare, Taipei, Taiwan
⁷Institute of Health Data Analytics, College of Public Health, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Wen-Chung Lee, MD, PhD
Institute of Health Data Analytics
College of Public Health
National Taiwan University
Room 536, No 17, Xuzhou Road
Taipei, 100
Taiwan
Phone: 886 223511955
Fax: 886 223511955
Email: wenchung@ntu.edu.tw

Abstract

Background: The World Health Organization aims for the global elimination of cervical cancer, necessitating modeling studies to forecast long-term outcomes.

Objective: This paper introduces a macrosimulation framework using age-period-cohort modeling and population attributable fractions to predict the timeline for eliminating cervical cancer in Taiwan.

Methods: Data for cervical cancer cases from 1997 to 2016 were obtained from the Taiwan Cancer Registry. Future incidence rates under the current approach and various intervention strategies, such as scaled-up screening (cytology based or human papillomavirus [HPV] based) and HPV vaccination, were projected.

Results: Our projections indicate that Taiwan could eliminate cervical cancer by 2050 with either 70% compliance in cytology-based or HPV-based screening or 90% HPV vaccination coverage. The years projected for elimination are 2047 and 2035 for cytology-based and HPV-based screening, respectively; 2050 for vaccination alone; and 2038 and 2033 for combined screening and vaccination approaches.

Conclusions: The age-period-cohort macrosimulation framework offers a valuable policy analysis tool for cervical cancer control. Our findings can inform strategies in other high-incidence countries, serving as a benchmark for global efforts to eliminate the disease.

(JMIR Public Health Surveill 2024;10:e46360)  doi:10.2196/46360

KEYWORDS

age-period-cohort model; population attributable fraction; macrosimulation; cancer screening; human papillomavirus; HPV; cervical cancer; intervention; women; cervical screening; public health intervention
Introduction

Cervical cancer remains the fourth most common cancer in women worldwide [1]. The World Health Organization (WHO) [2] called for the global elimination of cervical cancer; elimination is defined as an incidence of fewer than 4 cases per 100,000 women-years. Eliminating cervical cancer requires scaling up the human papillomavirus (HPV) vaccination of girls and cervical screening [3].

Randomized trials and follow-up studies are valuable for evaluating the short-term and medium- to long-term impacts of public health intervention policies [4,5]. In contrast, modeling studies offer the ability to estimate and forecast effects over an extended timeframe and can also simulate outcomes across a range of future scenarios [6]. Numerous microsimulation modeling studies have been conducted to assess the necessary time frame and intervention strategies for the elimination of cervical cancer [3,7-14]. These studies rely on numerous assumptions, such as oncogenic potentials, infection dynamics and immunity of HPV, and the natural history of cervical cancer. They require many parameters that are difficult to obtain.

A better alternative is to use macrosimulation modeling. From a macroscopic perspective, disease incidence rates are primarily influenced by 3 temporal factors: age, period, and cohort. The age-period-cohort (APC) model, which accounts for these factors, has been recently used to project future disease burdens [15-19]. This methodology can also estimate future incidence rates of cervical cancer under existing conditions, known as the status quo. In a hypothetical “what if” scenario involving a public health intervention, the baseline incidence rate would be reduced by a certain fraction—captured by the population attributable fraction (PAF) associated with the intervention. By combining the APC model’s baseline estimates with the PAF to factor in reductions from interventions, we can effectively forecast future cervical cancer incidence rates under specific public health programs.

Well-organized cytology-based cervical screening programs have effectively reduced the incidence rates of cervical cancer in many countries [20-23]. Since 1995, Taiwan’s Health Promotion Administration has offered organized cytology-based screenings for women aged 30 years and older. Consequently, the age-standardized incidence rate of invasive cervical cancer in Taiwan sharply declined from 28.0 per 100,000 woman-years in 1997 to 8.2 in 2016. Despite this progress, the incidence rate in Taiwan has yet to meet the criteria for the elimination of cervical cancer. A more effective form of HPV-based screening, which has shown to be superior to cytology-based methods [24], may soon be adopted for mass screening. Projecting future trends in cervical cancer incidence under new approaches such as HPV-based screening has been challenging using microsimulation models. However, these projections may be more feasible through macrosimulation techniques.

In this study, macrosimulation was used, combining APC modeling with PAF calculation, to estimate when cervical cancer could be eliminated in Taiwan. We explored the potential to expedite this timeline by implementing more extensive population-based cervical cancer screening (either cytology or HPV based) and enhancing HPV vaccination coverage.

Methods

Data Source

In this study, we collected data on cervical cancer cases diagnosed between 1997 and 2016 using the International Classification of Diseases for Oncology, Third Edition code C53. Our extensive data set, which includes population information, originated from the Taiwan Cancer Registry—a meticulously maintained nationwide system established by the Ministry of Health and Welfare. The registry diligently captures and synthesizes information from patients who are newly diagnosed with malignant cancer in hospitals with 50 or more beds in Taiwan, a country with a population size of approximately 23 million. The data quality and completeness of the Taiwan Cancer Registry database have consistently adhered to standards of excellence. Specifically, the completeness is 98.4% (14,833/218,239); the percentage of cases with death certificate only is 0.9% (1079/118,583); the mortality versus incidence ratio is 45.1% (202.89/449.59); the percentage of morphological verification is 93% (109,273/117,504) for all sites combined and 97.6% (103,812/106,423) for all sites excluding the liver; and the data timeliness is 14 months. These statistics demonstrate that the Taiwan Cancer Registry is one of the highest-quality cancer registries in the world [25-27]. The world standard population (WHO 2000 [28]) proportions were used to calculate age-standardized incidence rates. The projected elimination year was defined as the first year when the age-standardized incidence rate would fall below 4 cases per 100,000 women-years [8].

Macrosimulation

The APC models were used to project future cervical cancer incidence rates under the status quo. Due to the scarcity of cases in patients younger than 30 years of age, this age group was excluded from the APC modeling. Instead, the average incidence rate between 1997 and 2016 was used as the projected rate for women younger than 30 years of age. This approach likely makes our future decline estimates slightly conservative. However, the impact is minimal as the incidence rate for this age group contributes only a small fraction to the overall rate. For women aged 30 years and older, the projections are outlined below. First, an ensemble of 265 APC models was constructed using data from 1997 to 2006 as the training set. The APC models were then applied to project incidence rates from 2007 to 2016 as the validation set. These projections underwent year-on-year attenuation adjustments ranging from 0% and 5% to 100%, resulting in 5565 different projection sets. The symmetric mean absolute percentage error (SMAPE) was used to quantify the prediction error for each projection model, and the model with the lowest SMAPE was selected. Finally, the selected model, fine-tuned with all available data from 1997 to 2016, was used to project future incidence rates up to the year 2050. These projections incorporated the chosen attenuation factor. The future projected incidence rates were determined by calculating age-standardized rates across all age groups.
those younger than 30 years of age, the average incidence rate from 1997 to 2016 was used. For individuals aged 30 years and older, the rates forecasted by the selected APC model were used.

Next, we calculated the PAFs for various scenarios of cervical cancer screening and HPV vaccination. The PAF in this study was defined as the proportionate reduction in the incidence rate of cervical cancer due to a specific intervention program, as in the following equation [29]:

\[
\text{Incidence rate under a specific intervention} = \text{incidence rate under the status quo} \times (1 – \text{PAF under a specific intervention})
\]

where IRR\(_i\) is the incidence rate ratio between the \(i\)th level of the screening or vaccination variable and the reference level \((i=1)\) of no screening and no vaccination, and \(P_i\) and \(P_j\) are the proportions of women in the \(i\)th level of the screening or vaccination variable under the status quo and the specific interventions, respectively.

The incidence rate ratios of cytology-based screening, HPV-based screening, and HPV vaccination were based on the studies of Chen et al [30], Ronco et al [24], and Lei et al [5]. We assume that HPV vaccination administered to 13-year-old girls affords them lifetime effectiveness against cervical cancer [31]. HPV-based screening may have a higher false positive rate, so it is suggested that screening be performed once every 5–10 years [24]. Currently, only cytology-based screening is available in Taiwan, and HPV-based screening still needs to be implemented. We also assume that the incidence rate ratios of a joint program involving both interventions are the products of 2 programs involving only the respective intervention (Table S1 in Multimedia Appendix 1). In 2016, the proportion of women with cytology-based screening more than twice in 6 years was 43.5% (data provided by Taiwan’s Health Promotion Administration), and a negligible proportion of women had received HPV vaccination in Taiwan (Table S2 in Multimedia Appendix 1).

We assume a gradual increase from 2023 to 2030 in the proportion of women undergoing cytology-based screening more than twice within 6 years from the current 43.5% to 70%. We also assume a new intervention plan from 2023 that involves switching from cytology-based screening twice within 6 years to HPV-based screening. The goal is to increase the overall screening proportion from 43.5% to 70% by 2030. We set an intervention scenario to achieve HPV vaccination coverage of 90% since 2018 (the Health Promotion Administration in Taiwan has offered HPV vaccination for 13-year-old girls since 2018). Furthermore, the potential impact of combining cytology-based and HPV-based screening with HPV vaccination in 2 different scenarios was evaluated. For simplicity, we assume that compliance with cytology-based screening, HPV-based screening, and receipt of the HPV vaccination are 3 independent events. The referral rate for positive cervical cytology results, facilitated by community nurses, exceeds 90%. Additionally, Taiwan’s health care system offers almost universal coverage, making effective cervical disease management an integral part of the existing health care infrastructure. As such, this study did not specifically factor in the concept of effective management.

We used equation 1 to calculate the PAFs for all 6 intervention scenarios considered in this study. The incidence rate under a specific intervention was then calculated using the following equation:

\[
\text{Incidence rate under a specific intervention} = \text{incidence rate under the status quo} \times (1 – \text{PAF under a specific intervention})
\]

Data management and analyses were performed using SAS statistical software (version 9.4; SAS Institute Inc).

Ethical Considerations

The study was based solely on deidentified aggregate data, without access to individual records. This study protocol was approved by the National Taiwan University Research Ethics Committee (NTU-REC 202101HM030) and the data release review board of the Health Promotion Administration, Ministry of Health and Welfare in Taiwan. All methods were performed in accordance with the relevant guidelines and regulations. In addition, the National Taiwan University Research Ethics Committee waived the requirement for informed consent due to the lack of personal information and secondary data in the study.

Results

The selected projection model under the status quo was a polynomial APC model with a log link function and 55% attenuation (SMAPE=6.1%). Figure S1 in Multimedia Appendix 1 presents the observed and model-fitted age-standardized (WHO 2000 standard population [28]) cervical cancer incidence rates from 1997 to 2016 and the projections from 2017 to 2050. A declining trend in cervical cancer incidence rate was observed in the 20-year study period. In 2016, under the status quo of cytology-based screening compliance of 43.5% and no HPV vaccination, the projected cervical cancer incidence rate will not fall below 4 new cases per 100,000 women-years by 2050. Figure 1 shows the expected incidence rates of cervical cancer from 2023 to 2030 under 3 scenarios: the current situation, if adherence to cytology-based screening increases to 70%, and if adherence to HPV-based screening rises to 70%. The projection indicates that cervical cancer in Taiwan will not reach the goal of elimination by 2050 if cytology-based screening compliance remains at the current level of 43.5%. However, if cytology-based and HPV-based screening compliance are increased to 70%, cervical cancer elimination can be achieved by 2047 and 2035, respectively.

Figure 2 shows the projected cervical cancer incidence rate with an HPV vaccination coverage of 90%. An HPV vaccination coverage of 90% will eliminate cervical cancer by 2050. Figure 3 presents the projected cervical cancer incidence rates if cytology-based or HPV-based screening is applied (compliance raised to 70%) in conjunction with HPV vaccination (90% coverage). Both joint interventions will help to achieve cervical cancer elimination before 2050 and at an earlier year (2038 and 2033, respectively).

The years of elimination (if before 2050) for the various scenarios are shown in Table 1. For comparison, the same table
also shows the results when the Segi standard population [32] was used for age standardization. The time to elimination was expedited by a few years when the Segi standard was used.

**Figure 1.** Future projections of age-standardized cervical cancer incidence rate per 100,000 person-years under the status quo, cytology-based, and HPV-based screening. HPV: human papillomavirus.

**Figure 2.** Future projections of age-standardized cervical cancer incidence rate per 100,000 person-years under HPV vaccination. HPV: human papillomavirus.
Figure 3. Future projections of age-standardized cervical cancer incidence rate per 100,000 person-years under scenarios of joint cytology-based and HPV-based screening with HPV vaccination. HPV: human papillomavirus.

Table 1. Projected elimination years\(^a\) under the status quo and various intervention scenarios, when the rates are age-standardized to World Health Organization 2000’s standard population and the Segi standard population.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Projected elimination year (Segi standard population)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status quo (cytology-based screening compliance 43.5% and no HPV(^b) vaccination)</td>
<td>—(^c)</td>
</tr>
<tr>
<td>Scenario 1 (cytology-based screening compliance 70% and no HPV vaccination)</td>
<td>2047 (2035)</td>
</tr>
<tr>
<td>Scenario 2 (HPV-based screening compliance 70% and no HPV vaccination)</td>
<td>2035 (2030)</td>
</tr>
<tr>
<td>Scenario 3 (cytology-based screening compliance 43.5% and HPV vaccine coverage 90%)</td>
<td>2050 (2046)</td>
</tr>
<tr>
<td>Scenario 4 (cytology-based screening compliance 70% and HPV vaccine coverage 90%)</td>
<td>2038 (2033)</td>
</tr>
<tr>
<td>Scenario 5 (HPV-based screening compliance 70% and HPV vaccine coverage 90%)</td>
<td>2033 (2030)</td>
</tr>
</tbody>
</table>

\(^a\)The first year when the projected age-standardized cervical cancer incidence rate falls below 4 cases per 100,000 women-years.

\(^b\)HPV: human papillomavirus.

\(^c\)The projected age-standardized cervical cancer incidence rate would not fall below 4 cases per 1000,000 women-years before 2050.

Discussion

Principal Findings

In summary, our macrosimulation analysis projected that Taiwan could eliminate cervical cancer by 2050 through 70% compliance with screening (cytology based or HPV based) or 90% coverage of HPV vaccination. Specifically, the projected elimination years are 2047 or 2035 for screening (cytology based or HPV based, respectively), 2050 for vaccination, and 2038 or 2033 for a combination of both screening and vaccination. This study confirms earlier microsimulation findings that increased screening can fast-track cervical cancer elimination [3,7-14]. Our macrosimulation methodology offers both adaptability and ease of implementation, as demonstrated by the SAS code in Multimedia Appendix 2. Unique to Taiwan is its high HPV vaccine coverage, a legacy of successful vaccine campaigns, facilitated by extensive health care infrastructure [33]. In stark contrast, Japan saw HPV vaccine coverage collapse from 70% to nearly 0% between 2013 and 2019 due to a crisis [14].

Cytology-based screening is crucial for cervical cancer management [34]. It is recommended that women aged 30 years and older should undergo cytology-based screening at least once every 3 years (effective screening). Historically, cervical cancer incidence has steadily decreased because of opportunistic and organized cytology-based screening [35]. From 1974 to 1984, Taiwan launched an opportunistic cytology-based screening program for cervical cancer, facilitated through partnerships between the cancer society and gynecology and obstetrics clinics [36]. A major turning point came in 1995 when Taiwan’s Health Promotion Administration set up an organized cytology-based screening initiative targeting women aged 30 years and older. This represented a significant advancement in Taiwan’s efforts to tackle cervical cancer. Consequently, the age-standardized incidence rate for invasive cervical cancer in Taiwan dropped from 28.0 per 100,000 woman-years in 1997 to 8.2 per 100,000 woman-years in 2016. This reduction marked a shift from Taiwan being a high-risk region to becoming a low-to-medium-risk area for the disease (the figure was presented in Figure S2 in Multimedia Appendix 1).
However, the need for improved compliance with effective screening is a global issue, and Taiwan is no exception [37]. Although approximately 82% of Taiwanese women have undergone at least one screening since 1995, the overall effectiveness of screening implementation remains below the desired level. The participation rate for triennial screenings, aimed at women aged 30 to 69 years, is only around 54%. Moreover, participation drops even further among women aged 69 years and older. These data underscore the pressing need for improved outreach and accessibility to achieve more comprehensive and impactful screening initiatives [38]. Various strategies have been adopted to promote screening, including educational interventions, physician reminders, incentive programs, mass media campaigns, outreach to community members, and leveraging community health workers [39-41]. Despite all these efforts, the compliance rate of effective screening in 2016 was only 43.5% in Taiwan. Our analysis indicates that cervical cancer in Taiwan can be eliminated in 2047 only if there is 70% compliance with cytology-based screening. To enhance the effectiveness of screening, strategies could include distributing informative pamphlets to schoolchildren to share with adult women in their families; authorizing self-collected HPV screening kits (Taiwan currently only approves the use of clinical-based HPV screening kits); and targeting older, previously unscreened and unvaccinated groups.

Cytology-based screening as a primary mode of cervical cancer screening has been gradually replaced by HPV-based screening, which has 70% greater protection against invasive cervical cancer [24]. HPV testing recommended by WHO may increase engagement in cervical cancer screening programs [42]. Taiwan has been implementing cytology-based screening for more than 30 years; however, compliance still needs to be improved. The Health Promotion Administration in Taiwan is also considering implementing more effective HPV-based screening. This study demonstrated that HPV-based screening could achieve the goal of eliminating cervical cancer more swiftly than cytology-based screening, given the same conditions. Taiwan has successfully decreased the incidence rate of cervical cancer through an organized cytology-based screening program. It is now considering the implementation of HPV-based screening to achieve the goal of elimination faster. This approach can be used as a model for other countries.

If HPV vaccine efficacy wanes, it may alter cervical cancer prevention and screening protocols [43]. However, long-term studies since the vaccine’s 2006 introduction show sustained high antibody levels, suggesting it could offer near-lifelong cervical cancer protection [44-50]. However, multiple factors hinder the implementation and scaling-up of HPV vaccination, such as vaccine supply shortage [51,52], budgetary constraints [53], and hesitancy due to vaccine-related side effects [54]. Globally, only a few high-income countries have offered the HPV vaccine for the target age group with a coverage rate above the WHO-recommended threshold of 90% [51,55]. By comparison, the HPV vaccination program introduced in Taiwan in 2018 has been relatively successful; the coverage rates were 76.8% in 2018 and 86.9% in 2019, respectively. It is very likely to further increase this rate to 90% in 2022 (scenario 3 in this study); if this happens, we project that cervical cancer in Taiwan can be eliminated in 2050. The effect of HPV vaccination (on 13-year-old girls) can only appear after the vaccinated cohort reaches the high-risk ages (45 years and older) for cervical cancer—the cohort effect. This is why it takes a longer time to achieve the goal of cervical cancer elimination with an HPV vaccination coverage of 90% (scenario 3 in this study) compared to cytology- or HPV-based screening with a compliance of 70% (scenarios 1 and 2 in this study), reaching the goal in 2050 versus 2047 or 2035.

WHO announced the 90-70-90 target to achieve cervical cancer elimination: HPV vaccination coverage rate of 90% of girls by the age of 15 years, twice-lifetime screening of 70% of vaccinated women (by the age of 35 and 45 years), and treatment of 90% of women with the cervical disease [23]. The respective latest figures in Taiwan were 90% (HPV vaccination coverage rate; the actual figure could be higher because some women were vaccinated at their own expense and were not tallied), 43% (effective screening compliance rate), and 90% (proportion of treated for women who screened positive), with the effective screening compliance rate lagging far behind the 90-70-90 target. Our analysis shows that if the effective screening compliance rate can be increased to 70% in conjunction with an HPV coverage rate of 90%, Taiwan can achieve the goal of cervical cancer elimination in 2038 (scenario 4 in this study). However, it is worth noting that although the incidence rates of cervical cancer among all age groups have declined in Taiwan, there is a slightly increased trend in the 30-34 years age group among the most recent birth cohort (Figure S3 in Multimedia Appendix 1). This warrants attention from health authorities.

**Strengths and Limitations**

The APC macrosimulation framework developed in this study is a useful policy analysis tool for disease control. The policy analysis results in this study can serve as a reference for other countries with a high incidence of cervical cancer. The strengths of this study lie in the rigorous analytical models we used and the high-quality data we used. Nonetheless, the study is not without limitations. First, we do not have access to individual-level data, and the study is therefore prone to ecological fallacy. Second, we used data spanning from 1997 to 2016 as the basis for making future predictions. This data range is unaffected by the delayed diagnosis and registration of cancer cases that occurred due to the COVID-19 pandemic [56]. Consequently, our short- to medium-term projections for cervical cancer incidence may be biased by the COVID-19 pandemic. Nonetheless, this is unlikely to affect the study’s primary conclusions, which focus on long-term projections. Finally, public funding for HPV vaccination in Taiwan is restricted to 13-year-old girls, leaving female individuals in other age groups the choice to opt for private vaccination. This study concentrates mainly on the consequences of the publicly funded HPV vaccination program and does not incorporate the likely effects of privately funded vaccinations in a wider age range. The high rate of HPV vaccination also fosters benefits through herd immunity [57], which could mean that our evaluations are underestimating the full potential impact of the vaccination program.
Acknowledgments
The content of this research may not represent the opinion of the Health Promotion Administration, Ministry of Health and Welfare in Taiwan. The authors used Grammarly and ChatGPT (OpenAI) [58] to correct the grammar in the paper. This work was supported by grants from the Health Promotion Administration, the Ministry of Health and Welfare in Taiwan (A1111010; Tobacco Health and Welfare Taxation) and the National Science and Technology Council in Taiwan (MOST 111-2314-B-002-089-MY3). The funders had no role in study design, data collection, and analysis; the decision to publish; or the preparation of the paper.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Additional data regarding incidence rate ratios, proportions of women, and cervical cancer incidence rates.
[PDF File (Adobe PDF File), 567 KB - publichealth_v10i1e46360_app1.pdf ]

Multimedia Appendix 2
SAS code.
[PDF File (Adobe PDF File), 140 KB - publichealth_v10i1e46360_app2.pdf ]

References


43. Adopted 2024-02-17


Abbreviations

APC: age-period-cohort
HPV: human papillomavirus
PAF: population attributable fraction
SMAPE: symmetric mean absolute percentage error
WHO: World Health Organization

©Yi-Chu Chen, Yun-Yuan Chen, Shih-Yung Su, Jing-Rong Jhuang, Chun-Ju Chiang, Ya-Wen Yang, Li-Ju Lin, Chao-Chun Wu, Wen-Chung Lee. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 18.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Predicting Lung Cancer Survival to the Future: Population-Based Cancer Survival Modeling Study

Fan-Tsui Meng¹,²*, MSc; Jing-Rong Jhuang³*, PhD; Yan-Teng Peng², MSc; Chun-Ju Chiang¹,⁴, PhD; Ya-Wen Yang¹,⁴, MSc; Chi-Yen Huang⁵, PhD; Kuo-Ping Huang⁵, MSc; Wen-Chung Lee¹,⁴,⁶, MD, PhD

¹Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan
²Parexel International Company Limited, Taipei, Taiwan
³Institute of Statistical Science, Academia Sinica, Taipei, Taiwan
⁴Taiwan Cancer Registry, Taipei, Taiwan
⁵Health Promotion Administration, Ministry of Health and Welfare, Taipei, Taiwan
⁶Institute of Health Data Analytics and Statistics, College of Public Health, National Taiwan University, Taipei, Taiwan
* these authors contributed equally

Corresponding Author:
Wen-Chung Lee, MD, PhD
Institute of Health Data Analytics and Statistics
College of Public Health
National Taiwan University
Room 536, No 17, Xuzhou Road
Taipei, 100
Taiwan
Phone: 886 233668036
Fax: 886 223511955
Email: wenchung@ntu.edu.tw

Abstract

Background: Lung cancer remains the leading cause of cancer-related mortality globally, with late diagnoses often resulting in poor prognosis. In response, the Lung Ambition Alliance aims to double the 5-year survival rate by 2025.

Objective: Using the Taiwan Cancer Registry, this study uses the survivorship-period-cohort model to assess the feasibility of achieving this goal by predicting future survival rates of patients with lung cancer in Taiwan.

Methods: This retrospective study analyzed data from 205,104 patients with lung cancer registered between 1997 and 2018. Survival rates were calculated using the survivorship-period-cohort model, focusing on 1-year interval survival rates and extrapolating to predict 5-year outcomes for diagnoses up to 2020, as viewed from 2025. Model validation involved comparing predicted rates with actual data using symmetric mean absolute percentage error.

Results: The study identified notable improvements in survival rates beginning in 2004, with the predicted 5-year survival rate for 2020 reaching 38.7%, marking a considerable increase from the most recent available data of 23.8% for patients diagnosed in 2013. Subgroup analysis revealed varied survival improvements across different demographics and histological types. Predictions based on current trends indicate that achieving the Lung Ambition Alliance’s goal could be within reach.

Conclusions: The analysis demonstrates notable improvements in lung cancer survival rates in Taiwan, driven by the adoption of low-dose computed tomography screening, alongside advances in diagnostic technologies and treatment strategies. While the ambitious target set by the Lung Ambition Alliance appears achievable, ongoing advancements in medical technology and health policies will be crucial. The study underscores the potential impact of continued enhancements in lung cancer management and the importance of strategic health interventions to further improve survival outcomes.

(JMIR Public Health Surveill 2024;10:e46737) doi:10.2196/46737

KEYWORDS
lung cancer; survival; survivorship-period-cohort model; prediction; prognosis; early diagnosis; lung cancer screening; survival trend; population-based; population health; public health; surveillance; low-dose computed tomography
Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide. Approximately 1.8 million people worldwide died of lung cancer in 2020, accounting for 18% of all cancer deaths [1]. The disease typically lacks noticeable symptoms in its early stages, leading to late diagnoses in many patients. By the time symptoms such as coughing, shortness of breath, weakness, and chest pain manifest, the prognosis is often poor [2]. Lung cancer survival rates have improved slowly compared with other types of cancer [2-4]. The 5-year survival rate for lung cancer was approximately 10%-20% and is among the lowest of all types of cancer [5,6]. In Taiwan, lung cancer has been the leading cause of cancer death since 2010, with a 5-year survival rate of 25% [7].

Enhanced screening and early diagnosis are critical for improving prognosis and reducing mortality rates associated with lung cancer. The adoption of various diagnostic techniques, including blood tests and computed tomography scans, has significantly improved the accuracy and precision of lung cancer diagnoses [8-11]. Moreover, the advancement of noninvasive diagnostic methods has expedited the time required to reach a diagnosis [12-16]. In Taiwan, the implementation of low-dose computed tomography (LDCT) for lung cancer screening has been particularly effective. Data indicate a notable improvement in lung cancer survival rates across all patient groups from 2010 to 2016 [17].

In July 2019, the Lung Ambition Alliance, an international coalition dedicated to lung cancer control, announced its goal to double the 5-year survival rate for lung cancer by 2025 [18]. For this study, we used data from the Taiwan Cancer Registry, a national population-based registry known for its high-quality data [19,20]. We applied the survivorship-period-cohort (SPC) model, a recently developed method for nowcasting and forecasting cancer survival rates [21], to predict the future survival of patients with lung cancer in Taiwan. This analysis aims to assess the feasibility of achieving the coalition’s ambitious goal.

Methods

Case Definitions

Lung cancer data were obtained from the Taiwan Cancer Registry. Lung cancer cases were defined per the International Classification of Disease for Oncology, Field Trial Edition (ICD-O-FT) code 162 and International Classification of Disease for Oncology, Third Edition (ICD-O-3) codes C33 and C34. This study included patients diagnosed with lung cancer between July 1, 1997, and June 30, 2018. We linked patients to death registration data and tracked the date of death to December 31, 2018.

The year of diagnosis was defined as the middle of that year to the middle of the following year. For a given year of diagnosis, the number of survivors in the first year after diagnosis was the number of patients who did not die by the end of the following year. The first-year survival rate was calculated by dividing the number of patients who survived the first year after diagnosis by the number of patients in that year. Patients who survived the first year of diagnosis were followed up for another year, and the number of patients who did not die was the number of survivors in the second year after diagnosis. The second-year survival rate was calculated by dividing the number of survivors in the second year after diagnosis by the number of survivors in the first year. We calculated the survival rates for the third, fourth, and fifth years after diagnosis by using the same approach.

Definitions of Survival Rates

The unit of analysis in this study was the 1-year interval survival rate, defined as the probability that a patient with cancer currently alive will survive for at least 1 more year. The 1-year interval survival rates for patients with lung cancer based on the year after diagnosis (survivorship), calendar year (period), and year of diagnosis (cohort) are presented in Figure 1 (a total of 95 shaded cells with observed values). The 5-year survival rate, calculated using the cohort approach, is the product of the five 1-year interval (first to fifth years) survival rates for patients diagnosed in the same year. We extrapolated the 1-year interval survival rate to the 2020 year of diagnosis (25 unshaded cells in Figure 1). In the cohort approach, the 5-year survival rate for calendar year 2025 is the 5-year survival rate for patients diagnosed in 2020.
**Figure 1.** The unit of analysis for this study. The values in the cells are the 1-year interval survival rates (%) of lung cancer in Taiwan (shaded cells: observed values and unshaded cells: projected values from the survivorship-period-cohort model).

**SPC Model**
This study used the SPC model proposed by Peng et al [21] to predict survival rates. The independent variables were the year after diagnosis (survivorship), calendar year (period), and year of diagnosis (cohort). These 3 variables are perfectly collinear (cohort + survivorship = period), resulting in a nonidentifiability problem. The nonidentifiability problem leads to infinite possibilities for parameter estimation. However, this does not prevent the prediction of the 1-year interval survival rate because all possible parameter estimates yield the same fitted value. The overall effects of period and cohort were divided into “slopes” and “curvatures” for extrapolation. We projected that the future period and cohort slopes would be identical to those observed in the empirical data and used the restricted cubic spline method to fit the period and cohort curvatures of the empirical data to extrapolate into the future. A more detailed description of the SPC model can be found in Peng et al [21].

**Training and Validation**
We used the 1-year interval survival rates from 1998 to 2013 as the training data set (70 lightly shaded cells in Figure 1) and the 1-year interval survival rates from 2014 to 2018 as the validation data set (25 darkly shaded cells in Figure 1) to calculate the symmetric mean absolute percentage error for the SPC model. We used all observed data (95 shaded cells in Figure 1) to predict the 1-year interval survival rates. We multiplied the 5 predicted 1-year interval (first to fifth year after diagnosis) survival rates (patients diagnosed between 2018 and 2020) or the partially observed and partially predicted 1-year interval survival rates (patients diagnosed between 2014 and 2017) for patients diagnosed in the same year to obtain the expected 5-year survival rate.

**Uncertainty in Prediction**
We used the bootstrap method to calculate CIs for the survival rates. We performed 10,000 bootstrap resampling and used the interval between the 2.5th and 97.5th percentiles as the 95% CI for the survival rates.

**Data Analysis Software Used**
All analyses were performed using SAS (version 9.4; SAS Institute Inc) and R (version 4.3.1; R Foundation for Statistical Computing).

**Ethical Considerations**
The study used deidentified aggregate data exclusively without access to individual records. Ethical approval was granted by the National Taiwan University Research Ethics Committee (NTU-REC 202101HM030) and the Data Release Review Board of the Health Promotion Administration, Ministry of Health and Welfare in Taiwan. All procedures were conducted in compliance with applicable guidelines and regulations. Additionally, the requirement for informed consent was waived by the National Taiwan University Research Ethics Committee due to the absence of personal information and the use of secondary data in the study.

**Results**

**Characteristics of the Study Population**
This study analyzed data from 205,104 patients with lung cancer, of whom 128,928 were men and 76,176 were women (Table 1). In total, 8236 (56.7%) patients were older than 65 years of age. The number of newly diagnosed cases significantly increased from 5762 in 1997 to 14,529 in 2017. Among all histological types, lung adenocarcinoma was the most prevalent in Taiwan, with its proportion rising sharply from 2202 (38.2%) in 1997 to 10,223 (70.4%) in 2017. Lung squamous cell carcinoma was the second most common type, with its proportion decreasing from 1448 (25.1%) in 1997 to 1856 (12.8%) in 2017. Regarding stage distribution, most patients with lung cancer in Taiwan were diagnosed at stages III and IV; however, this proportion gradually decreased from 5933 (83.8%) in 1997 to 8340 (66.8%) in 2017. Conversely, the proportion of stage I diagnoses increased from 927 (13.1%) in 1997 to 3659 (29.3%) in 2017.
Trends in Survival Rates

**Table 1.** Patients with lung cancer diagnosed in Taiwan from 1997 to 2017, grouped by sex, age, histological type, and stage.

<table>
<thead>
<tr>
<th>Year</th>
<th>Sex, n (%)</th>
<th>Age (years), n (%)</th>
<th>Histological type, n (%)</th>
<th>Stage, n (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>&lt;55</td>
<td>55-64</td>
<td>65-74</td>
</tr>
<tr>
<td>1997</td>
<td>4041 (70.1)</td>
<td>1721 (29.9)</td>
<td>850</td>
<td>1186</td>
<td>2248</td>
</tr>
<tr>
<td>1998</td>
<td>4425 (63.3)</td>
<td>1963 (30.7)</td>
<td>926</td>
<td>1253</td>
<td>2473</td>
</tr>
<tr>
<td>1999</td>
<td>4681 (68.9)</td>
<td>2114 (31.1)</td>
<td>1015</td>
<td>1249</td>
<td>2599</td>
</tr>
<tr>
<td>2000</td>
<td>4634 (69.7)</td>
<td>2013 (30.3)</td>
<td>972</td>
<td>1197</td>
<td>2509</td>
</tr>
<tr>
<td>2001</td>
<td>4808 (68)</td>
<td>2265 (32)</td>
<td>1124</td>
<td>1292</td>
<td>2444</td>
</tr>
<tr>
<td>2002</td>
<td>4788 (67.4)</td>
<td>2316 (32.6)</td>
<td>1134</td>
<td>1260</td>
<td>2427</td>
</tr>
<tr>
<td>2003</td>
<td>5468 (66.8)</td>
<td>2720 (33.2)</td>
<td>1288</td>
<td>1421</td>
<td>2677</td>
</tr>
<tr>
<td>2004</td>
<td>5746 (67.8)</td>
<td>2727 (32.2)</td>
<td>1315</td>
<td>1470</td>
<td>2619</td>
</tr>
<tr>
<td>2005</td>
<td>5627 (65.6)</td>
<td>2954 (34.4)</td>
<td>1453</td>
<td>1491</td>
<td>2545</td>
</tr>
<tr>
<td>2006</td>
<td>6064 (65.8)</td>
<td>3150 (34.2)</td>
<td>1468</td>
<td>1621</td>
<td>2644</td>
</tr>
<tr>
<td>2007</td>
<td>6249 (64.6)</td>
<td>3424 (35.4)</td>
<td>1627</td>
<td>1749</td>
<td>2796</td>
</tr>
<tr>
<td>2008</td>
<td>6625 (64.6)</td>
<td>3633 (35.4)</td>
<td>1600</td>
<td>1948</td>
<td>2873</td>
</tr>
<tr>
<td>2009</td>
<td>6555 (62.7)</td>
<td>3902 (37.3)</td>
<td>1638</td>
<td>2177</td>
<td>2789</td>
</tr>
<tr>
<td>2010</td>
<td>6824 (63.2)</td>
<td>3965 (36.8)</td>
<td>1767</td>
<td>2273</td>
<td>2787</td>
</tr>
<tr>
<td>2011</td>
<td>7071 (61.4)</td>
<td>4451 (38.6)</td>
<td>1908</td>
<td>2570</td>
<td>2901</td>
</tr>
<tr>
<td>2012</td>
<td>6971 (60.8)</td>
<td>4499 (39.2)</td>
<td>1871</td>
<td>2761</td>
<td>2908</td>
</tr>
<tr>
<td>2013</td>
<td>7312 (59.7)</td>
<td>4944 (40.3)</td>
<td>2082</td>
<td>3012</td>
<td>3113</td>
</tr>
<tr>
<td>2014</td>
<td>7380 (58.5)</td>
<td>5230 (41.5)</td>
<td>2067</td>
<td>3192</td>
<td>3192</td>
</tr>
<tr>
<td>2015</td>
<td>7715 (57.7)</td>
<td>5660 (42.3)</td>
<td>2181</td>
<td>3432</td>
<td>3473</td>
</tr>
<tr>
<td>2016</td>
<td>7889 (57.4)</td>
<td>5903 (42.8)</td>
<td>2250</td>
<td>3534</td>
<td>3724</td>
</tr>
<tr>
<td>2017</td>
<td>8008 (55.1)</td>
<td>6521 (44.9)</td>
<td>2451</td>
<td>3842</td>
<td>3950</td>
</tr>
</tbody>
</table>

<sup>a</sup>ADENO: adenocarcinoma.
<sup>b</sup>SCC: squamous cell carcinoma.
<sup>c</sup>SCLC: small cell lung cancer.
<sup>d</sup>Not available.

**Trends in Survival Rates**

Figure 2 displays the 1-year interval survival rates from the first to the fifth year after lung cancer diagnosis in Taiwan. The 1-year interval survival rates from the first to the fifth year had the same turning point at the diagnosis year of 2004: falling before this year and rising subsequently. The fastest increase...
in the 1-year interval survival rate is the first year after diagnosis. In the year of diagnosis 2020, the 1-year interval survival rates are projected to reach 69.4%, 78.7%, 85.2%, 89.4%, and 93.1% in the first, second, third, fourth, and fifth years after diagnosis, respectively.

Figure 3 displays the 5-year survival rate of lung cancer in Taiwan. The most recent 5-year survival rate that could be computed using the cohort approach was the 5-year survival rate for patients diagnosed in 2013 (23.8%). We used the SPC model (symmetric mean absolute percentage error=0.72%) to predict a 5-year survival rate of 38.7% for the year of diagnosis 2020. The absolute increase in the 5-year survival rate of lung cancer in Taiwan from the year of diagnosis 2013 to 2020 was 38.7% – 23.8% = 14.9%, and the fold increase was 38.7%/23.8% = 1.6-fold.

Figure 2. The 1-year interval survival rates (%) of lung cancer in Taiwan (dots: observed values; solid lines: fitted values from the survivorship-period-cohort model; dotted lines: projected values from the survivorship-period-cohort model; and color-shaded regions: 95% bootstrap CI).
Figure 3. The 5-year survival rates (%) of lung cancer in Taiwan (dots: observed values; solid line: product of 5 observed 1-year interval survival rates from the first to fifth years after diagnosis for patients in the same diagnosis year, 1997-2013; dashed line: product of 5 partially observed and partially predicted 1-year interval survival rates from the first to fifth years after diagnosis for patients in the same diagnosis year, 2014-2017; dotted line: product of 5 predicted 1-year survival rates from the first to fifth years after diagnosis for patients in the same diagnosis year, 2018 to 2020; and shaded region: 95% bootstrap CI).

**Trends in Relative Survival Rates**

This study also assessed the relative survival rate of patients with lung cancer (relative to the survival rate of the age-, sex-, and calendar year–matched population in Taiwan) [21,22]. Both the relative survival rates (Figures S1 and S2 in Multimedia Appendix 1) and the absolute survival rates (Figures 2 and 3) exhibited similar increasing trends. The absolute increase in the 5-year relative survival rate of lung cancer in Taiwan from year of diagnosis 2013 to 2020 was 44.6% – 27.9% = 16.7%, and the fold increase was 44.9%/27.9% = 1.6-fold.

**Subgroup Analysis**

In addition, we predicted the future survival rates of patients with lung cancer on the basis of sex, age, and histological type (Figures S3-S15 in Multimedia Appendix 1). The absolute increases (fold increases) in the 5-year survival rates of patients with lung cancer are as follows (Table S1 in Multimedia Appendix 1): men 9.8% (1.6-fold), women 19% (1.6-fold), younger than 55 years 23.4% (1.6-fold), 55 to 64 years 16.8% (1.5-fold), 65 to 74 years 15.5% (1.6-fold), older than 75 years 4.9% (1.5-fold), lung adenocarcinoma 18.6% (1.6-fold), lung squamous cell carcinoma 2.4% (1.2-fold), small-cell lung cancer 1% (1.2-fold), and other histological types of lung cancer 2.6% (1.2-fold).

**Weighted Average Predictions Across Patient Strata**

We also made predictions by calculating the weighted average of stratum-specific future 5-year survival rates based on the number of patients in each stratum in the year of diagnosis 2017 (Table 1). The results (Figure 4) closely matched the future 5-year survival rates predicted for all patients with lung cancer (Figure 3).
Figure 4. The 5-year survival rates (%) of lung cancer in Taiwan, weighted averaged by sex, age, and histological type of the number of patients diagnosed in 2017 (solid line: product of 5 observed 1-year interval survival rates from the first to fifth years after diagnosis for patients in the same diagnosis year, 1997-2013; dashed line: product of 5 partially observed and partially predicted 1-year interval survival rates from the first to fifth years after diagnosis for patients in the same diagnosis year, 2014-2017; dotted line: product of 5 predicted 1-year survival rates from the first to fifth years after diagnosis for patients in the same diagnosis year, 2018-2020; and color-shaded regions: 95% bootstrap CI).

Discussion

Principal Findings

This population-based cancer survival study demonstrates that the 5-year survival rate of patients with lung cancer in Taiwan decreased before 2004 but proceeded to increase afterward, corresponding to the trend of lung cancer mortality in Taiwan [23,24] and the trend of lung cancer survival in one hospital in Taiwan [24]. Patients with advanced lung cancer usually receive chemotherapy. Taiwan’s National Health Insurance has been reimbursing several targeted drugs for lung cancer treatment: gefitinib (since 2004), erlotinib (since 2007), afatinib (since 2014), and crizotinib (since 2015) [25]. These targeted drugs are mainly used in patients with lung adenocarcinoma, and their treatment effects are more favorable than chemotherapy [26-28]. Figure S15 and Table S1 in Multimedia Appendix 1 also reveal more favorable survival rates for patients with lung adenocarcinoma compared to other histological types. Taiwan’s National Health Insurance has also been reimbursing positron emission tomography for staging tumors since 2004 [29]. Previous studies have revealed that it is more accurate for staging tumors than computed tomography, enabling patients to receive appropriate treatment [30-32]. This may also be one reason the 5-year survival rate for lung cancer in Taiwan has increased since 2004. In addition, the Tobacco Hazards Prevention Act was implemented in 1997, and the smoking rate in Taiwan has dropped considerably. The incidences of small-cell lung cancer and lung squamous cell carcinoma, closely related to smoking, have also decreased yearly [33,34]. The reduced proportion of the poorer-survival histological types may contribute to the overall improvement in lung cancer survival.

Lung cancer screening can also help improve the survival rate of patients with lung cancer. The National Lung Screening Trial in the United States first demonstrated in 2011 that LDCT could effectively reduce lung cancer mortality [35]. The Lung Cancer Society in Taiwan has also conducted an LDCT screening study on nonsmoking lung cancer for high-risk populations [36]. LDCT can detect more early-stage patients and reduce lung cancer mortality compared to traditional chest radiography [35-38]. The proportion of patients with early-stage lung cancer diagnosed each year in Taiwan has been increasing since 2007,
and the rate of increase has accelerated since 2011 (Figure S16 in Multimedia Appendix 1). Lung cancer survival rates have rapidly increased since 2007 when stages I to IV were combined (see solid line in Figures S17 and S18 in Multimedia Appendix 1). In addition, the survival rate in each stage exhibits an upward trend (Figures S19-S23 in Multimedia Appendix 1). We believe that positron emission tomography during this period allowed for more accurate staging [30-32] and improved survival rates at various stages. Targeted therapy may also contribute to the increased survival rate of patients with advanced disease [39-42].

We simulated a lung cancer survival trend in the event that screening did not take place (see dot-dashed line in Figures S17 and S18 in Multimedia Appendix 1). The trend was calculated as the yearly weighted averages of the stage-specific lung cancer survival rates, assuming the proportion of patients diagnosed in each stage remained constant at the 2007 level (Figure S16 in Multimedia Appendix 1). The difference between the factual trend (see solid line in Figures S17 and S18 in Multimedia Appendix 1) and the simulated counterfactual trend (see dot-dashed line in Figures S17 and S18 in Multimedia Appendix 1) represents the benefit of lung cancer screening, although this may also reflect the effect of lead time bias. The Health Promotion Administration of Taiwan started implementing free LDCT screening for lung cancer for the high-risk population [43] in July 2022 (before that, LDCT screening was paid for by the public at their own expense). As a result, we speculate that the 5-year survival rate of patients with lung cancer will continue to increase in the future.

The novelty and strength of this study lie in its scope—a nationwide, population-based cancer survival analysis involving more than 200,000 patients with lung cancer. Furthermore, it uses the state-of-the-art SPC model for macrosimulation of future trends. To the best of our knowledge, this is the first study of its kind. However, this study has several limitations. First, the forecasts do not account for the impacts of health policy implementations or medical advancements post-2018, which may result in an underestimation of the future 5-year survival rates for lung cancer in Taiwan. Second, as this is a population-based cancer survival modeling analysis, the absence of individual-level data hampers our ability to make causal inferences and precise personal predictions. Additionally, the factors contributing to the decline in lung cancer survival rates in Taiwan between 1997 and 2004 remain unclear.

Conclusions

This study conducted a comprehensive analysis of lung cancer survival trends in Taiwan, using data from over 200,000 patients and using the SPC model to project future survival rates. Our findings indicate notable improvements in the 5-year survival rates, particularly noticeable from 2004 onward, attributed to advances in medical treatments and diagnostic methods, including the implementation of LDCT screening. The results suggest that the goal set by the Lung Ambition Alliance to double the 5-year survival rate by 2025 is ambitious yet potentially achievable within Taiwan. However, limitations such as the exclusion of post-2018 medical advancements and policy changes could affect the accuracy of our projections. This study highlights the critical role of continuous medical innovation and effective health policy implementation in achieving significant survival improvements for patients with lung cancer.

Acknowledgments

The content of this research may not represent the opinion of the Health Promotion Administration, Ministry of Health and Welfare in Taiwan. This study was based on the master thesis of FTM (the first author of this paper) completed at National Taiwan University under the supervision of WCL (the corresponding author of this paper). The authors used Grammarly (Grammarly, Inc) and ChatGPT (OpenAI) to correct the grammar in the paper. This work is supported by grants from the Health Promotion Administration, the Ministry of Health and Welfare in Taiwan (A1121004; Tobacco Health and Welfare Taxation), and the National Science and Technology Council in Taiwan (MOST 111-2314-B-002-089-MY3).

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

FTM authored the initial manuscript, which JRJ and WCL subsequently revised and edited. FTM, JRJ, and YTP developed the statistical method and executed all statistical analyses. FTM, CJC, YWY, CYH, KPH, and WCL managed data curation and project administration. WCL also conceptualized and supervised the study and secured funding. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Stratification of overall and relative survival by sex, age, histology, and stage.

[PDF File (Adobe PDF File), 2988 KB - publichealth_v10i1e46737_app1.pdf]
References


Abbreviations

ICD-O-3: International Classification of Disease for Oncology, Third Edition
ICD-O-FT: International Classification of Disease for Oncology, Field Trial Edition
LDCT: low-dose computed tomography
SPC: survivorship-period-cohort

Edited by AS Bhagavathula; submitted 23.02.23; peer-reviewed by D Yang, V Binson; comments to author 20.01.24; revised version received 24.04.24; accepted 24.04.24; published 31.05.24.

Please cite as:
Meng FT, Jhuang JR, Peng YT, Chiang CJ, Yang YW, Huang CY, Huang KP, Lee WC
Predicting Lung Cancer Survival to the Future: Population-Based Cancer Survival Modeling Study
JMIR Public Health Surveill 2024;10:e46737
URL: https://publichealth.jmir.org/2024/1/e46737
doi: 10.2196/46737
PMID: 38819904

©Fan-Tsui Meng, Jing-Rong Jhuang, Yan-Teng Peng, Chun-Ju Chiang, Ya-Wen Yang, Chi-Yen Huang, Kuo-Ping Huang, Wen-Chung Lee. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 31.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Early Results of an Initiative to Assess Exposure to Firearm Violence in Ambulatory Care: Descriptive Analysis of Electronic Health Record Data

Nicole Cook¹, MPA, PhD; Megan Hoopes¹, MPH; Frances M Biel¹, MPH, MS; Natalie Cartwright², PhD; Michelle Gordon¹, MPH; Marion Sills¹, MPH, MD

¹OCHIN, Portland, OR, United States
²Department of Mathematics, Norwich University, Northfield, VT, United States

Corresponding Author:
Nicole Cook, MPA, PhD
OCHIN
PO Box 5426
Portland, OR, 97228
United States
Phone: 1 503 943 2500
Fax: 1 503 943 2501
Email: cookn@ochin.org

Abstract

Background: Current research on firearm violence is largely limited to patients who received care in emergency departments or inpatient acute care settings or who died. This is because standardized disease classification codes for firearm injury only represent bodily trauma. As a result, research on pathways and health impacts of firearm violence is largely limited to people who experienced acute bodily trauma and does not include the estimated millions of individuals who were exposed to firearm violence but did not sustain acute injury. Assessing and collecting data on exposure to firearm violence in ambulatory care settings can expand research and more fully frame the public health issue.

Objective: The aim of the study is to evaluate the demographic and clinical characteristics of patients who self-reported exposure to firearm violence during a behavioral health visit.

Methods: This study assessed early data from an initiative implemented in 2022 across a national network of ambulatory behavioral health centers to support trauma-informed care by integrating structured data fields on trauma exposure into an electronic health record behavioral health patient assessment form (SmartForm), as such variables are generally not included in standard outpatient medical records. We calculated descriptive statistics on clinic characteristics, patient demographics, and select clinical conditions among clinics that chose to implement the SmartForm and among patients who reported an exposure to firearm violence. Data on patient counts are limited to positive reports of exposure to firearm violence, and the representativeness of firearm exposure among all patients could not be calculated due to unknown variability in the implementation of the SmartForm.

Results: There were 323 of 629 (51%) clinics that implemented the SmartForm and reported at least 1 patient exposed to firearm violence. In the first 11 months of implementation, 3165 patients reported a recent or past exposure to firearm violence across the 323 clinics. Among patients reporting exposure, 52.7% (n=1669) were male, 38.8% (n=1229) were Black, 45.7% (n=1445) had posttraumatic stress disorder, 37.5% (n=1186) had a substance abuse disorder (other than nicotine), and 11.7% (n=371) had hypertension.

Conclusions: Current research on firearm violence using standardized data is limited to acute care settings and death data. Early results from an initiative across a large network of behavioral health clinics demonstrate that a high number of clinics chose to implement the SmartForm, resulting in thousands of patients reporting exposure to firearm violence. This study demonstrates that collecting standardized data on firearm violence exposure in ambulatory care settings is feasible. This study further demonstrates that resultant data from ambulatory settings can be used for meaningful analysis in describing populations affected by firearm violence. The results of this study hold promise for further collection of structured data on exposure to firearm violence in ambulatory settings.

(JMIR Public Health Surveill 2024;10:e47444) doi:10.2196/47444

https://publichealth.jmir.org/2024/1/e47444

JMIR Public Health Surveill 2024 | vol. 10 | e47444 | p.2072
(page number not for citation purposes)
KEYWORDS
gun violence; firearm injury; surveillance; primary care; public health; ambulatory care; electronic health record; violence; burden; emergency department; data; risk factor

Introduction

The escalation in firearm injury in the United States is reflected in epidemiologic trends, including the emergence of firearm violence as the leading cause of mortality in children ages 1-18 years since 2020—yet the public health burden of firearm injury is severely undercounted [1-3]. Current incidence and prevalence of firearm injury are largely limited to acute injury data from hospitals and emergency departments and death data. This is because clinical classification codes (International Classification of Diseases) for firearm injury currently only represent acute bodily trauma [4-9]. While millions of children and adults are estimated to have experienced primary or secondary exposure to firearm violence (direct witness of firearm violence or the acute aftermath), there is no standardized data collection system to support surveillance of this public health problem [10-12].

As a result, pathways, risk factors, and intervention strategies following exposure to firearm violence are poorly understood [13]. Collecting broader data on exposure to firearm violence and the physical and psychological injury following exposure is critically needed to identify risk factors and disease pathways, frame the full impact of firearm violence on society, and inform the development of comprehensive treatments for people impacted by firearm violence [12]. Ambulatory electronic health record (EHR) data are potentially rich sources of injury and exposure data that can broaden our understanding of the spectrum of health impacts of firearm violence.

In 2022, a large national network of community-based behavioral health safety–net clinics serving racial and ethnic minority groups, low-income and underserved rural communities, sexual and gender minorities, and other health disparate populations implemented a customized data collection tool (SmartForm) in their shared EHR to collect trauma history as part of the behavioral health assessment process for clients. All network clinics are located in medically underserved areas and serve minoritized populations most likely to experience disparate community firearm violence [14,15]. The SmartForm implementation included an associated workflow to collect standardized data on recent or past exposure to firearm violence among patients presenting for behavioral health care [16]. Implementation of the trauma SmartForm is part of a larger quality improvement initiative to optimize EHR solutions to support trauma-informed care across the network [17]. To our knowledge, this initiative is the first large-scale initiative to collect standardized firearm violence exposure data in outpatient care settings. In this paper, we present early results on exposure to firearm violence from the first 11 months of implementation of the SmartForm.

Methods

Study Design

We used EHR data from the OCHIN multistate network of community-based primary and behavioral health care clinics [16]. OCHIN is a nonprofit health care innovation center that offers a fully hosted, highly customized instance of Epic practice management and EHR solutions to 140 members representing 1071 clinic sites, 629 (59%) of whom provide behavioral health services. All OCHIN members are trained on common workflows. Data on “exposure to gun violence” are incorporated as an optional standardized, reportable field as part of the behavioral health trauma SmartForm within the patient history section of the EHR used by behavioral health clinicians. While associated workflows are recommended, clinics and clinicians within the network have full autonomy to determine their use (or nonuse) of the SmartForm. As a quality improvement initiative, it is accepted that the use of the SmartForm is not consistent across all network clinics.

EHR reportable data include “recent exposure to gun violence” and “past exposure to gun violence.” Our data set represents cross-sectional data from February 2022, when the trauma SmartForm was implemented in the EHR, to December 2022. The SmartForm in use during this time only collected patient-reported “yes” responses to firearm violence exposure; there was no ability to document “no” responses. As a result, there is no denominator of total patients assessed for exposure and no practical method for estimating the prevalence of self-reported firearm violence exposure within the population.

We identified all active patients with documentation of a recent or past exposure to firearm violence and extracted data on demographics (age, sex, etc) and the prevalence of a select, predefined list of active health problems including behavioral health diagnoses, diabetes, and hypertension [18]. The list of active health problems was extracted from the patient problem list within the EHR.

Ethical Considerations

This project used a deidentified data set from a data repository of clinical and administrative data of all patients seen in the OCHIN network. The data repository is under institutional review board oversight with Advarra (Pro00060082) and includes a waiver of consent and authorization. This project was reviewed by Norwich University Institutional Review Board (00005859) and OCHIN and met exemption criteria under 45 CFR 46.

Results

During the study period, there were 129 active OCHIN member organizations with 629 behavioral health clinic sites. Among those, 91 (70.5%) member organizations and 323 (51.3%) clinics implemented the SmartForm with at least 1 patient reporting exposure to firearm violence per clinic. Between February 1 and December 31, 2022, these clinics documented between 19 and 33,067 (median 1066, IQR 356-4309) patients with firearm violence exposure, with a median age of 17 (IQR 7-64) years. Of these, a median of 43% (IQR 39%-49%; range 2%-92%) of patients were male. Across the 323 clinics, 3165 behavioral
health patients had patient-reported and clinician-documented recent or past exposure to firearm violence noted in their EHR between February 2022 and December 2022 (Table 1). Among patients with noted firearm exposure, 52.7% (n=1669) were male, 41.4% (n=1308) were between 12 and 34 years of age, 11.5% (n=364) self-reported having a sexual orientation other than straight, 38.3% (n=1229) were Black or African American, and 42.5% (n=1334) were uninsured. The median age of patients with noted exposure to firearm violence was 20 (IQR 7-64; range 3-89) years. When described by behavioral health diagnosis, 45.7% (n=1445) had a diagnosis of posttraumatic stress disorder, and 37.5% (n=1886) had a substance abuse disorder (excluding nicotine dependence). A large percentage of patients with a documented exposure had hypertension (n=371, 11.7%), and 66.7% (n=2110) had 7 or more clinic visits (for either behavioral health or primary care) during 2022. Each clinic determines which social risks, if any, they measure. Among patients screened for social determinants of health, 69.8% (n=1213) reported one or more social risks such as food insecurity or transportation insecurity.
Table 1. Demographic and active conditions of patients with self-reported exposure to recent or past firearm violence across 323 ambulatory behavioral health clinics (February to December 2022; N=3165).

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>Patients reporting exposure, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-11</td>
<td>56 (1.8)</td>
</tr>
<tr>
<td>12-17</td>
<td>193 (6.1)</td>
</tr>
<tr>
<td>18-24</td>
<td>334 (10.6)</td>
</tr>
<tr>
<td>25-34</td>
<td>781 (24.7)</td>
</tr>
<tr>
<td>35-49</td>
<td>1027 (32.4)</td>
</tr>
<tr>
<td>50+</td>
<td>774 (24.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assigned sex\textsuperscript{a}</th>
<th>Patients reporting exposure, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>1496 (47.3)</td>
</tr>
<tr>
<td>Male</td>
<td>1669 (52.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender identity</th>
<th>Patients reporting exposure, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>1433 (45.3)</td>
</tr>
<tr>
<td>Woman</td>
<td>1295 (40.9)</td>
</tr>
<tr>
<td>Transman and transwoman</td>
<td>43 (1.4)</td>
</tr>
<tr>
<td>Genderqueer; reported other gender identity</td>
<td>51 (1.6)</td>
</tr>
<tr>
<td>Decline or unknown</td>
<td>343 (10.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sexual orientation</th>
<th>Patients reporting exposure, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight</td>
<td>2226 (70.3)</td>
</tr>
<tr>
<td>Gay, lesbian, bisexual, multiple; reported other sexual orientation</td>
<td>364 (11.5)</td>
</tr>
<tr>
<td>Decline or unknown</td>
<td>575 (18.2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>Patients reporting exposure, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian or Alaska Native</td>
<td>77 (2.4)</td>
</tr>
<tr>
<td>Asian</td>
<td>49 (1.5)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1229 (38.8)</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>18 (0.6)</td>
</tr>
<tr>
<td>White</td>
<td>1368 (43.2)</td>
</tr>
<tr>
<td>Multiple races</td>
<td>51 (1.6)</td>
</tr>
<tr>
<td>Decline or unknown</td>
<td>373 (11.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Patients reporting exposure, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic or Latinx</td>
<td>543 (17.2)</td>
</tr>
<tr>
<td>Non-Hispanic or Latinx</td>
<td>2388 (75.5)</td>
</tr>
<tr>
<td>Unknown</td>
<td>234 (7.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preferred language</th>
<th>Patients reporting exposure, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>2958 (93.5)</td>
</tr>
<tr>
<td>Spanish</td>
<td>146 (4.6)</td>
</tr>
<tr>
<td>Other or unknown</td>
<td>61 (1.9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Payor type</th>
<th>Patients reporting exposure, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicaid</td>
<td>1406 (44.4)</td>
</tr>
<tr>
<td>Uninsured or other public insurance</td>
<td>1344 (42.5)</td>
</tr>
<tr>
<td>Private</td>
<td>225 (7.1)</td>
</tr>
<tr>
<td>Medicare</td>
<td>190 (6)</td>
</tr>
</tbody>
</table>
Patients reporting exposure, n (%)  

<table>
<thead>
<tr>
<th>Category</th>
<th>Count (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Veteran status (among adults age 18 years and older)</td>
<td>63 (2.2)</td>
</tr>
<tr>
<td>Homeless</td>
<td>131 (4.1)</td>
</tr>
<tr>
<td><strong>Ever screened for social determinants of health</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Reported one or more adverse social determinants of health (those ever screened)</td>
<td>1213 (69.8)</td>
</tr>
<tr>
<td>Assigned primary care provider</td>
<td>2047 (64.7)</td>
</tr>
</tbody>
</table>

**Number of clinic visits in 2022**  

<table>
<thead>
<tr>
<th>Range</th>
<th>Count (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2</td>
<td>247 (7.8)</td>
</tr>
<tr>
<td>3-6</td>
<td>808 (25.5)</td>
</tr>
<tr>
<td>7+</td>
<td>2110 (66.6)</td>
</tr>
</tbody>
</table>

**Active behavioral health conditions**  

- PTSD<sup>c</sup> and reaction to severe stress (F43)  
  - Count (Percentage): 1445 (45.7)  
- Other anxiety disorders (F41)  
  - Count (Percentage): 1268 (40.1)  
- Substance use disorders (excluding nicotine dependence; F10-F16, F18, and F19)  
  - Count (Percentage): 1186 (37.5)  
- Major depressive disorder (F33)  
  - Count (Percentage): 753 (23.8)  
- Depressive episode (F32)  
  - Count (Percentage): 683 (21.6)  

**Active medical conditions**  

- Hypertension (I10)  
  - Count (Percentage): 371 (11.7)  
- Lipid disorders (E78)  
  - Count (Percentage): 316 (10)  
- Type 2 diabetes (E11)  
  - Count (Percentage): 164 (5.2)  

<sup>a</sup>A small number of patients with other or unknown sex were proportionally distributed to the female and male categories to avoid reporting small cell counts.

<sup>b</sup>Social determinants of health is a broad category, and each clinic has the autonomy to determine which, and how many, social determinants of health they collect for patients. Adverse social determinants of health include social risks including, but not limited to, food insecurity, housing insecurity, and transportation insecurity.

<sup>c</sup>PTSD: posttraumatic stress disorder.

**Discussion**

**Principal Findings**

Current firearm injury research, largely limited to acute injury and death, has many shortcomings [19] that can potentially be overcome by expanding the collection of standardized data on firearm violence exposure in ambulatory care settings. A review of EHR data from safety-net clinics within the OCHIN network, which initiated a standardized question on firearm violence exposure as part of collecting information on patient history during behavioral health care visits, demonstrates that data on firearm violence can be collected as part of routine behavioral health care. EHR data include patient-level variables not currently available in common firearm injury data sets (Healthcare Cost & Utilization Project; National Vital Statistics System/Web-Based Injury Statistics Query and Reporting System) used for research, including gender identity, sexual orientation, social determinants of health, experiencing homelessness, health care use, behavioral health, and chronic disease comorbidities [20-22].

**Strengths and Limitations**

Expanding the collection of firearm violence exposure data in outpatient EHRs opens new opportunities in firearm violence research; these novel data may be leveraged to support more precise prevalence estimates of firearm violence exposure and injury, assess population-level associations between exposure and medical or behavioral health outcomes, and support the application of machine learning to develop predictive analytics for treatment planning [1,12,23].

Some limitations of this analysis arise from the nonmandatory nature of the data collection form. As OCHIN implemented the EHR trauma SmartForm as part of a quality improvement behavioral health EHR optimization initiative, clinics and clinicians within the network had full autonomy to determine their use (or nonuse) of the SmartForm. As the SmartForm is not consistently used across all OCHIN member clinics, the data presented here are representative of neither OCHIN’s patient population nor clinician’s behavior. In addition, the SmartForm currently in use only collects patient-reported “yes” responses to firearm violence exposure; there is no place to document “no” responses. As a result, there is no denominator of total patients assessed for exposure and no practical method for estimating the prevalence of self-reported firearm violence exposure within the population.

This study broadens the discussion around firearm injury by demonstrating that standardized data on firearm violence exposure can be collected in outpatient settings. Such data can...
extend our knowledge of the burden of firearm violence, which is typically limited to acute care settings including inpatient and emergency department data or mortality data, by incorporating a broader understanding of both exposure to firearm violence and firearm injury through the availability of rich demographic, clinical, psychological, and social risk data collected in ambulatory care EHRs.

**Future Directions**

Results from this study demonstrate the potential value of screening for firearm violence exposure and suggest how collection of this novel data may be leveraged to support firearm injury surveillance, understand physical and psychological firearm violence sequelae, and support treatment. Ongoing work should include identifying opportunities for more systematic screening of exposure to firearm violence across health care settings.

**Acknowledgments**

This work was supported by the AIM-AHEAD (Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity) Coordinating Center, funded by the National Institutes of Health (NIH). Research reported in this publication was supported by the office of the director, NIH Common Fund (award 1OT2OD032581-01). The work is solely the responsibility of the authors and does not necessarily represent the official view of AIM-AHEAD or the NIH.

**Data Availability**

The data sets generated and analyzed during this study are not publicly available due to source patient-level data from multiple health systems, which have restrictions regarding the availability and rerelease of data under cross-institution agreements. Data are available from the corresponding author on reasonable request and with permission of all relevant parties.

**Conflicts of Interest**

None declared.

**References**


Abbreviations

EHR: electronic health record

©Nicole Cook, Megan Hoopes, Frances M Biel, Natalie Cartwright, Michelle Gordon, Marion Sills. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 05.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Original Paper

Revealing the Mysteries of Population Mobility Amid the COVID-19 Pandemic in Canada: Comparative Analysis With Internet of Things–Based Thermostat Data and Google Mobility Insights

Kirti Sundar Sahu¹, MPH, PhD; Joel A Dubin², PhD; Shannon E Majowicz¹, PhD; Sam Liu³, PhD; Plinio P Morita¹,⁴,⁵,⁶,⁷, PEng, PhD

¹School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada
²Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada
³School of Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada
⁴Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada
⁵Research Institute of Aging, University of Waterloo, Waterloo, ON, Canada
⁶Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada
⁷eHealth Innovation, University Health Network, Toronto, ON, Canada

Corresponding Author:
Plinio P Morita, PEng, PhD
School of Public Health Sciences
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567 ext 31372
Fax: 1 519 746 6776
Email: plinio.morita@uwwaterloo.ca

Abstract

Background: The COVID-19 pandemic necessitated public health policies to limit human mobility and curb infection spread. Human mobility, which is often underestimated, plays a pivotal role in health outcomes, impacting both infectious and chronic diseases. Collecting precise mobility data is vital for understanding human behavior and informing public health strategies. Google’s GPS-based location tracking, which is compiled in Google Mobility Reports, became the gold standard for monitoring outdoor mobility during the pandemic. However, indoor mobility remains underexplored.

Objective: This study investigates in-home mobility data from ecobee’s smart thermostats in Canada (February 2020 to February 2021) and compares it directly with Google’s residential mobility data. By assessing the suitability of smart thermostat data, we aim to shed light on indoor mobility patterns, contributing valuable insights to public health research and strategies.

Methods: Motion sensor data were acquired from the ecobee “Donate Your Data” initiative via Google’s BigQuery cloud platform. Concurrently, residential mobility data were sourced from the Google Mobility Report. This study centered on 4 Canadian provinces—Ontario, Quebec, Alberta, and British Columbia—during the period from February 15, 2020, to February 14, 2021. Data processing, analysis, and visualization were conducted on the Microsoft Azure platform using Python (Python Software Foundation) and R programming languages (R Foundation for Statistical Computing). Our investigation involved assessing changes in mobility relative to the baseline in both data sets, with the strength of this relationship assessed using Pearson and Spearman correlation coefficients. We scrutinized daily, weekly, and monthly variations in mobility patterns across the data sets and performed anomaly detection for further insights.

Results: The results revealed noteworthy week-to-week and month-to-month shifts in population mobility within the chosen provinces, aligning with pandemic-driven policy adjustments. Notably, the ecobee data exhibited a robust correlation with Google’s data set. Examination of Google’s daily patterns detected more pronounced mobility fluctuations during weekdays, a trend not mirrored in the ecobee data. Anomaly detection successfully identified substantial mobility deviations coinciding with policy modifications and cultural events.
Conclusions: This study’s findings illustrate the substantial influence of the Canadian stay-at-home and work-from-home policies on population mobility. This impact was discernible through both Google’s out-of-house residential mobility data and ecobee’s in-house smart thermostat data. As such, we deduce that smart thermostats represent a valid tool for facilitating intelligent monitoring of population mobility in response to policy-driven shifts.

KEYWORDS
population-level health indicators; internet of things; public health surveillance; mobility; risk factors; chronic diseases; chronic; risk; surveillance; mobility; movement; sensor; population

Introduction
The dynamics of human mobility are currently undergoing a remarkable transition. Rapid global urbanization, sedentary lifestyles, infectious diseases, air pollution, and climate change are some of the factors driving shifts in human mobility. Whether examined at the individual [1,2] or population level [3], human mobility patterns are associated with multiple public health and social issues. Dwindling mobility patterns are linked to chronic diseases including dementia and age-related physical decline. For infectious diseases, individual mobility is linked to the spread of infections such as COVID-19. Human mobility data can therefore be an effective tool for comprehending the complexities of human health and behavior [4].

The COVID-19 pandemic has served as an ideal case study to develop tools and knowledge in the field of human mobility research [5]. The World Health Organization declared COVID-19 an international public health emergency in late January 2020 and subsequently declared it a pandemic on March 11, 2020. To curb the spread of COVID-19 and avoid overwhelming health care institutions [6], many countries implemented restrictions on human mobility including social distancing, self-isolation, closure of nonessential services, work-from-home policies, and travel restrictions [7,8]. In the early stages of the pandemic, digital data sources found that the decrease in human mobility was closely paralleled by a reduction in the incidence of COVID-19 [1,7,9-15]. Thus, the COVID-19 pandemic provides an ideal opportunity to measure and extract meaning from human mobility and how it is affected by restrictive policies.

The study of human mobility has historically been retrospective with limited study participants. Interestingly, it was the technological advancements in smartphones and wearable devices that have had the largest impact on the field [16]. These innovations now provide direct access to human location, trajectories, opinions, and interactions [17]. Human mobility can now be tracked passively in real time from various sources: GPS-enabled smartphones, texts or photos via photo-sharing platforms like Twitter (X Corp) and Flickr (Altaba, Ludicorp), geolocation-enabled internet posts, public transport cards, satellite, flight traffic, and even credit card transactions. The data collected have a previously unprecedented level of detail, immediacy, and precision. These large data sets, equivalent to “big data” in volumes, are now being analyzed to describe human movement patterns, characteristics (such as sleep, stress, and activity), and interactions.

Google Maps is the most popular navigation app in the United States and Canada. The app surpassed 23 million downloads in 2020 with 154.4 million monthly users. Google passively generates and collects over 20 million pieces of mobility data per day. The information has now been made available to researchers and policy makers through Google’s open-source “COVID-19 Community Mobility Reports” [18]. The reported geographic movement, called macromobility, is grouped by category: retail and recreation, grocery and drug stores, parks, transit stations, workplaces, and residences [18]. While macromobility data are readily available through sources such as Google’s “COVID-19 Community Mobility Reports,” our understanding of micromobility at the population level remains limited. This study aims to fill this gap by analyzing smart thermostat data, thereby providing a more comprehensive picture of human mobility patterns. This study leverages the unprecedented access to real-time human mobility data provided by smart thermostats, a source that has not been extensively used in previous research. The high level of detail, immediacy, and precision of these data allows for a more granular understanding of human mobility patterns, particularly in-house mobility (micromobility), which has been less explored compared to macromobility. Unlike Google’s macromobility data, smart thermostat data can provide insights into in-house mobility patterns. This is particularly relevant in the context of the COVID-19 pandemic, where stay-at-home orders and work-from-home policies have significantly altered in-house mobility patterns.

Modular smart home thermostats offer a novel source of data on in-house human behaviors [19,20]. Not only can they report on indoor temperature, humidity, and air quality, but embedded motion sensors can capture mobility within the home. More than 90% of Canadian households had thermostats in 2018, with most opting for programmable thermostats. Smart thermostats, often known as internet of things (IoT) devices, are a type of programmable thermostat that may be connected to the internet. ecobee has the second-highest market share for smart thermostats in Canada. ecobee has a program called Donate Your Data (DYD), in which subscribers can opt to make their anonymized data available for research purposes. DYD currently has 1 million users including over 172,000 households in Canada. Here, we sought to compare mobility data obtained from ecobee smart thermostats to the “gold standard” mobility data from Google during the COVID-19 pandemic in Canada. We explored day-by-day, week-by-week, and month-by-month seasonality patterns and applied anomaly detection to both data sets.
Methods

Data Sources
Google Mobility data were collected for each province in Canada from February 15, 2020, to February 14, 2021 [18]. Google Maps uses aggregated and anonymized data to determine the daily total number of visits to specific destinations Textbox 1 visited by individuals who have enabled their location history [18]. Daily values are compiled across individuals and are compared to the baseline value for that day of the week to determine changes in mobility. The baseline is the median of corresponding days over the 5 weeks from January 3 to February 6, 2020 [18]. Out of the 6 categories, we focused on residential data. We curated mobility data from ecobee’s DYD program.

Textbox 1. Google mobility data categories and their description as described on the website [18].

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery and pharmacy</td>
<td>“Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies.”</td>
</tr>
<tr>
<td>Parks</td>
<td>“Mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens.”</td>
</tr>
<tr>
<td>Transit stations</td>
<td>“Mobility trends for places like public transport hubs such as subway, bus, and train stations.”</td>
</tr>
<tr>
<td>Retail and recreation</td>
<td>“Mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theatres.”</td>
</tr>
<tr>
<td>Residential</td>
<td>“Mobility trends for places of residence.”</td>
</tr>
<tr>
<td>Workplaces</td>
<td>“Mobility trends for places of work.”</td>
</tr>
</tbody>
</table>

Ethical Considerations
Per ethical research standards, this study exclusively used secondary data, and therefore, no additional consent from human subjects was necessary for this research. The original informed consent procedures or institutional review board approvals for primary data collection explicitly permitted the secondary analysis conducted in this study. Furthermore, robust privacy and confidentiality safeguards were implemented to ensure the anonymity and deidentification of study data. Ethics approval for this study was duly obtained from the University of Waterloo Office of Research Ethics (#31377), attesting to our commitment to upholding the highest ethical standards in research.

Data Preparation
ecobee’s data were transferred from Google’s BigQuery to the Microsoft Azure cloud services platform [25]. Data were prepared using Azure Databricks data analytics platform and Jupyter Notebook (Jupyter Team) using Python; detailed explanation of the whole process has been mentioned in our previous studies [20,23]. Data cleaning, analysis, and visualization were done in R studio (version 1.4.1106; Posit Software, PBC) with R software (version 4.0.5; The R Foundation) and data analysis packages tidyverse [26] and timetk [27].

Using ecobee data, we established a baseline mobility value akin to the Google data preparation approach detailed earlier for seamless comparison. To ascertain daily in-house mobility, we computed the total number of activated sensors within each 24-hour timeframe by summing all sensor statuses. A table containing date records alongside the average count of activated sensors per date was then saved as a CSV file.

As the timestamp of the DYD data set was in Coordinated Universal Time format, the time zones in the data set were converted by locating time zone information from the geolocation of the households in the metadata. In Quebec and Alberta, all the cities were in the same time zone. In the province of Ontario, 6 cities have different time zones than EST: namely, Drayton, Kenora, Kenora-Unorganized, Mitchell, Red Lake, and Sioux Lookout cities (1% of the DYD data set for Ontario). For British Columbia, 28 households were from different time zones than Pacific Standard Time and were excluded from the analysis. Once cleaned for the time zone, time-series data

https://publichealth.jmir.org/2024/1/e46903 JMIR Public Health Survell 2024 | vol. 10 | e46903 | p.2081
(page number not for citation purposes)
analysis was performed on the adjusted data on the households included in this study. These numbers are presented in Table 1.

<table>
<thead>
<tr>
<th>Location</th>
<th>Households before time zone cleaning, n</th>
<th>Households excluded for time zone difference, n</th>
<th>Households after time zone cleaning, n</th>
<th>The proportion of household data by province, n/N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>21,690</td>
<td>_a</td>
<td>21,690</td>
<td>_</td>
</tr>
<tr>
<td>Ontario</td>
<td>7145</td>
<td>11</td>
<td>7134</td>
<td>7134/21,690 (32.9)</td>
</tr>
<tr>
<td>Alberta</td>
<td>3989</td>
<td>0</td>
<td>3989</td>
<td>3989/21,690 (18.4)</td>
</tr>
<tr>
<td>British Columbia</td>
<td>449</td>
<td>28</td>
<td>421</td>
<td>421/21,690 (1.9)</td>
</tr>
<tr>
<td>Quebec</td>
<td>708</td>
<td>0</td>
<td>708</td>
<td>708/21,690 (3.2)</td>
</tr>
</tbody>
</table>

*aNot available.

Data Analysis

To assess mobility variations, we aggregated daily province-level movement into weekly, monthly, and day-of-the-week periods and compared them to the respective baselines in both data sets. Time series plots were generated for each province using Google’s residential and ecobee mobility data. The statistical significance of the relationship between the 2 data sources was determined through Pearson and Spearman correlation coefficients [28].

We conducted seasonal diagnostic tests using the `timetk` [27] package in R software, tailored for time-series data analysis. Distinct approaches were employed to investigate daily, weekly, and monthly seasonality. To assess statistical significance [29] 1-way ANOVA was used.

For anomaly detection, we initially conducted seasonal and trend decomposition using the Loess method [27,30,31]. After removing the trend and seasonality components, we performed anomaly detection on the residual data. Anomalies were identified based on the IQR, specifically the difference between the 75th and 25th percentiles, establishing the distribution of the remaining data. Default boundaries were set at 3x above or below the IQR, designating values beyond these limits as anomalies.

To examine data granularity, we analyzed mobility data by days of the week for both Google and ecobee data sets. Additionally, we aggregated mobility data into 1-week intervals for all 4 provinces to explore seasonality in mobility changes following the onset of the COVID-19 pandemic.

Results

Positive Association Between Google and ecobee Mobility Data

The comparison of Google residential mobility data and ecobee mobility data across each province over a year revealed a positive association (Figure 1). Notably, the Google mobility data exhibited a recurrent weekly pattern, which was closely mirrored by the ecobee data (Figure 1). Both data sources exhibited a significant uptick in residential mobility, both indoors and outdoors, commencing around March 11, 2020, aligning with the declaration of the COVID-19 pandemic in Canada. It is worth noting that given Canada’s provincial health regulation framework [32], the official implementation dates of pandemic-related policies varied among the 4 provinces under investigation (March 17, 2020, for Ontario; March 16, 2020, for Quebec; March 17, 2020, for Alberta; and March 19, 2020, for British Columbia).
All provinces exhibited a notable increase in the percentage of mobility change coinciding with the commencement of the pandemic, as indicated by both Google and ecobee mobility data. Subsequently, from March 2020 to September 2020, a gradual decline in this trend was observed across all provinces in both data sets. Notably, a subsequent surge in residential mobility was linked to the onset of the second pandemic wave and the reinstatement of pandemic-related stay-at-home measures on specific dates (December 26, 2020, for Ontario; December 17, 2020, for Quebec; December 13, 2020, for Alberta; and October 19, 2020, for British Columbia).

For each province, we calculated the correlation between Google’s residential mobility data and ecobee’s mobility data. The trend line displayed a linear relationship between the 2 data sets (see Figure 2). Pearson and Spearman correlation coefficients revealed a statistically significant association, ranging from 0.67 to 0.73 (refer to Table 2). Consequently, the data obtained from ecobee’s smart home thermostat are demonstrably equivalent to Google’s mobility data.
Figure 2. Correlation between residential mobility data from ecobee and Google for the province of (A) Ontario, (B) Alberta, (C) Quebec, and (D) British Columbia, Canada. A high resolution version is available in Multimedia Appendix 1.

Table 2. Correlation between Google and ecobee mobility data for Ontario, Alberta, Quebec, and British Columbia.

<table>
<thead>
<tr>
<th>Province</th>
<th>Households, n</th>
<th>Pearson correlation coefficient (95% CI)</th>
<th>Spearman rank correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontario</td>
<td>7134</td>
<td>0.73 (0.67-0.77)</td>
<td>0.75</td>
</tr>
<tr>
<td>Alberta</td>
<td>3989</td>
<td>0.73 (0.69-0.78)</td>
<td>0.76</td>
</tr>
<tr>
<td>Quebec</td>
<td>708</td>
<td>0.67 (0.61-0.73)</td>
<td>0.70</td>
</tr>
<tr>
<td>British Columbia</td>
<td>421</td>
<td>0.69 (0.64-0.74)</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Variations in Mobility by Day of the Week

There was a significant difference in daily mobility patterns across all 4 provinces when examining the Google residential mobility data set (Figure 3). Greater mobility changes were observed on the weekdays compared to weekends. Further, a 1-way ANOVA test (Table 3) showed that the day of the week had a statistically significant impact on residential mobility for all 4 provinces (Google; all $P<0.001$). On the contrary, these differences in daily mobility were only observed in Quebec when we used ecobee’s mobility data.
Figure 3. Analysis of the (A) Google residential and (B) ecobee mobility data in days of a week in Ontario, Quebec, Alberta, and British Columbia. A high resolution version is available in Multimedia Appendix 1.

Table 3. The ANOVA test compares the day of the week’s impact on Google and ecobee mobility for the 4 provinces of Canada.

<table>
<thead>
<tr>
<th>Province</th>
<th>Google</th>
<th></th>
<th>ecobee</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum of squares</td>
<td>F test (df)</td>
<td>P value</td>
<td>Sum of squares</td>
</tr>
<tr>
<td>Ontario</td>
<td>17.86 (6, 358)</td>
<td>&lt;.001</td>
<td>0.458 (6, 358)</td>
<td>.84</td>
</tr>
<tr>
<td>Weekday</td>
<td>3467</td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>11,579</td>
<td></td>
<td>1306</td>
<td></td>
</tr>
<tr>
<td>Quebec</td>
<td>9.357 (6, 358)</td>
<td>&lt;.001</td>
<td>2.364 (6, 358)</td>
<td>.03</td>
</tr>
<tr>
<td>Weekday</td>
<td>2426</td>
<td></td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>15,471</td>
<td></td>
<td>3278</td>
<td></td>
</tr>
<tr>
<td>Alberta</td>
<td>8.223 (6, 358)</td>
<td>&lt;.001</td>
<td>1.123 (6, 358)</td>
<td>.35</td>
</tr>
<tr>
<td>Weekday</td>
<td>1456</td>
<td></td>
<td>22.9</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>10,566</td>
<td></td>
<td>1216</td>
<td></td>
</tr>
<tr>
<td>British Columbia</td>
<td>6.955 (6, 358)</td>
<td>&lt;.001</td>
<td>0.371 (6, 358)</td>
<td>.90</td>
</tr>
<tr>
<td>Weekday</td>
<td>897</td>
<td></td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>7673</td>
<td></td>
<td>2404</td>
<td></td>
</tr>
</tbody>
</table>

Monthly Variations in Mobility

To deepen our understanding of the pandemic-related changes in population behavior, we aggregated the Google and ecobee data to analyze month-by-month mobility changes. Interestingly, in contrast to the differences observed between Google and ecobee mobility data for days of the week, month-by-month patterns were similar between the 2 data sets (Figure 4). Across all provinces, the change in mobility above baseline spiked in April 2020 consistent with the implementation of COVID-19 policies to curb social mobility. Further, the variability within the data, specifically for ecobee data, was reduced drastically from April 2020 onwards. The increase in residential and in-home mobility slowly declined from April 2020 to September 2020. A subsequent rise in mobility from October 2020 to December 2020 corresponded to the second pandemic wave and the reimplementation of pandemic-related stay-at-home policies. Furthermore, 1-way ANOVA test showed a statistically significant change in both residential (Google) and in-house (ecobee) month-by-month mobility across all 4 provinces (all P<.001; Table 4).
Figure 4. Analysis of the (A) Google residential and (B) ecobee mobility data in month-by-month for Ontario, Alberta, Quebec, and British Columbia. A high resolution version is available in Multimedia Appendix 1.

Table 4. The ANOVA test compares month-by-month impact on Google and ecobee mobility for the 4 provinces of Canada.

<table>
<thead>
<tr>
<th>Province</th>
<th>Google</th>
<th></th>
<th>ecobee</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum of squares</td>
<td>F test (df)</td>
<td>P value</td>
<td>Sum of squares</td>
</tr>
<tr>
<td>Ontario</td>
<td>23.65 (11, 353)</td>
<td>&lt;.001</td>
<td>26.18 (11, 353)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Month 6383</td>
<td></td>
<td>591.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residual 8663</td>
<td></td>
<td>724.8</td>
<td></td>
</tr>
<tr>
<td>Quebec</td>
<td>36.64 (11, 353)</td>
<td>&lt;.001</td>
<td>21.61 (11, 353)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Month 9541</td>
<td></td>
<td>1372</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residual 8357</td>
<td></td>
<td>2037</td>
<td></td>
</tr>
<tr>
<td>Alberta</td>
<td>31.16 (11, 353)</td>
<td>&lt;.001</td>
<td>30.76 (11, 353)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Month 5923</td>
<td></td>
<td>606.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residual 6099</td>
<td></td>
<td>632.7</td>
<td></td>
</tr>
<tr>
<td>British Columbia</td>
<td>28.83 (11, 353)</td>
<td>&lt;.001</td>
<td>24.49 (11, 353)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Month 4062</td>
<td></td>
<td>1047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Residual 4508</td>
<td></td>
<td>1372</td>
<td></td>
</tr>
</tbody>
</table>

Week-by-Week Mobility Changes

Beginning in February 2020, the initial 5 to 6 weeks had a similar level of mobility across all 4 provinces (Figure 5). Google residential mobility declined from week 7 to week 10. Although there was a high level of data variability at week 7, this was lost in subsequent weeks. ecobee in-house mobility followed a similar pattern but with a lag period of 1 week and a large degree of variability. Beginning at week 12 (corresponding to the week of March 16, 2020), Google residential data witnessed a sharp spike in residential mobility across all 4 provinces. This timing correlates to the date the pandemic was declared in Canada. A similar trend was observed for the ecobee data, however, mobility appeared to increase starting in week 11. Similar to the trends observed in month-by-month data (Figure 4), the elevated mobility steadily declined toward week 25 and stabilized above baseline until approximately week 40 (Figure 5). There was a subsequent steady increase in mobility from week 40 until the end of the year corresponding to the timing of the second wave of the pandemic in Canada. The ecobee mobility data showed a similar trend with the exception that the decline in mobility took place over a longer period and with a shorter period of stabilization before rising again. An ANOVA analysis for the week-by-week mobility data showed a statistically significant difference between weeks for both data sets and for all 4 provinces (all P<.001; Table 5).
**Figure 5.** Analysis of the (A) Google residential and (B) ecobee mobility data in week-by-week for Ontario, Alberta, Quebec, and British Columbia. A high resolution version is available in Multimedia Appendix 1.

**Table 5.** The ANOVA test comparing the week-by-week impact on Google and ecobee mobility for the 4 provinces of Canada.

<table>
<thead>
<tr>
<th>Province</th>
<th>Google</th>
<th></th>
<th></th>
<th></th>
<th>ecobee</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum of squares</td>
<td>$F$ test ($df$)</td>
<td>$P$ value</td>
<td>Sum of squares</td>
<td>$F$ test ($df$)</td>
<td>$P$ value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ontario</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week</td>
<td>10,140</td>
<td>12.4 (52, 312)</td>
<td>&lt;.001</td>
<td>992.9</td>
<td></td>
<td>18.43 (52, 312)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>4906</td>
<td>323.2</td>
<td></td>
<td>323.2</td>
<td></td>
<td>13.79 (52, 312)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Quebec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week</td>
<td>13,582</td>
<td>18.88 (52, 312)</td>
<td>&lt;.001</td>
<td>2375</td>
<td></td>
<td>13.79 (52, 312)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>4316</td>
<td>1034</td>
<td></td>
<td>1034</td>
<td></td>
<td>17.84 (52, 312)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Alberta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week</td>
<td>8961</td>
<td>17.57 (52, 312)</td>
<td>&lt;.001</td>
<td>927.3</td>
<td></td>
<td>17.84 (52, 312)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>3061</td>
<td>311.9</td>
<td></td>
<td>311.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>British Columbia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week</td>
<td>6558</td>
<td>19.49 (52, 311)</td>
<td>&lt;.001</td>
<td>1675.8</td>
<td></td>
<td>13.53 (52, 312)</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>2012</td>
<td>742.9</td>
<td></td>
<td>742.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Social Mobility Policy Restrictions and Festive Periods Create Mobility Anomalies**

To determine whether Google and ecobee data analysis could pick out behavioral anomalies in mobility associated with policy changes, we performed an anomaly detection analysis (Figure 6). Neither Google residential data nor ecobee in-house mobility data showed any anomaly within 2020 for Ontario. For Quebec, Alberta, and British Columbia, anomalies were found at the beginning period in both the Google and ecobee data. Notably, these corresponded with the dates of COVID-19–related policy changes. In Alberta and British Columbia, ecobee’s in-house mobility analysis captured anomalies in May and June 2020, which correspond to phase-wise reopening plans and the lifting of social restrictions. Interestingly, Google residential data were able to capture anomalies corresponding to festive periods such as Christmas and New Year’s Eve in Quebec, Alberta, and British Columbia. These anomalies were not seen in the ecobee data. Overall, these results demonstrate the ability of Google’s residential data and ecobee’s thermostat data to capture notable shifts in population behavior as a result of policy changes and cultural festivities.
Discussion

This study aimed to investigate the potential application of in-house mobility data obtained from ecobee sensors in comparison to residential Google mobility data. The most significant differences between data sets were observed, particularly when mobility was analyzed by the day of the week. While Google’s residential data exhibited a significant correlation with the day of the week across all 4 provinces, ecobee’s data showed a significant impact only in Quebec. However, when scrutinized on a monthly and weekly basis, consistent findings were observed across all provinces and data sets. A notable surge in in-house and residential mobility occurred between March and April 2020, coinciding with the initiation of pandemic-related policy changes in Canada. These mobility shifts gradually diminished until September 2020, and the onset of the second wave of the pandemic, coupled with the reinstatement of social policies, corresponded with increased mobility from October to December 2020. In summary, a statistically significant association between the 2 data sets was identified. Anomaly detection analysis provided evidence supporting the capability of both data sets to detect deviations in population mobility, capturing events such as emergency declarations, reopening phases, and festive days such as December 25 (Christmas) and January 1 (New Year’s Eve). These findings underscore the benefits of employing public health surveillance mechanisms, especially during health emergencies of pandemic proportions.

Historically, monitoring individual mobility has been hampered by insufficient sample sizes, difficulty in collecting data, recall bias, and privacy concerns [33-35]. The advent of smartphones, wearable devices, and the IoT has paved the way for researchers to capture individual geolocation and movement within the home (micromobility) [36-40]. For both micro- and macromobility, it is necessary to integrate data from multiple sources [41]. Thermostat-based IoT data, such as the ecobee DYD data reported here, can provide new opportunities for calculating population-level mobility indicators. IoT is a modern, passive sensing tool that can quantify in-house movement. Understanding in-house movement is particularly relevant given that people now spend more than 80% of their time indoors [4,42]. The use of these large data sets, combined with geographical data and timestamps [16,17], has the potential to unravel different dimensions of human behavior and lifestyle.

In-house mobility data can measure changes in sleep [5,43], physical activity [44], sedentary behaviors [45-47], and movement patterns [48] with greater detail and granularity. Motion sensors are a next-generation tool with a wealth of opportunities for application in the field of public health. Integrated into smart cities, motion sensors have the power to detect overall human mobility from various sources. Data collection will enable effective planning and implementation of responsive and preventative public health strategies [37].

Monitoring real-time population mobility is important in public health as it plays a significant role in both chronic and infectious diseases. A dwindling mobility pattern is a predisposition to various chronic diseases including dementia and age-related physical decline. For infectious diseases, individual mobility is directly proportional to the transmission of infections such as COVID-19 [49]. Although the “stay-at-home” and “work-from-home” strategies were promoted globally to curb the spread of COVID-19, it is still unclear to what extent individuals complied with policy restrictions. Mobility data analysis has the potential to provide real-time information about the impact of such policies on individual and population behavior [12,50,51].

However, a notable limitation of our study was the absence of sociodemographic information in the ecobee data set. Consequently, our analysis was confined to spatiotemporal dimensions, and we were unable to examine the impact of sociodemographic features, including age, gender, and occupation. Varied perceptions surrounding infectious diseases, vaccinations, social mobility, and government policies across cultural and socioeconomic groups highlight the need for caution when generalizing the results [52]. Although the data were collected across 4 Canadian provinces, encompassing 87% of the population, the demographic represented by those with smart home thermostats likely skews toward a specific group—namely, young, tech-savvy individuals with higher socioeconomic status, who may be more inclined to work from home. Therefore, caution should be exercised in generalizing the reported results. Additionally, challenges arise in separating mobility patterns in multiperson households and eliminating sensor activation due to factors such as animals, rapid airflow, or other noises.

In conclusion, the real-time monitoring of population-level mobility using smartphones and IoT sensors has emerged as a...
recent development in public health, primarily in response to the COVID-19 pandemic. This study investigates the utility of IoT-based mobility data, specifically from smart thermostats, for assessing individual mobility within the context of social isolation policies. The findings demonstrate a close alignment between thermostat mobility data and the data presented in the Google Mobility Report, affirming its value as a mobility monitoring tool. The acquisition of real-time mobility data from smart thermostats has the potential to enhance our understanding of the intricate social determinants of health, providing valuable insights for the formulation of public health policies.

Acknowledgments
The authors would like to thank ecobee for sharing the data with us. The authors would also like to thank the teammates at the Ubiquitous Health Technology Lab (UbiLab) for their help in data management, quality checking of the code, and visualizations.

Data Availability
The Google mobility data can be readily accessed by the public [18]. In contrast, access to ecobee data is restricted to researchers and scientists through a data-sharing agreement. If you are interested in accessing ecobee data, please email the research division of ecobee.

Authors' Contributions
KSS and PPM conceived and designed this study, KSS performed data extraction, cleaning, and analysis. KSS wrote the paper under the supervision of PPM, JAD, SEM, and SL. All the authors have reviewed and approved the final manuscript and provided edits and feedback to strengthen it.

Conflicts of Interest
KSS is an Editorial Board member of JMIR Public Health and Surveillance at the time of this publication. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
High resolution images of Figures 1-6.

References


23. Jalali N, Sahu KS, Oetomo A, Morita PP. Usability of smart home thermostat to evaluate the impact of weekdays and seasons on sleep patterns and indoor stay: observational study. JMIR mHealth uHealth 2022;10(4):e28811 [FREE Full text] [doi: 10.2196/28811]


47. Sahu et al. JMIR PUBLIC HEALTH AND SURVEILLANCE. Sahu et al. JMIR Public Health Survell 2024 | vol. 10 | e46903 | p.2091 https://publichealth.jmir.org/2024/1/e46903


Abbreviations

DYD: Donate Your Data
IoT: internet of things
Search Engines and Generative Artificial Intelligence Integration: Public Health Risks and Recommendations to Safeguard Consumers Online

Amir Reza Ashraf1, PharmD; Tim Ken Mackey2,3,4, MAS, PhD; András Fittler1, PharmD, PhD

1Department of Pharmaceutics, Faculty of Pharmacy, University of Pécs, Pécs, Hungary
2Global Health Program, Department of Anthropology, University of California, La Jolla, CA, United States
3Global Health Policy and Data Institute, San Diego, CA, United States
4S-3 Research, San Diego, CA, United States

Corresponding Author:
Tim Ken Mackey, MAS, PhD
Global Health Program
Department of Anthropology
University of California
9500 Gilman Dr.
MC 0505
La Jolla, CA, 92093
United States
Phone: 1 9514914161
Email: tmackey@ucsd.edu

Abstract

Background: The online pharmacy market is growing, with legitimate online pharmacies offering advantages such as convenience and accessibility. However, this increased demand has attracted malicious actors into this space, leading to the proliferation of illegal vendors that use deceptive techniques to rank higher in search results and pose serious public health risks by dispensing substandard or falsified medicines. Search engine providers have started integrating generative artificial intelligence (AI) into search engine interfaces, which could revolutionize search by delivering more personalized results through a user-friendly experience. However, improper integration of these new technologies carries potential risks and could further exacerbate the risks posed by illicit online pharmacies by inadvertently directing users to illegal vendors.

Objective: The role of generative AI integration in reshaping search engine results, particularly related to online pharmacies, has not yet been studied. Our objective was to identify, determine the prevalence of, and characterize illegal online pharmacy recommendations within the AI-generated search results and recommendations.

Methods: We conducted a comparative assessment of AI-generated recommendations from Google’s Search Generative Experience (SGE) and Microsoft Bing’s Chat, focusing on popular and well-known medicines representing multiple therapeutic categories including controlled substances. Websites were individually examined to determine legitimacy, and known illegal vendors were identified by cross-referencing with the National Association of Boards of Pharmacy and LegitScript databases.

Results: Of the 262 websites recommended in the AI-generated search results, 47.33% (124/262) belonged to active online pharmacies, with 31.29% (82/262) leading to legitimate ones. However, 19.04% (24/126) of Bing Chat’s and 13.23% (18/136) of Google SGE’s recommendations directed users to illegal vendors, including for controlled substances. The proportion of illegal pharmacies varied by drug and search engine. A significant difference was observed in the distribution of illegal websites between search engines. The prevalence of links leading to illegal online pharmacies selling prescription medications was significantly higher (P=0.001) in Bing Chat (21/86, 24%) compared to Google SGE (6/92, 6%). Regarding the suggestions for controlled substances, suggestions generated by Google led to a significantly higher number of rogue sellers (12/44, 27%; P=.02) compared to Bing (3/40, 7%).

Conclusions: While the integration of generative AI into search engines offers promising potential, it also poses significant risks. This is the first study to shed light on the vulnerabilities within these platforms while highlighting the potential public health implications associated with their inadvertent promotion of illegal pharmacies. We found a concerning proportion of AI-generated recommendations that led to illegal online pharmacies, which could not only potentially increase their traffic but also further
exacerbate existing public health risks. Rigorous oversight and proper safeguards are urgently needed in generative search to mitigate consumer risks, making sure to actively guide users to verified pharmacies and prioritize legitimate sources while excluding illegal vendors from recommendations.

*(JMIR Public Health Surveill 2024;10:e53086)* doi:10.2196/53086

**KEYWORDS**

generative artificial intelligence; artificial intelligence; comparative assessment; search engines; online pharmacies; patient safety; generative; safety; search engine; search; searches; searching; website; websites; Google; Bing; retrieval; information seeking; illegal; pharmacy; pharmacies; risk; risks; consumer; consumers; customer; customers; recommendation; recommendations; vendor; vendors; substance use; substance abuse; controlled substances; controlled substance; drug; drugs; pharmacueutics; pharmaceuticals; pharmaceuticalal; medication; medications

**Introduction**

The internet has evolved into an increasingly popular platform for searching for health information and purchasing medications, with more people opting to turn to online marketplaces due to convenience and cost considerations. The online pharmacy market has experienced exponential growth during the past decade in parallel with the rapid proliferation of global e-commerce. The global online pharmacy market was valued at an estimated US $68 billion in 2021, with a compound annual growth rate of 16.8%, with research indicating that the internet (including social media) is now frequently used to purchase medicines online [1,2]. Properly regulated online pharmacies, often accessible via search engine results, dispense prescription and nonprescription medicines directly to patients especially benefiting individuals in remote areas and patients who are disabled or housebound. The COVID-19 pandemic further amplified behaviors associated with purchasing medicines via the internet; thus, most countries now have regulations in place to govern the delivery of medicinal products remotely.

However, the increasing global demand for online medication purchases has also attracted malicious actors, leading to the proliferation of illegal online pharmacies—websites that fail to meet national or international regulations and have not undergone regulatory review and verification. Illegal online pharmacies use extensive rogue digital marketing strategies and search engine optimization to boost their ranking and visibility on search engine results pages (SERPs) [3]. Due to the uncontrolled nature of the internet, patients often encounter both legitimate and illegitimate vendors while conducting searches for medicines online. While several national and international verification or accreditation systems exist, such as the National Association of Boards of Pharmacy (NABP)’s Digital Pharmacy Accreditation program and the “.pharmacy” domain registry (USA) [4] and the European Commission’s EU logo for online sale of medicines (EU) [5], patients and health professionals continue to have issues verifying the credibility of online pharmacy websites appearing in search engine results [6].

Illegal online vendors endanger health by selling medicines without requiring a valid prescription and supplying substandard and falsified medicines [7] that could lead to dangerous patient outcomes [3,8]. This illegal practice has broad public health consequences, including eroding trust in health care delivery, compromising pharmacy supply chain safety, and potentially contributing to antimicrobial resistance due to the presence of substandard and adulterated products [9]. Despite persistent warnings from researchers and regulators who have called for reform and enhanced monitoring, the continued online presence of illegal pharmacies remains largely unchecked. Law enforcement efforts have had limited effectiveness in keeping up with the growing number and diversity of illicit marketplaces, public awareness campaigns show limited efficacy in changing consumer behavior, and search engine providers have yet to enforce more stringent controls on their organic search results [10,11].

This lack of accountability, awareness, and inaction has facilitated the rampant growth of illicit online drug sales for a variety of therapeutic classes (eg, antibiotics, controlled substances, and weight loss drugs) [9,12,13]. A recent study revealed that compromised results redirecting to active illicit online pharmacies were present in search query results of several European countries, with the most affected regions having up to one-third of the SERP links associated with illegal online pharmacies [14]. Other recent public health threats include fake COVID-19 products offered via the internet during the pandemic [15,16]. Although no “magic bullet” exists, effective regulation of these websites likely lies in the hands of search engine providers, as these companies have effective methodologies to screen advertisements and prevent vendors of illegal products from using paid promotion for their services. However, unpaid organic results (ie, that are not sponsored ads) are seemingly uncontrolled.

Interest and commercial adoption of generative artificial intelligence (AI)-based conversational chat features and applications are rapidly expanding throughout society. Yet, improper integration of generative AI into search engine results could further complicate and exacerbate the illegal online pharmacy issue. As of June 2023, Google continued to dominate the global search market with 84.6 billion monthly visits, while Microsoft Bing was a distant second with 1.2 billion monthly visits according to web analytics data by Similarweb Ltd [17]. With the emergence of generative AI, especially after witnessing the surging popularity of OpenAI’s ChatGPT, search engine giants have rushed to integrate generative AI into their search interfaces, giving rise to Microsoft Bing’s Chat feature, also known as Microsoft Copilot, and Google’s Search Generative Experience (SGE). After Microsoft launched Bing Chat in February 2023, Bing search crossed 100 million daily active users for the first time in its history [18].

https://publichealth.jmir.org/2024/1/e53086

JMIR Public Health Surveill 2024 | vol. 10 | e53086 | p.2094

*(page number not for citation purposes)*
These recent developments will transform the way global users search for and interact with health information online. Large language models (LLMs) and generative AI, when implemented and used responsibly, have the potential to revolutionize search by delivering accurate, safe, and personalized results through a user-friendly experience. However, they also carry potential risks and ethical considerations, particularly when it comes to public health, as recently highlighted by the World Health Organization that has called for caution in using these technologies [19-21]. LLMs, lacking the ability to reason, may produce results with critical mistakes and have demonstrated significant drawbacks, such as generating misinformation and falsifying data, potentially leading to patient injury that in turn raises liability concerns [22] while concurrently heightening the need for a comprehensive framework to address present compliance and reliability issues, especially in regulated settings like health care [23]. Other published studies have examined the use, impact, and potential threat of LLMs in pharmacy education and practice (eg, answering clinical pharmacy questions), their use in medical consultations regarding drug-to-drug interactions and drug-related questions related to risk, and evaluated LLM-generated responses to prompts containing vaccine conspiracies and misconceptions [24-33]. However, no study to our knowledge has specifically evaluated LLMs in the context of popular search engine integration, and how they may generate content that could direct consumers to illegal websites selling medication online.

Hence, several questions arise that warrant further inquiry in the context of patient safety, information quality, and potential consumer exposure to harmful medication access associated with LLMs. The focus of this study is to conduct an exploratory study to identify whether these novel search tools will influence consumer interaction with the online pharmacy market and whether they will assist or potentially harm consumers by exposing them to illegal websites through SERPs. In response, this study conducts a structured comparative analysis using different prompts to conduct a comparative assessment of AI-generated recommendations of Google and Bing for hypothetical users looking to buy prescription drugs online.

Methods

Overview

In this study, we conduct a structured analysis of 2 generative AIs to assess the impact of commercial implementation of this new technology on search engine results associated with the online pharmaceutical market. The 2 platforms reviewed included Google SGE using converse mode, which is based on Google’s Bard generative AI chatbot, and Microsoft Bing’s Chat feature, which uses OpenAI’s GPT-4 LLM. The rationale for choosing a comparative analysis of these 2 platforms was to specifically conduct an assessment of the differences and similarities associated with generative AI integration into search engine results, with an emphasis on evaluating sourcing recommendations made that could be seen by the public.

For our investigation, we selected 6 globally popular medicinal products by proprietary name and their corresponding active pharmaceutical ingredients (APIs) with representation across multiple therapeutic categories (eg, medicines used to treat diabetes, erectile dysfunction, and bacterial infections) to assess the quality of AI-generated recommendations under a general prompt for a user seeking advice on where to buy prescription drugs from the internet and specifically requesting information on a recommended website. An additional group of controlled substances (eg, Xanax and fentanyl) were included to assess how AI-generated responses to this prompt would differ for drugs specifically subject to international control and regulation, as generally controlled substances are not permitted to be sold via internet sources, may not be trafficked online, and are subject to enhanced dispensing regulations in various countries.

AI Prompt and Website Classification

The following standard prompt was used on each generative AI platform with both the proprietary name and the APIs listed in the “Results” section: “Where can I buy [Proprietary Name/API] online?” and “I am looking for an online pharmacy to buy [Proprietary Name/API]. Please suggest a website where I can buy it!” All results were then analyzed and crosschecked with NABP’s not recommended list (maintained by the US NABP that provides a list of sites that may be fraudulent or illegitimate) and LegitScript database (an online service that monitors global online pharmacies for compliance with applicable laws and regulations and classifies illegal and legitimate websites) to determine the legitimacy of the online pharmacies recommended, a verification approach similar to those used in prior published studies on the topic [9,34,35]. Website evaluation and categorization were conducted independently by 2 authors (ARA and AF) and finalized following the initial categorization through a collaborative discussion of individual results to reach consensus in the event of any disagreement.

Initial categorization showed almost perfect agreement for interrater reliability with a κ value of 0.98. An internet pharmacy domain was classified as legitimate if it was approved by either or both the NABP and LegitScript databases. Additionally, pages were visually inspected to identify the presence of an official internet pharmacy verification logo with a functional redirection link to the website of the competent national authority. An internet pharmacy website was categorized as illegal whether the databases classified the domain as rogue or not recommended or if there was a clear indication of illegal activity, such as the sale of prescription-only medicines without requiring a valid medical prescription. In cases where users were redirected to third-party websites from the initial link, the classification was done based on the evaluation of the final destination website offering medicines for sale. Links leading to inaccessible sites (eg, error 404) underwent multiple periodic evaluation attempts and were categorized as nonreliable if domains remained inaccessible.

Generative AI searches were conducted between July 10, 2023, and July 12, 2023, using Microsoft Edge desktop browser (version 114.0.1823.37) for Bing Chat and Google Chrome desktop browser (version 114.0.5735.198) for the Google SGE platform.

Data Analysis

Data were analyzed using the SPSS Statistics (version 26; IBM Corp) program. Descriptive statistics were used to describe the
prevalence of link categories in AI-generated search engine results for each prompt. The initial level of agreement between the 2 authors’ (ARA and AF) categorization of websites was assessed with Cohen κ statistic to measure interrater reliability. Both nominal and frequency data were analyzed using a chi-square analysis, in which P values <.05 were regarded as statistically significant.

**Ethical Considerations**

All information collected from this study was from the public domain, and the study did not involve any interaction with users or user-related data.

**Results**

A total of 262 links were provided by the generative search engine replies to our queries, with 136 generated from Google SGE and 126 from Microsoft Bing Chat. Of the links provided, 47.33% (124/262) suggested an active online pharmacy website that dispensed medications. It is important to note that a larger proportion of the results provided by both search engines did recommend legitimate pharmacies (82/262, 31.29%), with Google SGE at 25.74% (35/136) and Bing Chat at 37.3% (47/126). However, we also observed a notable presence of recommended links to illegal or unlicensed online pharmacies on both platforms. Specifically, 13.23% (18/136) of Google SGE’s responses and 19.04% (24/126) of links provided in Bing Chat’s generative replies were found to direct users to known illegal online pharmacies. (Table 1 and Figure 1 for example of Google SGE recommendation for illegal online seller of antidiabetic drug semaglutide that has been reported as counterfeited and sold online, including a recommendation to the semaspace website, which has been issued a warning letter from the US Food and Drug Administration for introducing misbranded and unapproved semaglutide and has subsequently been shut down.) The remaining 61.02% (83/136) of Google’s and 43.65% (55/126) of Bing Chat’s recommendations were for informational sites, articles, or other online sources, that is, telemedicine consultation websites, not directly selling medications to consumers.

A closer examination of the results for prescription medications queried reveals distinct differences between the 2 search engines’ generative feature recommendations. This suggests that both have likely implemented some form of additional controls to filter illegal sellers from results or that these recommendations are filtered or reviewed by other training or referenced data, although correct classification is not consistent (Multimedia Appendix 1 for additional examples of illegal sellers in recommendations). Although the overall occurrence of legitimate pharmacy websites was higher (P=.08) in Bing Chat (38/86, 44%) compared to Google SGE (29/92, 31%), the number of recommendations leading to illegal online sellers was significantly higher (P=.001) for Bing Chat (21/86, 24%) compared to Google SGE (6/92, 6%). The proportion of links to rogue websites was notably higher for the antibiotic amoxicillin (9/24, 37%) and the proton pump inhibitor omeprazole (7/19, 37%) in Bing Chat. However, Google’s generative AI search results showed an absence (0%) of illegal seller recommendations for these medications. Instead, Google’s generative AI search results included several illegal websites (3/23, 13%) offering the sale of sildenafil or Viagra, a commonly counterfeited erectile dysfunction medication [36]. In contrast, Bing Chat appeared to exclude illegal sellers of this drug (Table 1).

Specific to controlled substance recommendations, these narcotic medications hold a high potential for abuse and dependence and are subject to special regulatory and legal requirements at the national (eg, national controlled substance acts) and international (eg, United Nation conventions and treaties) levels and are generally not available for purchase and dispensing online. Despite these prohibitions, suggestions for where to purchase controlled drugs were returned using the simple prompt used in this study, which led to a significantly higher (P=.02) number of rogue sellers in Google SGE’s suggestions (12/44, 27%) compared to 7% (3/40) from Bing Chat. Notably, for the popular anxiolytic alprazolam or Xanax, a substantially higher number of illegal pharmacy suggestions (10/20, 50%) was observed compared to legitimate pharmacies (2/20, 10%) in Google SGE results. Xanax is also a controlled substance subject to abuse and counterfeiting [37]. The results of recommendations for controlled substances carry heightened consumer risk due to the high potential for abuse and known counterfeiting of versions of these drugs laced with fentanyl, which has led to overdose deaths due to poisoning [38].

Bing Chat provides a generative response to every query and also provides sources by default for key parts of the generated response. However, these links do not always directly relate to the topic of the AI-generated text, and in some instances, these may even be contradictory. For instance, when we asked Bing Chat, “Where can I buy fentanyl online?” the generated response began with, “I’m sorry, but I cannot help you with that.” This was followed by a well-reasoned explanation that “fentanyl is highly addictive and dangerous and can cause serious harm or even death.” Subsequently, it explained that “it is illegal to buy or sell fentanyl without a prescription,” and added, “I strongly advise you to avoid buying fentanyl online or anywhere else and seek professional help if you are struggling with addiction.” Finally, Bing AI offered help in finding resources for addiction treatment (Figure 2, screenshot on the left).
Table 1. Recommendations by generative AI\(^a\)-powered searches conducted using Microsoft’s Bing Chat and Google search generative experience for prescription medicine purchase–focused search terms.

<table>
<thead>
<tr>
<th>Indication (ATC code)(^b) and API(^c) and proprietary name</th>
<th>Google</th>
<th>Bing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Links provided (n=136), n</td>
<td>Legitimate pharmacy (n=35), n</td>
</tr>
<tr>
<td>Prescription-only medications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penicillin with extended spectrum (J01CA04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Amoxil</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Proton pump inhibitor (A02BC01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omeprazole</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Prilosec</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Glucagon-like peptide-analogue (A10BJ06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semaglutide</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>Ozempic</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Drug used in erectile dysfunction (G04BE03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sildenafil</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Viagra</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Controlled substances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiolytic (N05BA12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alprazolam</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Xanax</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Phenylpiperidine derivative (N02AB03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fentanyl</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Duragesic</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>

\(^a\)AI: artificial intelligence.

\(^b\)ATC code: Classification of the substance according to the World Health Organization’s anatomical therapeutic chemical (ATC) system, table indicates level-4 ATC terminology based on the ATC/DDD (defined daily dose) index.

\(^c\)API: active pharmaceutical ingredient name.
This response is perfectly appropriate and demonstrates that the AI recognized the inherent danger of the situation from the user’s query and generated a sound and constructive response. This indicates that chatbots can be programmed to produce highly aligned responses reflective of public health concerns about sourcing medications online. However, it is disconcerting to note that the links provided in the “Learn more” section are not effectively monitored. The first link given to the user for this prompt led to an illegal online pharmacy (Figure 2, screenshot on the right). This is a notable weakness of Bing Chat. The majority of concerns we observed were in the hyperlinks within the generated response or recommended links below the response in the “Learn more” section. This issue could be attributed to the lack of stringent oversight over reviewing whether organic search results generated by the search engine provider include illegal sellers, which consequently surface in generative AI-related responses. This laxity allows illegal pharmacies to rank high within the organic SERPs and, in turn, find their way into the recommendations offered to users.

At the time of the study evaluation, the Google SGE was still in early experimental access in the United States and was not available in other locations. Contrary to Bing Chat, Google SGE did not generate extensive detailed generative responses to all user queries, and at times, the generative response was simply limited to “Here are some results,” followed by recommended links. As Google SGE also provides links along with its responses, and since it relies on the organic results ranking high on the SERPs to recommend links to users, it also returned questionable recommendations as observed in Bing Chat’s responses. Specifically, we encountered instances where illegal pharmacy websites were directly recommended to the user both within the generative text and in the recommended links for both platforms.
Discussion

Principal Findings

Our study found that one-third (44/124, 35.48%) of the recommendations made for purchasing medications online from an active online pharmacy site made by 2 popular generative AIs directed users to rogue online pharmacies and that recommendations were also made for online sources of controlled substances. These findings are in line with previously published data on traditional, non-AI–generated results, including this study on the prevalence of illegal internet pharmacy links in Google search results of 12 European countries, where we identified 19.8% (380/1920) were compromised [14].

Our recent findings signal a concerning public health issue intersecting with emerging technology, particularly salient as these LLM applications enjoy widespread and rapidly growing appeal, with ChatGPT reaching 100 million users just 2 months after its launch, making it the fastest-growing consumer application in history [39]. With tens of millions of users prompting responses to these generative AI systems daily, the potential for user exposure to known unsafe and fraudulent online pharmacy websites needs further study and action.

Specifically, the inadvertent promotion of illegal and rogue online pharmacy websites by generative AI platforms may be linked to the rogue search engine optimization techniques used by bad actors to gain high rankings on SERPs. This presents a new potential vulnerability that could be exploited to influence generative AI’s responses and recommendations for other popular health questions, similar to our observations of suggestions made for high-ranking SERPs for illegal or unlicensed pharmacies. Although the total number of illegal sellers recommended by these mainstream generative AI platforms was not overwhelmingly high, the mere presence of illegitimate vendors still represents a significant potential safety risk and could introduce challenging health and safety issues, as studies have shown individuals tend to prefer computer-generated advice over human advice as tasks become more complex [15] and that they rely more on algorithmically generated advice if it aligns closely with their initial guess [16]. This confirmation bias combined with potentially erroneous AI-generated advice or recommendations could lead users to make decisions that could jeopardize their health and well-being, particularly in the context of controlled substances and other medications known to be counterfeited. It is crucial that these risks are fully acknowledged and addressed, highlighting the urgent need for greater scrutiny of the way search engines index and rank websites, as well as the sources they use for training their AI models.

From a regulatory standpoint, it is imperative that governments around the world intensify their efforts with informed,
responsible policy making to address these emerging challenges and establish a robust legal framework for rigorous regulatory oversight of AI operations, including conversational generative search engine results. In April 2021, the European Commission proposed a draft regulation on AI known as the EU AI Act, a set of requirements and obligations to gain access to the EU market, which is the first regulation of its kind on AI. The draft incorporated sensible elements such as establishing a technology-neutral definition of AI systems in EU law, paired with a risk-level–based classification for these systems, the introduction of prohibitions on AI systems presenting “unacceptable” risks [40], and a public database to enable public scrutiny and democratic oversight of AI systems. However, the EU AI Act also has certain shortcomings, largely due to it being constructed from a mix of product safety regulations, fundamental rights protection, surveillance, and consumer protection laws from the 1980s [41]. The recent approval [42] of amendments and revisions to the draft is a promising starting point, but there remains much more work to be done.

Currently, tens of thousands of websites are offering medicines for sale, with numerous rogue vendors easily accessible via traditional search engine results not assisted by generative AI. It is already challenging for consumers to differentiate between illegal and legitimate internet pharmacies. As we have previously emphasized [3], regulators and search engine providers have a shared responsibility to implement additional guardrails for AI-generated recommendations in order to ensure the protection and promotion of well-being, consumer safety, and public health. These should include real-time verification solutions built into AI systems to confirm the safety and legitimacy of online pharmacies before featuring them in search results. Search engine providers also need to take a more proactive role in directing users toward licensed and reputable pharmacies, whose lists are available on the national authority websites of many countries. Despite these calls to action, the chronic issue of illegal online pharmacies infiltrating search engine results remains unresolved and may be exacerbated by inaccurate suggestions generated by LLMs that are now integrated into search engines, as demonstrated in this study.

Limitations

We performed a comparative analysis of 2 leading generative AI-integrated search platforms accessed by millions of users daily. However, this approach has some limitations. The rationale for opting against having a nongenerative conventional search comparison group was based on the extensive pre-existing literature, already indicating the prevalence of illegal online pharmacy links in search results before generative AI integration. The primary objective of this study was instead to specifically identify and characterize whether questionable recommendations occurred with generative AI search results.

Further, it is challenging to compare structured search queries on conventional search (eg, buy [Drug Name] without a prescription) with more conversational user queries (eg, Where can I buy [Drug Name] online?) as the latter are not mere keywords but nuanced prompts for the LLM, shaping its human-like conversational response. Due to the dynamic nature of generative AI systems, similar queries might yield varied results and are not longitudinally comparable. One might perceive our findings as anomalies that are part of the development process and easy to mitigate; however, we urge stakeholders to consider this as a cautionary case study that signals a potential paradigm shift that could alter current infodemiology and infoveillance methodologies, reshaping our approach to studying online health–related information-seeking behaviors. Future studies should further explore the influence of generative AI systems on consumer search patterns while seeking medications online compared to conventional search engine queries, online forums, social media, and other user-generated content.

Conclusions

The emergence of generative AI–integrated search is a promising development with the potential to fundamentally reshape our interactions with the digital world, and its impact on public health is both unavoidable and inevitable. Our research has uncovered a concerning new trend: links to both legal and illegal online pharmacies appeared together in generative AI responses being integrated into search engine results delivered to the public, highlighting the urgent need for more comprehensive and focused oversight. With proper integration of generative AI, search engines can strategically prioritize linking to verified, legal pharmacies within generated responses, addressing the longstanding issue of illegal online medicine vendors appearing in search results. Improving generative AI search results in this manner could enhance patient safety by ensuring access to accurate information and authentic and safe pharmaceutical products. However, the realization of this potential is heavily contingent upon the decisions made by technology stakeholders about the development and deployment strategies of AI-assisted technologies. Through meticulous planning and effective regulation, we can fully harness the power of AI while prioritizing the safety of the online pharmaceutical market to safeguard public health.

Acknowledgments

The research was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (BO/00238/20/5) and the Hungarian Scientific Research Fund OTKA (grant number NKFI-ID: 143684).

Conflicts of Interest

TKM is an employee and co-owner of the small business company S-3 Research LLC. S-3 Research has been previously funded by the National Institutes of Health—National Institute of Drug Abuse through a Small Business Innovation and Research contract.
Multimedia Appendix 1

Images illustrating inappropriate generative AI responses with potential medication safety and public health concerns.

[DOCX File, 1883 KB - publichealth_v10i1e53086_app1.docx ]

References


21. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 2023 Jul 06;6(1):120. [doi: 10.1038/s41746-023-00873-0] [Medline: 37414860]


38. DEA laboratory testing reveals that 6 out of 10 fentanyl-laced fake prescription pills now contain a potentially lethal dose of fentanyl. DEA.gov. URL: https://www.dea.gov/alert/dea-laboratory-testing-reveals-6-out-10-fentanyl-laced-fake-prescription-pills-now-contain [accessed 2023-08-31]


Abbreviations

AI: artificial intelligence
API: active pharmaceutical ingredient
ATC: anatomical therapeutic chemical
LLM: large language model
NABP: National Association of Boards of Pharmacy
SERP: search engine results page
SGE: search generative experience

© Amir Reza Ashraf, Tim Ken Mackey, András Fittler. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 21.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Original Paper

Ambulance Services Attendance for Mental Health and Overdose Before and During COVID-19 in Canada and the United Kingdom: Interrupted Time Series Study

Graham Law1, PhD; Rhiannon Cooper2,3, MSc; Melissa Pirrie2, PhD; Richard Ferron3,4, MHM; Brent McLeod5, MPH; Robert Spaight6, MSc; A Niroshan Siriwardena1, PhD; Gina Agarwal2,3, PhD; UK Canada Emergency Calls Data Analysis and GEospatial Mapping (EDGE) Consortium1,2,3,4,5,6

1Community and Health Research Unit, School of Health and Social Care, University of Lincoln, Lincolnshire, United Kingdom
2Department of Family Medicine, McMaster University, Hamilton, ON, Canada
3Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
4Niagara Emergency Medical Services, Niagara, ON, Canada
5Hamilton Paramedic Service, Hamilton, ON, Canada
6East Midlands Ambulance Service NHS Trust, Nottingham, United Kingdom

Corresponding Author:
Gina Agarwal, PhD
Department of Family Medicine
McMaster University
100 Main Street West
Hamilton, ON,
Canada
Phone: 1 905 525 9140
Email: gina.agarwal@gmail.com

Abstract

Background: The COVID-19 pandemic impacted mental health and health care systems worldwide.

Objective: This study examined the COVID-19 pandemic’s impact on ambulance attendances for mental health and overdose, comparing similar regions in the United Kingdom and Canada that implemented different public health measures.

Methods: An interrupted time series study of ambulance attendances was conducted for mental health and overdose in the United Kingdom (East Midlands region) and Canada (Hamilton and Niagara regions). Data were obtained from 182,497 ambulance attendance records for the study period of December 29, 2019, to August 1, 2020. Negative binomial regressions modeled the count of attendances per week per 100,000 population in the weeks leading up to the lockdown, the week the lockdown was initiated, and the weeks following the lockdown. Stratified analyses were conducted by sex and age.

Results: Ambulance attendances for mental health and overdose had very small week-over-week increases prior to lockdown (United Kingdom: incidence rate ratio [IRR] 1.002, 95% CI 1.002-1.003 for mental health). However, substantial changes were observed at the time of lockdown; while there was a statistically significant drop in the rate of overdose attendances in the study regions of both countries (United Kingdom: IRR 0.573, 95% CI 0.518-0.635 and Canada: IRR 0.743, 95% CI 0.602-0.917), the rate of mental health attendances increased in the UK region only (United Kingdom: IRR 1.125, 95% CI 1.031-1.227 and Canada: IRR 0.922, 95% CI 0.794-1.071). Different trends were observed based on sex and age categories within and between study regions.

Conclusions: The observed changes in ambulance attendances for mental health and overdose at the time of lockdown differed between the UK and Canada study regions. These results may inform future pandemic planning and further research on the public health measures that may explain observed regional differences.

(JMIR Public Health Surveill 2024;10:e46029) doi:10.2196/46029

KEYWORDS
COVID-19; mental health; overdose; emergency medical services; administrative data; Canada; the United Kingdom; ambulance; sex; age; lockdown; pandemic planning; emergency service
Introduction

As the SARS-CoV-2 (COVID-19) pandemic has spread across the globe, it presents a major threat to mental health in general and is related to alcohol and substance use [1-3]. Government responses, in the form of travel restrictions and economic support for the COVID-19 pandemic, vary greatly from country to country [4]. Sweeping border closures and strict public health measures were variably implemented by many governments in an effort to contain the virus. However, inconsistencies in government responses may have contributed to fear and uncertainty, while strict public health measures may lead to social isolation [5].

Anxiety, depression, poor sleep quality, and psychological distress are mental health symptoms that have increased since the start of the COVID-19 pandemic [6]. Stress and poor mental health are being exacerbated by the uncertain future of the pandemic, misinformation, and social isolation brought about by physical distancing measures [7,8]. Factors such as sex and age have also been identified as risk factors for experiencing negative mental health effects of the COVID-19 pandemic [6,9].

Being younger and identifying as female have both been linked to experiencing lower psychological well-being throughout the course of the pandemic. Relatedly, there are also reports of an increase in overdoses during the COVID-19 pandemic [10,11]. This may be due to changes in the availability of illicit drugs during lockdown [2], possibly resulting in a higher rate of overdoses when access was restored, and also challenges in physically accessing pharmaceutical therapies (eg, methadone) [2]. In addition, a systematic review in 2022 reported that mental health and overdose are consistently found to be associated across studies, although the causal mechanism remains unclear [12]. In the context of COVID-19, it has been reported that social isolation has put at risk the vulnerable population of substance users, increasing the strain on mental health and subsequently further increasing the probability of an overdose [2]. Therefore, mental health and overdose are 2 related, but different, outcomes that should be examined in tandem.

Currently, studies examining the relationship between mental health and COVID-19 rely on self-reported measures [6,9,13]. Other indicators of population mental health and overdoses are of interest, especially for policy and resource planning. Emergency medical service (EMS) data, such as attendances, have previously been used to understand public health trends, which can help planning and resource allocation [14].

Due to variation in population characteristics and the government response to COVID-19 between countries, differences may exist in mental health effects. Specifically, the United Kingdom and Canada are 2 countries that experienced differing national responses to the ongoing pandemic. A report by the chief public health officer of Canada found that self-perceived mental health status decreased during the COVID-19 pandemic when compared to 2018 and that overdoses were increasing [11]. A comparison of the mental health and overdose effects of COVID-19 between Canada and the United Kingdom is of interest, as population characteristics and EMS attendance differ. Understanding the need for other indicators of mental health and overdose besides self-reporting, EMS data will be used to elucidate the impact of COVID-19 on EMS attendances.

Little is known about the trend in attendances before and after lockdown, especially for mental health and overdoses internationally. This paper aims to compare differences in the trends and volumes of mental health and overdose attendances before and after lockdown between specific regions in Canada and the United Kingdom. Evidence suggests that differences exist between mental health and overdose effects between sexes and age groups [6,9]. Therefore, this study will examine and describe the number of mental health and overdose attendances to 911 or 999 EMS calls in regions of the United Kingdom and Canada according to population subgroups.

Methods

Study Design

We used an interrupted time series design, analyzing weekly ambulance attendances before and after the COVID-19 lockdown from regions of the United Kingdom (East Midlands) and Canada (Niagara and Hamilton, Ontario). The East Midlands Ambulance Service NHS Trust is a regional EMS that operates in the East Midlands of the United Kingdom. In Canada, 2 services were examined: Niagara Emergency Medical Services is a moderate-sized EMS that operates in the Niagara region of Ontario, Canada, and Hamilton Paramedic Service is a moderate-sized EMS in Hamilton, Ontario. We purposefully selected these regions in the United Kingdom and Canada for comparison since they have similar population densities and median household incomes, and it was feasible to obtain their EMS call records.

Population

In Canada, the study population was those attended to by Niagara Emergency Medical Services and Hamilton Paramedic Service in Ontario. The population of Niagara and Hamilton was obtained from the 2021 Canadian Census Profile [15]. In the United Kingdom, the study population was the East Midlands region of England, which is served by East Midlands Ambulance Service NHS Trust. The population for the East Midlands region was obtained from the Office for National Statistics using the midyear population estimates [16] by age and sex.

Ethical Considerations

In the United Kingdom, the study was given NHS favorable ethical opinion (reference 20/SC/0307) by the South Central—Berkshire B Research Ethics Committee and is listed on the Integrated Research Application System as 286198. Patient records for this study were anonymized by the ambulance service, and informed consent could not be feasibly obtained for this retrospective study of administrative data. In Canada, the study was reviewed by the Hamilton Integrated Research Ethics Board, and a waiver letter was issued since individual patient data were not released by the paramedic services, and only weekly aggregated administrative data for the entire region were provided for this study.
Data Collection

Overview

The weekly number of ambulance attendances (Sunday to Saturday) was collected covering 2 adjacent time periods: the weeks leading up to the lockdown (December 29, 2019, until March 21, 2020, in the United Kingdom and March 14, 2020, in Canada) and the weeks during and following the lockdown (March 22, 2020 and March 15 in the United Kingdom and Canada, respectively, until August 1, 2020). The lockdown occurred on March 23, 2020, in the United Kingdom and March 15, 2020, in Canada.

Mental Health and Overdose

Ambulance call records for mental health and overdose were identified by the ambulance services through clinical impression (provisional diagnosis) codes recorded by ambulance staff when attending the patient. The UK and Canadian systems for assigning problem codes to paramedic-patient encounters are similar, though the actual codes available to select from are slightly different in each country. Work to map the 2 countries’ systems of paramedic coding is underway [17]. Where more than one problem code existed (Canada), data for all problem codes (primary, secondary, and final) were extracted. The mental health records were identified in Canada as clinical impression or problem code 45 (behavior or psychiatric) and in the United Kingdom as the clinical impressions “admission under mental health act,” “anxiety,” “attempted suicide,” “deliberate self-harm,” “depression,” “panic attack,” “psychosis,” and “other mental health.” For overdose, the records were identified in Canada as clinical impression or problem codes 81 (drug or alcohol overdose), 81.1 (opioid overdose), or 81.2 (alcohol intoxication) and in the United Kingdom as clinical impression “intentional drug overdose (mental health),” “accidental overdose or poisoning (medical),” “effects of alcohol,” and “query intoxicated (medical).”

Sex and Age Groups

Sex was recorded in Canada and the United Kingdom by the crew in attendance. Age was recorded from self-report or estimated by the crew in attendance where this was not possible. Age groups were used to categorize the data into 18 years and younger, 18-44 years, 45-65 years, and 65 years and older.

Statistical Analysis

Using an interrupted time series approach (Figure 1), the number of weekly attendances was modeled as the outcome using negative binomial regression, and the results were reported as incidence rate ratios (IRRs). As shown in Figure 1, for each regression, the prelockdown trend of weekly calls was modeled (per week), and then a counterfactual scenario was imagined, where the trend in the data would have continued without the interruption. The counterfactual scenario provided a comparison for the evaluation of the impact of the lockdown by examining the change in level at the time of the lockdown (lockdown) and the change in the slope during lockdown (lockdown trend) relative to this counterfactual scenario.

Figure 1. Interrupted time series study design used to model the change in weekly emergency medical service calls at the time of lockdown and the period following lockdown.

Negative binomial regression was identified as the most appropriate model since there was overdispersion observed in the outcome variables. The models were fitted using time (in weeks relative to the lockdown start week), lockdown (a binary categorical variable of before lockdown and during lockdown), lockdown trend (time in weeks following lockdown, which indicates the slope change following lockdown), and seasonality (an adjustment made via the categorical variable of month) with an offset of total population size. All statistical analysis was conducted using R (R Foundation for Statistical Computing).

Results

Table 1 shows there were 4.9 million people living in the area served by the ambulance service in the East Midlands region.
of the United Kingdom and 1.0 million in the Canadian regions of Hamilton and Niagara. The proportion of female individuals in each location was similar (n=2,457,905, 50.5% vs n=522,488, 51.2%), as were the proportions of individuals aged 65 years and older (n=902,947, 18.6% vs n=207,184, 20.5%). Household incomes and population densities were found to be similar, as well (Table 1). In the Canadian study regions, the ambulance services provided 1 vehicle per 14,776 persons, while in the UK study region, it was 1 per 9809 persons.

### Table 1. Description of the UK and Canadian study regions, including population size by sex and age, geographic area, household income, and paramedic capacity.

<table>
<thead>
<tr>
<th></th>
<th>East Midlands region, United Kingdom (n=4,865,583)</th>
<th>Niagara and Hamilton regions, Canada (n=1,019,510)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2,457,905 (50.5)</td>
<td>522,488 (51.2)</td>
</tr>
<tr>
<td>Male</td>
<td>2,407,678 (49.5)</td>
<td>497,022 (48.8)</td>
</tr>
<tr>
<td><strong>Age (years), n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>1,007,895 (20.7)</td>
<td>206,428 (20.4)</td>
</tr>
<tr>
<td>18-44</td>
<td>1,623,597 (33.4)</td>
<td>306,517 (30.4)</td>
</tr>
<tr>
<td>45-65</td>
<td>1,331,144 (27.4)</td>
<td>289,575 (28.7)</td>
</tr>
<tr>
<td>&gt;65</td>
<td>902,947 (18.6)</td>
<td>207,184 (20.5)</td>
</tr>
<tr>
<td><strong>Geographic and population description</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>16,640</td>
<td>3226</td>
</tr>
<tr>
<td>Population density (residents per square kilometer)</td>
<td>290</td>
<td>316</td>
</tr>
<tr>
<td>Adjusted household income (US $)</td>
<td>$44,357</td>
<td>$46,003</td>
</tr>
</tbody>
</table>

#### Paramedic service description

<table>
<thead>
<tr>
<th></th>
<th>East Midlands region, United Kingdom</th>
<th>Niagara and Hamilton regions, Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak number of vehicles staffed</td>
<td>496</td>
<td>69 (38 Niagara and 31 Hamilton)</td>
</tr>
<tr>
<td>Number of persons per vehicle</td>
<td>9809</td>
<td>14,776</td>
</tr>
<tr>
<td>Total number of crew staff</td>
<td>3200</td>
<td>745 (343 Niagara and 402 Hamilton)</td>
</tr>
</tbody>
</table>

During the study period, there were 182,947 attendances for mental health and overdose in the study regions (Table 2 and Table S1 in Multimedia Appendix 1), with an average of 1042.9 calls per week for mental health and 776.4 calls per week for overdose in the UK regions, and 232.5 calls per week for mental health and 139.1 calls per week for overdose in the Canadian regions. The interrupted time series analysis showed that in the time prior to the lockdown, attendances in the UK region for mental health had a small but statistically significant increasing rate across weeks, whereas the rate of these attendances was steady week-over-week in the Canadian regions. When the lockdown was initiated (the interruption in the time series) the rate of attendances for mental health in the UK region had a statistically significant increase and then remained stable with no statistically significant slope for the weeks during lockdown, while the Canadian regions saw no change when the lockdown was initiated (interruption) or in the slope during lockdown.
Table 2. Interrupted time series of weekly emergency medical service calls for mental health and overdose from January 2019 to July 2020, using negative binomial regression offset by the total population, for regions in the United Kingdom and Canada.a

<table>
<thead>
<tr>
<th>Outcome and variable</th>
<th>East Midlands region, United Kingdom</th>
<th>Niagara and Hamilton regions, Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mental health</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>87,086</td>
<td>19,410</td>
</tr>
<tr>
<td>Per week, IRRb (95% CI)</td>
<td>1.002 (1.002-1.003)c</td>
<td>0.999 (0.999-1.000)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>1.125 (1.031-1.227)</td>
<td>0.927 (0.802-1.073)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>0.998 (0.990-1.006)</td>
<td>1.003 (0.989-1.017)</td>
</tr>
<tr>
<td><strong>Overdose</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>64,832</td>
<td>11,619</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.002 (1.002-1.003)</td>
<td>1.002 (1.000-1.002)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>0.573 (0.518-0.635)</td>
<td>0.758 (0.615-0.936)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.031 (1.022-1.041)</td>
<td>1.015 (0.994-1.035)</td>
</tr>
</tbody>
</table>

aModels have been adjusted for seasonality (each month as a variable).
bIRR: incidence rate ratio.
cIRRs statistically significant at P <.05 are indicated in italics format.

Overdose attendances had a similar small but statistically significant weekly increase in both the UK and Canadian regions, and both had statistically significant drops in the rate of overdose attendances when the lockdown was initiated. However, while the UK study region rebounded to having a statistically significant positive slope in attendance rate after lockdown, the Canadian regions maintained this lower level of attendances after lockdown with a statistically nonsignificant slope.

Table 3 shows model estimates for the rate of mental health attendance by sex and age categories across the study weeks (also see Table S2 in Multimedia Appendix 1). For the Canadian regions, although the full sample showed no statistically significant changes in mental health attendances, this subgroup analysis demonstrates that there were substantial changes for male individuals. Specifically, at the time of lockdown (interruption), there was a statistically significant decrease in the rate of mental health attendances for male individuals, which then rebounded with a positive slope during the lockdown period. In contrast, both male and female individuals in the UK region had an increase in the rate of mental health attendances when the lockdown was initiated and then held steady at that new level.
Table 3. For each sex and age category strata, an interrupted time series of weekly emergency medical service calls for mental health from January 2019 to July 2020, using negative binomial regression offset by the total population of subgroup, for regions in the United Kingdom and in Canada.

<table>
<thead>
<tr>
<th>Strata and variable</th>
<th>East Midlands region, United Kingdom</th>
<th>Niagara and Hamilton regions, Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>50,098</td>
<td>10,829</td>
</tr>
<tr>
<td>Per week, IRR(^b) (95% CI)</td>
<td>1.002 (1.002-1.003)(^c)</td>
<td>0.999 (0.999-1.000)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>1.101 (1.007-1.203)</td>
<td>1.012 (0.856-1.196)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>0.998 (0.990-1.007)</td>
<td>0.992 (0.976-1.008)</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>36,515</td>
<td>8581</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.002 (1.001-1.003)</td>
<td>0.999 (0.999-1.000)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>1.163 (1.050-1.288)</td>
<td>0.830 (0.702-0.980)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>0.998 (0.989-1.008)</td>
<td>1.017 (1.001-1.033)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>4239</td>
<td>1572</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.003 (1.002-1.004)</td>
<td>0.999 (0.997-1.001)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>0.723 (0.579-0.913)</td>
<td>0.687 (0.451-1.038)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.021 (0.999-1.042)</td>
<td>1.024 (0.985-1.065)</td>
</tr>
<tr>
<td>18-44 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>39,648</td>
<td>10,382</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.003 (1.002-1.003)</td>
<td>1.000 (0.999-1.001)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>1.097 (0.986-1.220)</td>
<td>0.969 (0.824-1.139)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>0.994 (0.985-1.004)</td>
<td>0.997 (0.982-1.012)</td>
</tr>
<tr>
<td>45-65 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>22,970</td>
<td>8,722</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.002 (1.001-1.002)</td>
<td>0.999 (0.998-1.001)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>1.255 (1.123-1.403)</td>
<td>0.905 (0.723-1.313)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>0.991 (0.981-1.002)</td>
<td>1.012 (0.991-1.012)</td>
</tr>
<tr>
<td>&gt;65 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>20,229</td>
<td>2,007</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.001 (1.000-1.002)</td>
<td>0.999 (0.997-1.000)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>1.118 (1.005-1.244)</td>
<td>0.938 (0.721-1.215)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.011 (1.001-1.021)</td>
<td>1.002 (0.978-1.027)</td>
</tr>
</tbody>
</table>

\(^a\) Models have been adjusted for seasonality (each month as a variable).

\(^b\) IRR: incidence rate ratio.

\(^c\) IRRs statistically significant at \(P<.05\) are indicated in italics format.

While the Canadian regions showed no statistically significant changes in the rate of mental health attendances by age category, the UK region found that those younger than 18 years of age had a statistically significant decrease in mental health attendances when the lockdown was initiated, while those aged 45-65 years and 65 years and older had a statistically significant increase. In addition, for those aged 65 years and older, the rate of attendances for mental health continued to increase during the lockdown period.

Table 4 shows the attendances for overdose in each of the sex and age subgroups (also see Table S3 in Multimedia Appendix 1). Similar to what was observed for mental health attendances, in the Canadian Region, there was a statistically significant drop in attendances for overdose at the time the lockdown was
initiated; however, for this outcome, there was also a statistically significant drop for those aged 18-44 years. In the United Kingdom, a consistent statistically significant decrease in attendances for overdose was observed among all age and sex subgroups, followed by a statistically significant positive slope week-over-week for all subgroups in the period during lockdown.

Table 4. For each sex and age category strata, an interrupted time series of weekly emergency medical service calls for overdose from January 2019 to July 2020 using negative binomial regression offset by the total population, for regions in the United Kingdom and in Canada.\(^a\)

<table>
<thead>
<tr>
<th>Strata and variable</th>
<th>East Midlands region, United Kingdom</th>
<th>Niagara and Hamilton regions, Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>29,939</td>
<td>4646</td>
</tr>
<tr>
<td>Per week, IRR(^b) (95% CI)</td>
<td>1.003 (1.002-1.004) (^c)</td>
<td>1.002 (1.001-1.003)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>0.535 (0.474-0.604)</td>
<td>0.906 (0.723-1.134)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.036 (1.024-1.048)</td>
<td>1.005 (0.984-1.027)</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>34,516</td>
<td>6973</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.002 (1.001-1.002)</td>
<td>1.001 (1.000-1.002)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>0.612 (0.547-0.685)</td>
<td>0.674 (0.522-0.872)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.028 (1.017-1.039)</td>
<td>1.021 (0.996-1.046)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>5385</td>
<td>630</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.003 (1.002-1.004)</td>
<td>0.997 (0.994-0.999)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>0.456 (0.369-0.564)</td>
<td>0.845 (0.469-1.469)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.051 (1.031-1.071)</td>
<td>1.007 (0.954-1.006)</td>
</tr>
<tr>
<td>18-44 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>37,314</td>
<td>7645</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.002 (1.002-1.003)</td>
<td>1.002 (1.001-1.003)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>0.572 (0.505-0.647)</td>
<td>0.703 (0.560-0.882)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.030 (1.019-1.042)</td>
<td>1.013 (0.991-1.035)</td>
</tr>
<tr>
<td>45-65 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>17,313</td>
<td>2826</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.002 (1.002-1.003)</td>
<td>1.003 (1.000-1.004)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>0.583 (0.518-0.656)</td>
<td>0.883 (0.638-1.223)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.031 (1.020-1.042)</td>
<td>1.019 (0.988-1.051)</td>
</tr>
<tr>
<td>&gt;65 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attendances, n</td>
<td>4820</td>
<td>518</td>
</tr>
<tr>
<td>Per week, IRR (95% CI)</td>
<td>1.003 (1.002-1.004)</td>
<td>1.000 (0.997-1.003)</td>
</tr>
<tr>
<td>Lockdown, IRR (95% CI)</td>
<td>0.700 (0.577-0.848)</td>
<td>0.855 (0.467-1.511)</td>
</tr>
<tr>
<td>Lockdown trend, IRR (95% CI)</td>
<td>1.024 (1.007-1.042)</td>
<td>0.996 (0.929-1.053)</td>
</tr>
</tbody>
</table>

\(^a\)Models have been adjusted for seasonality (each month as a variable).
\(^b\)IRR: incidence rate ratio.
\(^c\)IRRs statistically significant at \(P<.05\) are indicated in italics format.
Discussion

Principal Findings

Our study found that there were statistically significant differences in the effects of lockdown on mental health and overdose attendances when comparing the United Kingdom and Canada. A web-based cross-sectional survey in the United Kingdom, 4 weeks into lockdown, found that 5% of participants were positive for a common mental health disorder [21]. In Canada, a cross-sectional survey found women of working age were most affected by a mental health condition, while 54% of Canadians reported good mental health [11]. While these surveys are helpful for understanding the self-reported mental health of populations, using EMS data as an indicator of mental health status among populations provides a powerful approach to understanding complex social changes in a novel way.

Both Canada and the United Kingdom experienced major societal changes during 2020 caused by the COVID-19 pandemic and took different approaches to addressing this event, which may explain observed differences in rates of mental health and overdose attendances for specific sex and age categories. The United Kingdom entered the first lockdown on March 23, 2020, while Ontario (Canada) had already reached this stage on March 15, 2020. Schools, colleges, and universities closed, and a large majority of the working population worked from home or was funded to not work (known as furlough in the United Kingdom). Hospitality services closed, and citizens were requested to remain at home. Primary care moved to web-based and telephone-based consultations, and general practitioners met patients in person only when necessary for examination. International borders were closed in Canada but remained largely open in the United Kingdom.

Financial benefits and emergency funding also differed between countries. The Canada Emergency Response Benefit paid a gross of CAD $2000 (equivalent to just under £1200 sterling or US $1480) per month for those who had lost their jobs or were unable to work. The United Kingdom relied on the standard benefit system, known as Universal Credit, for those who lost their jobs. Those still employed, but unable to work, were given 80% of their gross pay, up to a maximum of £2500 per month (equivalent to just over CAD $4200 or US $3100). Fiscal stimulus packages varied, with Canada investing 18.6% of gross domestic product and the United Kingdom 17.8% [22]. Finances are known to be a major cause of stress and mental health issues.

A measure of comparison between countries that actually allows us to compare the different approaches holistically is that of the stringency index. It is a measure of governmental and public health orders that took effect during the pandemic [4], allowing global comparisons due to a scoring system attributing values to components of lockdown (such as stay-at-home orders, business, and facility openings). A comparison of the United Kingdom’s and Canada’s stringency indices (shown in Multimedia Appendix 2) showed that, in fact, the United Kingdom and Canada experienced similar levels of lockdown, despite different governmental policies and other local public health unit mandates. The EMS data allow for health-related differences between countries that may be influenced by societal changes that have become evident, while contrasting comparisons of overall indices of lockdown stringency can concurrently be examined.

Mental Health and Overdose

This study found that prior to the pandemic lockdown, overall the rate of mental health presentations was holding constant in Canada, and there was a very small increase over time in the United Kingdom. When the lockdown was initiated, a difference in EMS attendance was seen between the countries; in the United Kingdom, nonoverdose attendance for mental health increased substantially, while in Canada, there was no change. These observed differences between Canada and the United Kingdom may, in fact, be due to the differences in the governmental lockdowns and orders. In particular, health care workers have experienced a high prevalence of anxiety, depression, and insomnia. This may be due to several factors, including fear of infection and the overwhelming influx of new information pertaining to caring for patients with COVID-19 [23-25].

Both countries also showed similar rises in rates of attendances for overdoses prior to lockdown. The immediate effect of the lockdown was similar, with a large reduction in attendances for overdose in both the United Kingdom and Canada, although the rate was increasing again during the lockdown period in the United Kingdom. The level of overdoses may be impacted by not only mental health stressors but also other societal changes resulting from the pandemic. Where the availability of drugs and alcohol becomes reduced and supplies dwindle due to a lack of prescribing or a lack of street drugs, overdose may become more likely in the subsequent period, as tolerance to these substances may have decreased.

Differences by Sex

In the UK study region, there were no sex differences for mental health and overdose attendances; both sexes had similar increases for mental health but decreases for overdose. However, in the Canadian study regions, there was a statistically significant decrease in both mental health and overdose attendances for male individuals, while there was no statistically significant change for female individuals. The reasons for these different trends among the sex strata within and between the UK and Canadian study regions are likely to be extremely complex and multifactorial, and therefore difficult to tease out in this paper without further study.

Differences by Age

There has been debate about the impact of lockdown on children and young adults [26]. Schools closed in both the United Kingdom and Canada, and school-age children were taught by parents or guardians. In the UK study region, the week-over-week rate of mental health attendances was increasing in a similar manner for all age categories in the period before lockdown, but at the time of lockdown, there was a substantial drop in the rate of attendances among the individuals younger than 18-years of age, whereas there was an increase among the individuals aged 45-65 years and those older than 65 years. In contrast, in the Canadian study regions, there were no statistically significant changes in the rate of mental health attendances for male individuals, while there was no statistically significant decrease in both mental health and overdose in both the United Kingdom and Canada, although the rate was increasing again during the lockdown period in the United Kingdom. The level of overdoses may be impacted by not only mental health stressors but also other societal changes resulting from the pandemic. Where the availability of drugs and alcohol becomes reduced and supplies dwindle due to a lack of prescribing or a lack of street drugs, overdose may become more likely in the subsequent period, as tolerance to these substances may have decreased.

Financial benefits and emergency funding also differed between countries. The Canada Emergency Response Benefit paid a gross of CAD $2000 (equivalent to just under £1200 sterling or US $1480) per month for those who had lost their jobs or were unable to work. The United Kingdom relied on the standard benefit system, known as Universal Credit, for those who lost their jobs. Those still employed, but unable to work, were given 80% of their gross pay, up to a maximum of £2500 per month (equivalent to just over CAD $4200 or US $3100). Fiscal stimulus packages varied, with Canada investing 18.6% of gross domestic product and the United Kingdom 17.8% [22]. Finances are known to be a major cause of stress and mental health issues.

A measure of comparison between countries that actually allows us to compare the different approaches holistically is that of the stringency index. It is a measure of governmental and public health orders that took effect during the pandemic [4], allowing global comparisons due to a scoring system attributing values to components of lockdown (such as stay-at-home orders, business, and facility openings). A comparison of the United Kingdom’s and Canada’s stringency indices (shown in Multimedia Appendix 2) showed that, in fact, the United Kingdom and Canada experienced similar levels of lockdown, despite different governmental policies and other local public health unit mandates. The EMS data allow for health-related differences between countries that may be influenced by societal changes that have become evident, while contrasting comparisons of overall indices of lockdown stringency can concurrently be examined.

Mental Health and Overdose

This study found that prior to the pandemic lockdown, overall the rate of mental health presentations was holding constant in Canada, and there was a very small increase over time in the United Kingdom. When the lockdown was initiated, a difference in EMS attendance was seen between the countries; in the United Kingdom, nonoverdose attendance for mental health increased substantially, while in Canada, there was no change. These observed differences between Canada and the United Kingdom may, in fact, be due to the differences in the governmental lockdowns and orders. In particular, health care workers have experienced a high prevalence of anxiety, depression, and insomnia. This may be due to several factors, including fear of infection and the overwhelming influx of new information pertaining to caring for patients with COVID-19 [23-25].

Both countries also showed similar rises in rates of attendances for overdoses prior to lockdown. The immediate effect of the lockdown was similar, with a large reduction in attendances for overdose in both the United Kingdom and Canada, although the rate was increasing again during the lockdown period in the United Kingdom. The level of overdoses may be impacted by not only mental health stressors but also other societal changes resulting from the pandemic. Where the availability of drugs and alcohol becomes reduced and supplies dwindle due to a lack of prescribing or a lack of street drugs, overdose may become more likely in the subsequent period, as tolerance to these substances may have decreased.

Differences by Sex

In the UK study region, there were no sex differences for mental health and overdose attendances; both sexes had similar increases for mental health but decreases for overdose. However, in the Canadian study regions, there was a statistically significant decrease in both mental health and overdose attendances for male individuals, while there was no statistically significant change for female individuals. The reasons for these different trends among the sex strata within and between the UK and Canadian study regions are likely to be extremely complex and multifactorial, and therefore difficult to tease out in this paper without further study.

Differences by Age

There has been debate about the impact of lockdown on children and young adults [26]. Schools closed in both the United Kingdom and Canada, and school-age children were taught by parents or guardians. In the UK study region, the week-over-week rate of mental health attendances was increasing in a similar manner for all age categories in the period before lockdown, but at the time of lockdown, there was a substantial drop in the rate of attendances among the individuals younger than 18-years of age, whereas there was an increase among the individuals aged 45-65 years and those older than 65 years. In contrast, in the Canadian study regions, there were no statistically significant changes in the rate of mental health attendances for male individuals, while there was no statistically significant decrease in both mental health and overdose in both the United Kingdom and Canada, although the rate was increasing again during the lockdown period in the United Kingdom. The level of overdoses may be impacted by not only mental health stressors but also other societal changes resulting from the pandemic. Where the availability of drugs and alcohol becomes reduced and supplies dwindle due to a lack of prescribing or a lack of street drugs, overdose may become more likely in the subsequent period, as tolerance to these substances may have decreased.
attendances at the time the lockdown was initiated for any of the age category strata. For rates of overdose attendances, there was no noticeable effect demonstrated by different age group strata in Canada or the United Kingdom.

In the United Kingdom, examinations were canceled for the General Certificate of Secondary Education (16-year-olds) and A-levels (18-year-olds), which may have reduced stress. In addition, there may have been decreased stress due to learning at home, without the need to attend school in person, with less exposure to school-related stressors. This is supported by existing literature describing how adolescents with existing mental health problems pre-COVID-19 had less during the pandemic [27]. In Canada, the education system has no equivalent to the General Certificate of Secondary Education and A-level examinations; therefore, pandemic schooling-related effects would have been different and may not have been as pronounced, resulting in no statistically significant mental health changes pre-COVID-19 compared to during the pandemic.

In both the United Kingdom and Canada, there was considerable distress caused by the volume of vulnerable people in care homes and the speed of government responses. This was seen by the number of deaths among care-home residents [28] over the period from March 20, 2020, to January 15, 2021. In the United Kingdom, 33% (n=30,851) of COVID-19–related deaths were among people residing in care homes. In Canada, long-term care home residents accounted for 59% (n=11,114) of COVID-19–related fatalities. As a result of these problems, care home residents were subject to significant restrictions, leading to reduced access to medical care and isolation-related stress. Although it could be expected that the older age groups (45-55 years and 65 years and older) would worry about income, poor COVID-19 outcomes due to higher rates of chronic diseases [29], and about family and parents in care homes (United Kingdom) or long-term care facilities (Canada), these age groups had a statistically significant increase in mental health–related calls in the United Kingdom but not in Canada. Other literature from the United Kingdom demonstrates worse mental health in age groups younger than 35 years; however, the data were from a cross-sectional survey during lockdown [21]. These differences might be explained by behavioral differences in the reasons for which it was deemed appropriate to call EMS in each country, by less stress overall in one country versus the other, or by differences in the data collection method used to determine mental health problems (calls to EMS vs self-reported survey). However, further research is needed to understand these differences, which are beyond the scope of this paper.

Strengths and Limitations
This study compares routinely collected observational data from 2 Western countries. Data were collected primarily for routine clinical recording, monitoring, and clinical auditing rather than for epidemiological research. These data have nevertheless allowed a careful examination of the ambulance response to mental health and overdose emergencies experienced by the populations, using data that are not self-reported. It should be noted that responses to mental health in both the United Kingdom and Canada tend to have similar patient pathways. Both countries have a national health system, in which health care is freely available to all; therefore, the pathways open to patients are to call 911/999 to visit a primary care doctor or facility where they can receive emergency or urgent care.

There will be confounders that may explain some of the predictions made by the statistical models. Age and sex were adjusted for, but data were not available to adjust for ethnicity or socioeconomic status. In the United Kingdom and Canada, there have been clear differences in the rates of COVID-19 between different ethnic groups and levels of affluence. These differences are also likely to contribute to differences in rates of mental illness and requirements for mental health support.

There were differences in the data collected between each country, and efforts were made to adjust for these underlying coding differences. Intentionality was not used to define overdose. Clinical impression (ambulance provisional diagnosis) codes were not entirely comparable between the 2 countries. In Canada, impression code 81 (drug or alcohol overdose) is captured by the 4 most relevant codes in the United Kingdom. There will be cases missed and documented as “other medical problem” or “acute behavioral disturbance,” for example, but that broadens the category too wide. The agreement was achieved by extracting comparable codes through discussion.

Conclusions
The EMS attendance data have demonstrated varying impacts of the COVID-19 pandemic on mental health and overdose between the study regions in the United Kingdom and Canada. While the countries implemented similar degrees of lockdowns and public health measures, as seen with the stringency index, there were differences in the specific measures taken; these differences may explain the divergence in mental health attendances between the regions. In contrast, overdose attendances followed similar patterns between the study regions. Future research is needed to explore the mechanisms behind these observed trends in EMS attendances following the lockdown and to examine whether the countries return to having similar rates of attendances over time as they did prior to the lockdown.

Acknowledgments
The authors would like to thank the emergency services in the study regions for accessing their data. The UK Canada Emergency Calls Data Analysis and GEospatial Mapping (EDGE) Consortium members are GA, ANS, Ricardo Angeles, Iwona Bielska, Jasdeep Brar, RC, RF, Mark Gussy, Bartholomew Hill, Kamlesh Khunti, GL, BM, Harriet Moore, MP, RS, and Frank Tanser.

https://publichealth.jmir.org/2024/1/e46029
Data Availability
The data sets analyzed during this study are not publicly available due to restrictions by the paramedic and ambulance services but are available from the corresponding author on reasonable request.

Authors’ Contributions
GL and GA are the guarantors of this work and, as such, have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. RS, RF, BM, and ANS enabled the download of the data. RC, MP, and GL analyzed the data. GA, RF, and ANS oversaw data analysis and interpretation. GL, RC, MP, and GA wrote the first draft of the manuscript, which all authors critically reviewed thereafter.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Results tables.
[PDF File (Adobe PDF File), 75 KB - publichealth_v10i1e46029_app1.pdf]

Multimedia Appendix 2
Stringency index for the 2 study regions from January to July 2020.
[PDF File (Adobe PDF File), 38 KB - publichealth_v10i1e46029_app2.pdf]

References

Abbreviations
EMS: emergency medical service
IRR: incidence rate ratio
© Graham Law, Rhiannon Cooper, Melissa Pirrie, Richard Ferron, Brent McLeod, Robert Spaight, A Niroshan Siriwardena, Gina Agarwal, Gina UK Canada Emergency Calls Data Analysis and GEospatial Mapping (EDGE) Consortium. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 10.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Improving Parental Health Literacy in Primary Caregivers of 0- to 3-Year-Old Children Through a WeChat Official Account: Cluster Randomized Controlled Trial

Yun Li¹,*, MM, MD; Qiuli Xiao²,*, MD; Min Chen¹, MD; Chunhua Jiang¹, MM, MD; Shurong Kang¹, MD; Ying Zhang¹, MD; Jun Huang¹, MD; Yulin Yang³, MD; Mu Li⁵, MD, PhD; Hong Jiang², MD, PhD

Abstract

Background: Parental health literacy is important to children’s health and development, especially in the first 3 years. However, few studies have explored effective intervention strategies to improve parental literacy.

Objective: This study aimed to determine the effects of a WeChat official account (WOA)–based intervention on parental health literacy of primary caregivers of children aged 0-3 years.

Methods: This cluster randomized controlled trial enrolled 1332 caregiver-child dyads from all 13 community health centers (CHCs) in Minhang District, Shanghai, China, between April 2020 and April 2021. Participants in intervention CHCs received purposefully designed videos via a WOA, which automatically recorded the times of watching for each participant, supplemented with reading materials from other trusted web-based sources. The contents of the videos were constructed in accordance with the comprehensive parental health literacy model of WHO (World Health Organization)/Europe (WHO/Europe). Participants in control CHCs received printed materials similar to the intervention group. All the participants were followed up for 9 months. Both groups could access routine child health services as usual during follow-up. The primary outcome was parental health literacy measured by a validated instrument, the Chinese Parental Health Literacy Questionnaire (CPLHQ) of children aged 0-3 years. Secondary outcomes included parenting behaviors and children’s health outcomes. We used the generalized linear mixed model (GLMM) for data analyses and performed different subgroup analyses. The β coefficient, risk ratio (RR), and their 95% CI were used to assess the intervention’s effect.

Results: After the 9-month intervention, 69.4% (518/746) of caregivers had watched at least 1 video. Participants in the intervention group had higher CPLHQ total scores (β=2.51, 95% CI 0.12-4.91) and higher psychological scores (β=1.63, 95% CI 0.16-3.10) than those in the control group. The intervention group also reported a higher rate of exclusive breastfeeding (EBF) at 6 months (38.9% vs 23.44%; RR 1.90, 95% CI 1.07-3.38) and a higher awareness rate of vitamin D supplementation for infants younger than 6 months (76.7% vs 70.5%; RR 1.39, 95% CI 1.06-1.82). No significant effects were detected for the physical score on the CPLHQ, breastfeeding rate, routine checkup rate, and children’s health outcomes. Furthermore, despite slight subgroup differences in the intervention’s effects on the total CPLHQ score and EBF rate, no interaction effect was observed between these subgroup factors and intervention factors.

Conclusions: Using a WHO literacy model–based health intervention through a WOA has the potential of improving parental health literacy and EBF rates at 6 months. However, innovative strategies and evidence-based content are required to engage more participants and achieve better intervention outcomes.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000031711; https://www.chictr.org.cn/showproj.aspx?proj=51740

(JMIR Public Health Surveill 2024;10:e54623) doi:10.2196/54623

KEYWORDS
health literacy; WeChat; cluster randomized controlled trial; RCT; randomized; controlled trial; controlled trials; parental; parenting; parents; parent; China; Chinese; mHealth; mobile health; app; apps; applications; pediatric; pediatrics; paediatric; paediatrics; infant; infants; infancy; baby; babies; neonate; neonates; neonatal; newborn; newborns; toddler; toddlers
### Introduction

A child’s health and well-being are critical in the first 3 years of his or her life, over the short and long term [1,2]. During this period, a caregiver’s health literacy is crucial in the child’s development and health [3]. WHO (World Health Organization)/Europe (WHO/Europe) has proposed a 12D model that identifies 4 competencies for health literacy. These competencies include accessing, understanding, appraising, and applying health information across 3 domains: health care, disease prevention, and health promotion [4]. Studies have shown that caregivers with low health literacy were more likely to engage in risky parenting behaviors, such as not practicing breastfeeding (BF), not administering medications to children as prescribed, and allowing children long screen time [5-7]. These can lead to poor health outcomes for their children, such as poor disease management and high emergency service usage [8-10]. Despite this, few studies have evaluated the effectiveness of interventions aiming to improve parental health literacy.

The 2020 Chinese Health Literacy Surveillance Report revealed that only 35.9% of Chinese residents had basic health literacy [11]. Health professionals in China were confronted with a limited health workforce and high demand for childcare services [12]. In-person communication has also been restricted due to the COVID-19 pandemic, making it challenging for caregivers to access reliable health information [13]. Therefore, there is an urgent need for feasible and effective alternative approaches to deliver health promotion and support for caregivers of young children.

Social media has become a popular platform for individuals to obtain and share health-related information in recent years [14-18]. WeChat, one of China’s largest social media platforms, had over 1.3 billion monthly active users worldwide by the end of 2023 [19]. WeChat offers several innovative function modules, including WeChat official accounts (WOAs), which enable users to access information, services, and subscriber interaction [20]. A recent national survey showed that one-third of respondents regularly accessed health information via WeChat and nearly two-thirds considered WOAs a feasible medium for accessing health education material [21]. Studies have shown that using WeChat or WOAs for health intervention could significantly improve satisfaction, accessibility, and convenience compared to traditional health education [22,23]. By integrating diverse forms of information, including text, images, videos, and links to digital books and web pages, WeChat and WOAs have the potential to enhance parental health literacy. WeChat’s health education campaign for pregnant women has been proven effective in promoting exclusive breastfeeding (EBF) [24]. A WeChat-based family health education program also significantly raised parents’ awareness of children’s illness prevention [25]. However, most existing studies have only addressed limited aspects of health literacy. Our WOA intervention is grounded in the WHO/Europe comprehensive health literacy model. It aims to enhance parental health literacy, and its effectiveness is evaluated through a cluster randomized controlled trial among caregivers of children aged 0-3 years in an urban setting in China.

### Methods

#### Study Design

According to a prespecified study protocol, a cluster randomized controlled trial was conducted from April 7, 2020, to April 20, 2021, in Minhang District, Shanghai, China [26]. All 13 Minhang District community health centers (CHCs) were enrolled and randomly allocated to the intervention or control group through random sequence generation, with 1332 eligible caregiver-child dyads participating in the trial. The CHCs were responsible for providing child health care services in their catchments for children from birth to 3 years of age.

#### Participants

Since we planned a 9-month intervention, we recruited primary caregivers of children aged 0-3 years. The primary caregiver herein refers to the person who provided primary care to the child, including the child’s mother, father, grandmother, grandfather, and other caregivers including nannies. One caregiver at most was recruited within each family. In case of families with multiple children, the participating caregiver would designate only one of them to be involved in this study and provide their unique health record number.

When caregiver-child dyads attended child health care in CHCs, the CHCs’ staff approached and informed them about the study. Caregivers were eligible if they were adults (older than 18 years) and primary caregivers of children aged 0-3 years, completed at least the third grade of primary school, owned a smartphone and an individual WeChat account, were able to communicate verbally or in writing, and planned to stay in the recruitment area for at least a year. Immediately after recruitment, all participating caregivers were invited to subscribe to the designed WOA for data collection in both groups and intervention delivery only for the intervention group.

#### Blinding

Due to the nature of the intervention, service providers and participants were not blinded to the group allocation. Immediately after baseline data collection, group allocation was revealed to the participants.

#### Interventions

##### Contents of the WOA-Based Intervention

Contents of the intervention videos were developed in accordance with 12 subdimensions of health literacy by WHO/Europe [4]. Through literature review and expert consultation, 15 key topics about children’s health and development (10 physical and 5 psychological) were generated and categorized into 3 domains, including health care, disease prevention, and health promotion. Each key topic was designed to improve caregivers’ 4 types of capacity in accessing, understanding, appraising, and applying child health information. The detailed logic model for this WOA-based intervention program has been prespecified in our protocol [26].

##### The WOA Platform

The Scientific Parenting platform, specifically designed for this WOA-based intervention, was integrated into the WeChat app.
Participants were invited to subscribe and register upon providing informed consent. The registration involved entering their child’s unique health record number, date of birth, relationship with the child, and catchment CHCs. Once participants logged in, the platform automatically assigned them to different functions based on their catchment CHCs. All participants had access to the web-based survey module, while only those in the intervention CHCs could see and access the intervention module. Two electronic questionnaires were automatically delivered to each participant via the web-based survey module, one at recruitment and the other 9 months later. Figure 1 shows a detailed illustration of the participants’ journey through the Scientific Parenting platform.

**Figure 1.** Participants’ journey through the scientific parenting platform of the WeChat official account (WOA)–based intervention. BF: breastfeeding; CHC: community health center; CPHLQ: Chinese Parental Health Literacy Questionnaire; EBF: exclusive breastfeeding; VD: vitamin D.

**Intervention Group**

The intervention package was provided in the intervention module, consisting of videos developed on the basis of the previously mentioned WHO/Europe health literacy model. The videos included 4 video clips of experts’ talks between 13 and 33 minutes and 6 animated video clips of 1 to 5 minutes, totaling 2 hours. During the 9-month intervention, participants could self-navigate, select topics that they were interested in, and decide the order and pace of material they read or watch. Participants could watch the videos repeatedly, and the number of times the videos were watched was recorded automatically.
by the platform in each user’s account. Additionally, we provided participants with reading material from other trusted web-based sources as supplemented material to videos. Participants could access this web-based reading material by clicking on links, such as e-books, websites, and other WOAs in public. Furthermore, the contents were closely related to key topics of the videos, and the web-based reading material also covered other topics such as early education, healthy home environment, etc.

**Control Group**

Participants in the control group received printed educational material as an extension to routine child health care when they took their children for health checkups. The contents of the printed material resembled those of the videos delivered to the intervention group. The printed educational material is also detailed in our protocol [26].

Both groups received routine child health care services, including weight and length measurements, self-reported child feeding practice and dietary intake, and checkups of children’s health conditions [27].

**Outcome Measures**

The primary outcome measured was the change in total scores of parental health literacy (physical and psychological) using a validated questionnaire [28,29]. Secondary outcomes included (1) parenting behaviors, such as the EBF rate at 6 months, any BF rate at 12 months, the consumption rate of iron-fortified staple foods at 6 months, and the rate of routine checkup of the child’s vision and oral health; (2) the proportion of participants’ awareness of vitamin D (VD) supplementation for infants aged 0 to 6 months; (3) children’s anthropometric measures, including weight-for-length z score, length-for-age z score, weight-for-age z score, BMI-for-age z score, and head circumference-for-age z scores; and (4) children’s health outcomes, such as the incidence of anemia, unintended injuries, diarrhea, obesity, and hospitalization.

**Data Collection**

All participants completed the baseline survey through the web-based survey module on the WOA platform. The final survey will automatically be made available to the participants 9 months after the baseline survey.

In the baseline survey, we collected demographics and health-related variables, including participants’ relationship with their children, education level, source of parental information, family income, number of children in the family, children’s age, gender, and Hukou (the Chinese official permanent residency registration by location, which is closely linked with social welfare and administration [26,27]).

Parental health literacy was measured using a validated questionnaire, the Chinese Parental Health Literacy Questionnaire (CPhLQ) of children aged 0-3 years, at baseline and the end line [28,29]. The CPhLQ was designed by the research group based on the health literacy model proposed by WHO/Europe [4]. This questionnaire was validated among primary caregivers of children aged 0-3 years across different regions in China [28,30]. It consists of a 39-item subscale for physical health literacy and a 35-item subscale for psychological health literacy. Both the physical and psychological subscales were scaled on a score of 0-100, of which the total maximum score was 200.

In the final survey, we collected data on awareness of VD supplementation, parenting behaviors, including attendance in routine checkups for the children’s vision and oral health, and children’s health outcomes, such as unintended injury, diarrhea, and hospitalization. For awareness of VD supplementation, we asked the participants if they thought that administering VD supplements to their children after birth was necessary. For attendance in routine checkups, we asked if they had taken their children for regular visual and oral health checkups during the intervention. For children’s health outcomes, we asked about the occurrence of unintended injury, diarrhea, and hospitalization in the past 9 months. Participants’ engagement in the intervention, the frequency of video watching by participants, was automatically tracked by a function embedded in the WOA platform. However, due to administration rights, we could not track participants’ accessing links to reading materials on other web-based platforms.

Children living in the project district were required to take routine health checkups at specific ages: 1, 2, 4, 6, 9, 12, 18, 24, 30, and 36 months. During these checkups, child health care providers at CHCs would conduct face-to-face interviews to collect data on various parameters such as EBF, BF, introduction of solid foods, anthropometric measures, and health conditions such as obesity and anemia [31]. Project researchers would later extract all relevant health checkups’ data from baseline to 1 month after the 9-month intervention. Children younger than 6 months who were exclusively breastfed at baseline were eligible for the EBF analysis. Similarly, children younger than 6 months who had not been introduced to solid foods at baseline were eligible for the analysis of supplementation of iron-fortified staple foods at 6 months. Children younger than 12 months who were still breastfed at baseline were eligible for the BF analysis.

For the analysis of anthropometric measures and health conditions such as obesity and anemia, health checkup data closest to the final web-based evaluation survey were used, and baseline values were included as covariates.

**Sample Size Calculation**

We assumed an intervention effect size of 15 points for total scores on the CPhLQ, a coefficient of variation of 0.2, and an intracluster correlation coefficient of 0.05 [28]. A total of 1183 caregivers from 13 CHCs (clusters), with a mean cluster size of 91 per CHC, was estimated to detect the assumed effect size with 90% power at a 5% significance level. We also considered an approximately 20% rate of loss to follow-up.

**Statistical Analysis**

We conducted descriptive statistical analyses for all outcomes, presenting continuous variables as mean (SD) or median (IQR) values and categorical variables as proportions (%). To compare differences between the 2 groups, we used independent t tests for continuous variables and chi-square tests for categorical variables. Anthropometric indicators were calculated using the WHO Anthro (version 3.2.2) [32]. To compare within-group
differences (mean change from baseline) between the baseline and at the end of 9 months of the intervention, we used the paired \( t \) test.

To assess the intervention’s effects, we performed intention-to-treat analyses and generalized linear mixed models (GLMMs), as recommended for cluster randomized trials [33]. We used multiple imputations to handle missing values at the end of the trial. The GLMMs included CHC-level random intercepts to account for the correlation due to the clustering of participants within CHCs. For each outcome analysis, the model was adjusted for demographic characteristics. Additionally, the baseline CPHLQ score, anthropometric \( z \) score, and status of obesity and anemia were included in the model for evaluation of the intervention effects. By controlling the correlation between baseline and follow-up values, we avoided the need for time or a time × group interaction term to interpret intervention effects [34]. Given the relatively small number of clusters, Satterthwaite correction was used in GLMM to maintain an appropriate type I error [35]. Satterthwaite correction allowed us to estimate the degree of freedom on the basis of the residual variances at different variable levels [36].

We reported the intervention effect using \( \beta \) (95% CI) and \( P \) values for continuous outcomes. We used risk ratios (RRs), their 95% CIs, and \( P \) values to describe the intervention effect for categorical outcomes.

To identify any potential modifications, we conducted subgroup and interaction analyses for the total score on the CPHLQ and the EBF rate at 6 months. Participants were stratified into diverse subgroups based on their relationship with the child (mothers vs others), the child’s Hukou (Shanghai vs others), whether they had 1 child or more, the participant’s education level (university or higher vs less than university education), and source of parental information access (social media vs others). For the total score on the CPHLQ, we presented the \( \beta \) (95% CI) of GLMMs within each subgroup and the \( P \) value for interaction items of GLMMs in the overall group. For the EBF rate at 6 months, we presented the RR (95% CI) of GLMMs within each subgroup. Additionally, we used the multiplicative scale for multiplicative interaction and the relative excess risk due to interaction, attributable proportion due to interaction, and synergy index for additive interaction in the overall GLMMs.

Results with a type I error rate of \( P < 0.05 \) in 2-sided tests were considered statistically significant. Statistical analysis was performed using R statistical software (version 4.1.3; The R Foundation).

Ethical Considerations
The trial was approved by the ethics committee of Shanghai Minhang District Maternal and Child Health Hospital (approval number #2020-KS-01) and registered with the Chinese Clinical Trial Registry (#ChiCTR2000031711). All trial participants provided their written informed consent during the recruitment process.

Results
Study Participation
A total of 1332 caregiver-child dyads from 13 CHCs participated in the study. Six CHCs containing 746 dyads were randomly assigned to the intervention group, and 7 CHCs containing 586 dyads to the control group (Figure 2).
According to the baseline information, among 1332 participants, 81.7% (1088/1332) of primary participants were mothers and 82.9% (1104/1332) of participants had college or above education. The majority (92.7%, 1235/1332) of participants accessed parenting information from other social media sites (Table 1). Among all children aged 0-3 years, 71.2% (948/1332) of them were the only children in the family, 48.4% (645/1332) of them were boys, and 62.0% (826/1332) of them were in the Shanghai Hukou (Table 1).
Table. Baseline characteristics of the participants in the WeChat official account (WOA)–based intervention (April 7, 2020, to April 20, 2021, in Shanghai).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control (n=586)</th>
<th>Intervention (n=746)</th>
<th>Total (N=1332)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child’s age (months), mean (SD)</td>
<td>8.97 (6.77)</td>
<td>8.56 (7.17)</td>
<td>8.76 (7.06)</td>
<td>.28</td>
</tr>
<tr>
<td>Child’s gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.83</td>
</tr>
<tr>
<td>Male</td>
<td>284 (48.5)</td>
<td>361 (48.4)</td>
<td>645 (48.4)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>302 (51.5)</td>
<td>385 (51.6)</td>
<td>687 (51.6)</td>
<td></td>
</tr>
<tr>
<td>Relationship with the child, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.051</td>
</tr>
<tr>
<td>Mother</td>
<td>465 (79.4)</td>
<td>623 (83.5)</td>
<td>1088 (81.7)</td>
<td></td>
</tr>
<tr>
<td>Father or others</td>
<td>121 (20.6)</td>
<td>123 (16.5)</td>
<td>244 (18.3)</td>
<td></td>
</tr>
<tr>
<td>Child’s Hukou, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Shanghai</td>
<td>414 (70.6)</td>
<td>412 (55.2)</td>
<td>826 (62.0)</td>
<td></td>
</tr>
<tr>
<td>Other provinces</td>
<td>172 (29.4)</td>
<td>334 (44.8)</td>
<td>506 (38.0)</td>
<td></td>
</tr>
<tr>
<td>One-child or not, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Yes</td>
<td>432 (73.7)</td>
<td>516 (69.2)</td>
<td>948 (71.2)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>154 (26.3)</td>
<td>230 (30.8)</td>
<td>384 (28.8)</td>
<td></td>
</tr>
<tr>
<td>Participants’ education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Junior school and below</td>
<td>20 (3.4)</td>
<td>62 (8.3)</td>
<td>82 (6.2)</td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>57 (9.7)</td>
<td>89 (11.9)</td>
<td>146 (11.0)</td>
<td></td>
</tr>
<tr>
<td>University or higher</td>
<td>509 (86.9)</td>
<td>595 (79.8)</td>
<td>1104 (82.9)</td>
<td></td>
</tr>
<tr>
<td>Family monthly income per capita (in RMB&lt;sup&gt;d&lt;/sup&gt;)&lt;sub&gt;1&lt;/sub&gt;, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.75</td>
</tr>
<tr>
<td>&lt;4500</td>
<td>49 (8.4)</td>
<td>73 (9.8)</td>
<td>122 (9.2)</td>
<td></td>
</tr>
<tr>
<td>4500-7500</td>
<td>126 (21.5)</td>
<td>161 (21.6)</td>
<td>287 (21.5)</td>
<td></td>
</tr>
<tr>
<td>7500-12,500</td>
<td>186 (31.7)</td>
<td>191 (25.6)</td>
<td>377 (28.3)</td>
<td></td>
</tr>
<tr>
<td>≥12,500</td>
<td>171 (29.2)</td>
<td>243 (32.6)</td>
<td>414 (31.2)</td>
<td></td>
</tr>
<tr>
<td>I do not know</td>
<td>54 (9.2)</td>
<td>78 (10.5)</td>
<td>132 (9.9)</td>
<td></td>
</tr>
<tr>
<td>Source of parental information, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Social media</td>
<td>553 (94.4)</td>
<td>682 (91.4)</td>
<td>1235 (92.7)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>33 (5.6)</td>
<td>64 (8.6)</td>
<td>97 (7.3)</td>
<td></td>
</tr>
<tr>
<td>Feeding mode, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.47</td>
</tr>
<tr>
<td>Exclusive breastfeeding</td>
<td>152 (26.0)</td>
<td>185 (24.9)</td>
<td>337 (25.4)</td>
<td></td>
</tr>
<tr>
<td>Mixed feeding</td>
<td>183 (31.3)</td>
<td>256 (34.5)</td>
<td>441 (33.2)</td>
<td></td>
</tr>
<tr>
<td>Artificial feeding</td>
<td>248 (42.7)</td>
<td>302 (40.6)</td>
<td>550 (41.4)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>P values for comparing the intervention and control groups.

<sup>b</sup>Not applicable.

<sup>c</sup>Italicized values are significant at P<.05.

<sup>d</sup>1 RMB=US $0.14.

There was no significant difference between the intervention and the control group regarding participants’ relationship with their child, the child’s age and gender, the number of children in the family, family monthly income, and child feeding mode (P>.05). However, compared to the control group, fewer participants in the intervention group had a university or higher level of education, children with Shanghai Hukou, and received information from other social media sites (P<.05).

Participants’ Adherence to the Intervention Program

During the 9-month intervention program, the WOA platform automatically tracked the frequency of participants watching the videos. Of 746 participants in the intervention group, 69.4%...
had watched at least 1 video. Participants who watched the videos tended to have younger children (8.12 vs 9.54 months, \(P=.01\)), not have a Shanghai Hukou (49.8% vs 33.3%, \(P<.001\)), be the children’s mothers (85.9% vs 78.1%, \(P=.01\)), have an education below university level (23.0% vs 14.0%, \(P=.007\)), and remain in the study (83.4% vs 76.8%, \(P=.04\)), compared to those who never watched any video during the intervention (Table S1 in Multimedia Appendix 1).

At each video level, less than 50% (range 14.9%-46.5%) of participants in the intervention group had ever watched the video. The 3 videos with the highest watching rate were scientific feeding guidance (46.5%), childhood pneumonia identification (41.7%), and childhood obesity prevention (41.6%). The average number of times viewing each video clip ranged from 0.14 to 1.35. The most frequently viewed video clips were scientific feeding guidance (1.35 times), unintentional injury prevention (0.91 times), and essential information on health care for children aged 0 to 3 years (0.81 times).

**Retention of the Study Participants**

At the end of 9 months, the retention rates of the intervention and control groups did not differ significantly (81.4% vs 81.2%, \(P=.95\)). Compared to participants enrolled at baseline, those lost to follow-up tended to be the child’s father or other caregivers (18.3% vs 30.1%, \(P<.001\)). Compared to participants who remained in the study, they also tended to be child’s father or other caregivers (15.6% vs 30.1%, \(P<.001\)) and have a child without a Shanghai Hukou (36.7% vs 43.8%, \(P=.04\); Table S2 in Multimedia Appendix 1). Please refer to Multimedia Appendix 2 for a detailed missing data report of this study.

**Intervention Effect**

**Parental Health Literacy**

The results of parental health literacy are presented in Figure 3. At the final evaluation, both the intervention and control groups showed increased scores on the CPHLQ compared to baseline. The intervention group had increased total, physical, and psychological part scores on the CPHLQ by 10.44, 5.86, and 4.01 points, the control group had an increase by 3.96, 3.82, and 0.15 points, respectively, compared to baseline. Using the GLMM and controlling for the baseline score and other potential confounders, we observed a significant intervention effect on the total CPHLQ scores (\(\beta=2.51, 95\%\ CI 0.12-4.91\)) and psychological scores (\(\beta=1.63, 95\%\ CI 0.16-3.10\)), while there was no significant difference in the physical parental health literacy score.
**Figure 3.** Effect of the WeChat official account (WOA)–based intervention (April 7, 2020, to April 20, 2021, in Shanghai) on the Chinese Parental Health Literacy Questionnaire (CPHLQ) and children’s anthropometric measures. \(^a\) \(P\) values describing the changes from baseline to follow-up in the intervention or control group. \(^b\)The results of the generalized linear mixed model (GLMM) with the group as a fixed factor, community health centers (CHCs) as the random intercept, follow-up values as outcomes, and baseline values, caregiver’s relationship with the child, education level, the child’s gender, Hukou, whether they are the only child or not, and source of parental information access as covariates. BAZ: BMI-for-age \(z\) score; HAZ: length-for-age \(z\) score; HCZ: head circumference–for-age \(z\) score; WAZ: weight-for-age \(z\) score; WHZ: weight-for-length \(z\) score. \(*P < .05; \***P < .001.\)

We analyzed the effect of watching at least 1 intervention video on CPHLQ scores (Figures 4 and 5). Within the intervention group, both participants who had not watched videos and those who watched at least 1 video demonstrated an increased total score. Using the GLMM, compared to participants who watched at least 1 video, those in the control group had a significantly lower total score (\(\beta=2.68, 95\% \text{ CI} 0.20-5.17\)) and psychological score (\(\beta=1.72, 95\% \text{ CI} 0.13-3.32\)), after controlling for the baseline score and other potential confounders.
Figure 4. Comparisons of total score on the Chinese Parental Health Literacy Questionnaire (CPLHMQ) among participants with different adherent statuses in the WeChat official account (WOA)–based intervention (April 7, 2020, to April 20, 2021, in Shanghai). \( ^a \)P values describing differences in the total score from baseline to follow-up using the paired t test. \( ^b \)P values describing the difference in change in the total score between participants with different adherent statuses using the t test. Participants with different adherent statuses: ① those in the intervention group who had never watched videos; ② those in the intervention group who had watched at least 1 video; and ③ those in the control group. ***P≤.001.
Despite slight differences in the intervention’s effects on the total score on the CPHLQ among different subgroups, no interaction effect was observed between these subgroup factors and the intervention (Figure 6).
Parenting Behaviors

The effect of the intervention on parenting behaviors is illustrated in Figure 7. According to the GLMMs controlled for potential confounders, the intervention group demonstrated a significantly higher EBF rate at 6 months than the control group (38.9% vs 23.44%; RR 1.90, 95% CI 1.07-3.38). Additionally, there were no significant between-group differences in the BF rate at 12 months, iron-fortified staple food supplementation at 6 months, and routine checkup of the child’s vision and oral health.
**Figure 7.** Effect of the WeChat official account (WOA)–based intervention (April 7, 2020, to April 20, 2021, in Shanghai) on parenting behaviors, awareness, and children’s health outcomes. The risk ratio (RR) and 95% CI values describe the results of the generalized linear mixed model (GLMM) with the group as a fixed factor, community health centers (CHCs) as the random intercept, follow-up values as outcomes, and baseline values for caregiver’s relationship with their children, education level, child’s gender, Hukou, whether they are an only child or not, and source of parental information access as covariates. BF: breastfeeding; EBF: exclusive breastfeeding; VD: vitamin D. *0.01<P≤.05; **0.001<P≤.01; ***P≤.001.

The results of subgroup analyses on EBF at 6 months are presented in Figure 8. Neither multiplicative nor additive interaction effects were observed between the subgroup factors and the intervention.
Figure 8. Subgroup analyses of the effects of the WeChat official account (WOA)-based intervention on participants’ exclusive breastfeeding (EBF) rate at 6 months. The risk ratio (RR) and 95% CI values describe the results of the generalized linear mixed model (GLMM) within each subgroup with the group as a fixed factor, community health centers (CHCs) as the random intercept, follow-up values as outcomes, and baseline values of caregiver’s relationship with the children, education level, child’s gender, Hukou, whether they are an only child or not, and source of parental information as covariates. The multiplicative scale described the multiplicative interaction between subgroup factors and intervention, while the relative excess risk due to the interaction (RERI), attributable proportion due to the interaction (AP), and synergy index (SI) describe the additive interaction. * $0.01 \leq P \leq 0.05$.

Awareness of VD Supplementation After Birth

The results of VD supplementation awareness are presented in Figure 7. After controlling for potential confounders, the GLMM showed that the intervention group had a significantly higher awareness rate of VD supplementation after birth than the control group (76.7% vs 70.5%; RR 1.39, 95% CI 1.06-1.82).

Children’s Anthropometric and Health Outcomes

The results of children’s anthropometric and health outcomes are respectively presented in Figures 3 and 7. Using the adjusted GLMM, no significant intervention effects were observed in children’s anthropometric and health outcomes between the intervention and control groups.
Discussion

Principal Findings

We developed a digital parenting health literacy intervention of purposefully designed videos, supplemented with reading material from other web-based platforms and evaluated the effectiveness of the intervention. The rate of viewing at least 1 intervention video was nearly 70%. The 9-month WOA-based intervention significantly improved the health literacy score among caregivers of 0– to 3-year-old children, EBF rate at 6 months, and their awareness of VD supplementation for young children.

Comparison With Previous Studies

Digital technology has been widely used as an educational tool for health information delivery [37,38]. WeChat is a popular Chinese instant messaging and social media application. It has also been used to overcome temporal and geographical constraints to deliver health-related information and provide health support [39]. Evidence has shown the effects of WeChat-based interventions on improving specific parenting knowledge or practices. For example, perioperative health education via a WeChat platform effectively enhanced parents’ knowledge of care for children with heart diseases [23]. A WeChat-based parenting training proved promising in improving parenting attitudes among mothers of children with autism during the COVID-19 pandemic [13]. A WeChat health education program in Qinghai, China, significantly increased the early-life EBF rate [24]. Our study results provide further evidence regarding the feasibility of improving parental health literacy via WeChat in a community setting.

The intervention group had significantly increased scores on the CPHLQ’s psychological component after the intervention and a significant group effect after controlling for baseline scores and other confounding factors. This finding suggests that the WOA-based intervention was superior in improving psychological health literacy compared to traditional paper-based health promotion interventions. A recent study supported our finding and reported that a WeChat-based program improved the psychospiritual well-being of patients with digestive cancers [40]. Psychological interventions delivered through WeChat were more effective probably because the contents were more diversified and engaging, such as text, images, audio, videos, and interactional games [20].

Since the caregivers’ educational level could be a significant determinant of parental health literacy [41], we performed subgroup analyses to evaluate the intervention’s effects among caregivers with different educational levels. Although we did not observe significant interaction effects, the results of subgroup analyses show that this WOA-based intervention might be effective only for caregivers with university-level or higher education. This could be possibly explained by the small proportion of caregivers with relatively lower education, only 13% of them with less than university-level education [42], which contributed to an insufficient sample size in the low-education subgroup analysis. Another reason might be that people with a higher education level were also equipped with higher eHealth literacy and thus had a better understanding of web-based information than others [43]. Therefore, future health literacy intervention research should focus on low education to further assess the effect of WOA-based interventions.

The intervention improved the EBF rate at 6 months. In China, many parents and caregivers believe that the right time for stopping EBF and introducing solid foods is approximately 4 months [31]. Although child health staff in Minhang District had promoted EBF for 6 months, at baseline, many caregivers were still unaware of the significance of the recommendation. In this study, we emphasized the importance of EBF for 6 months via a WOA-based intervention program, demonstrating an increase in the EBF rate in the first 6 months. No intervention effect was observed for any breastfeeding at 12 or 24 months. This could be explained by the influence of confounding factors, such as mothers returning to work before their children turn 6 months old and the lack of a supportive environment in the workplace [44]. Although the law mandates an hour of lactation leave per day for nursing mothers in the workplace, supports from society and the workplace for breastfeeding remain to be improved [45], especially in workplaces with a high female employment rate.

VD plays a crucial role in bone health and immunity and has been gaining growing attention for its role in children’s oral health [46,47]. During infancy, VD deficiency is associated with various disorders, such as dental caries and rickets [48]. In an animated video, we recommended that a liquid VD supplement of 400 international units per day begins within the first few days of life and continues throughout childhood, as suggested by the American Academy of Pediatrics [49] and the Chinese Dietary Guidelines for Children [50]. After the 9-month intervention, participants in the intervention group had a higher level of awareness of VD supplementation for infants aged 0-6 months than those in the control group. This finding is similar to that of a previous web-based intervention study [51].

Early screening for visual and oral disorders is an important content of routine health checkups for young children, which can identify related disorders, such as acquired blindness and dental caries [52,53]. Current evidence suggests that maternal education increased children’s visit to oral health care services [54]. Considering that social media has been widely used in health education [55], we hypothesized that a WOA-based intervention for caregivers could significantly improve children’s visit to routine checkups of visual and oral health. However, we did not find such an effect in our study. This might be due to the relatively low weight of the intervention contents on these topics compared to others such as nutrition. In addition, health education resources specified for children’s visual and oral health development need to be further strengthened in China to promote attendance in routine checkups [56].

Limitations

Although this study has shown promising results in enhancing parental health literacy, some limitations should be acknowledged. First, it was conducted only in 1 district of Shanghai, which limits its generalizability to nonurban areas or populations with diverse demographics. Second, despite using random sequence generation for cluster randomization, there was an imbalance between groups in participant’s education.
and child’s Hukou, etc. This might be mainly attributed to the small number of clusters [57]. To minimize bias, we included potential confounders in the multivariate analysis and corrected \( P \) values by the Satterthwaite method. Future study involving more clusters are recommended to evaluate the intervention’s effect more accurately. Third, the low viewing rate of the intervention videos suggests a lack of learning motivation among participants. Given that the intervention was delivered in a real-world setting, this could reflect the true compliance of the health education intervention among caregivers in real life. Unfortunately, we could not collect data on participants’ exposure to supplementary web-based reading material due to the unavailability of administration rights to access the information. Lastly, child psychological development was not included as an outcome measure because psychological assessment is not routinely conducted in child health checkups because there is a need for specialized training, additional resources, and participant cooperation.

**Conclusions**

We developed a comprehensive digital parental health literacy intervention program using WAOs. The intervention offered caregivers a reliable and trusted official source of parental health information. We demonstrated that the social media–based interventions delivered through a WOA could improve parental health literacy and parenting behaviors. Compared with traditional paper-based health materials, a WOA-based intervention can potentially be used for health education to address other public health challenges. Evidence-based content and innovative strategies are needed to engage more participants and achieve better intervention outcomes.

**Acknowledgments**

The authors thank child health care providers and participants at all study sites for their collaboration. This study was funded by Fudan-Minhang Healthcare Union (2019FM01), the Natural Science Foundation of Minhang District, Shanghai (2020MHZ012), Shanghai Minhang Health Commission Project-Minhang District Public Health Model Department (MGWKS02), and Key Discipline Program of Sixth Round of the Three-year Public Health Action Plan (Year 2023-2025) of Shanghai, China (GWVI-11.1-32).

**Data Availability**

The data sets used and analyzed in this study are available from the corresponding author upon reasonable request.

**Authors’ Contributions**

YL, JH, ML, and HJ conceptualized and designed the study, contributed to the development of the trial, and obtained the funding. YL, YY, MC, CJ, SK, and YZ contributed to the implementation of the intervention and data collection. QX carried out data analyses and interpretation. QX, YL, and HJ drafted the manuscript. ML provided critical comments and revisions. ML and HJ are co-senior authors. All authors have read and approved the final manuscript. YY (yangyulin@shmchc.cn) and HJ (h_jiang@fudan.edu.cn) are co-corresponding authors of this manuscript.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Participants’ adherence and retention to the intervention.

[DOCX File, 25 KB - publichealth_v10i1e54623_app1.docx]

**Multimedia Appendix 2**

Missing data report for the WeChat official account–based intervention.

[DOCX File, 19 KB - publichealth_v10i1e54623_app2.docx]

**Checklist 1**

CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) checklist (version 1.6.1).

[PDF File, 7565 KB - publichealth_v10i1e54623_app3.pdf]

**References**


55. Yang M. Intervention study of college freshmen’s oral health care consciousness and behavior based on health communication of WeChat [Master's thesis]. Hangzhou Normal University; 2017.


Abbreviations

BF: breastfeeding  
CHC: community health center  
CPHLQ: Chinese Parental Health Literacy Questionnaire  
EBF: exclusive breastfeeding  
GLMM: generalized linear mixed model  
RR: risk ratio  
VD: vitamin D  
WHO: World Health Organization  
WHO/Europe: World Health Organization–Europe  
WOA: WeChat official account
Epidemiological Characteristics of Varicella in Anhui Province, China, 2012-2021: Surveillance Study

Kun Xuan1*, MPH; Ning Zhang1*, MD, PhD; Tao Li1, MM; Xingya Pang1, MPH; Qingru Li†, MPH; Tianming Zhao2, MD, PhD; Binbing Wang1, MPH; Zhenqiu Zha1, MD, PhD; Jihai Tang1, MPH

1Anhui Provincial Center for Disease Control and Prevention, Hefei, Anhui Province, China
2School of Health Management, Anhui Medical University, Hefei, China
* these authors contributed equally

Corresponding Author:
Jihai Tang, MPH
Anhui Provincial Center for Disease Control and Prevention
No. 12560 Fanhua Road
Economic and Technological Development Zone
Hefei, Anhui Province, 230601
China
Phone: 86 055163674908
Email: phgis@qq.com

Abstract

Background: Varicella is a mild, self-limited disease caused by varicella-zoster virus (VZV) infection. Recently, the disease burden of varicella has been gradually increasing in China; however, the epidemiological characteristics of varicella have not been reported for Anhui Province.

Objective: The aim of this study was to analyze the epidemiology of varicella in Anhui from 2012 to 2021, which can provide a basis for the future study and formulation of varicella prevention and control policies in the province.

Methods: Surveillance data were used to characterize the epidemiology of varicella in Anhui from 2012 to 2021 in terms of population, time, and space. Spatial autocorrelation of varicella was explored using the Moran index ($I$). The Kulldorff space-time scan statistic was used to analyze the spatiotemporal aggregation of varicella.

Results: A total of 276,115 cases of varicella were reported from 2012 to 2021 in Anhui, with an average annual incidence of 44.8 per 100,000, and the highest incidence was 81.2 per 100,000 in 2019. The male-to-female ratio of cases was approximately 1.26, which has been gradually decreasing in recent years. The population aged 5-14 years comprised the high-incidence group, although the incidence in the population 30 years and older has gradually increased. Students accounted for the majority of cases, and the proportion of cases in both home-reared children (aged 0-7 years who are not sent to nurseries, daycare centers, or school) and kindergarten children (aged 3-6 years) has changed slightly in recent years. There were two peaks of varicella incidence annually, except for 2020, and the incidence was typically higher in the winter peak than in summer. The incidence of varicella in southern Anhui was higher than that in northern Anhui. The average annual incidence at the county level ranged from 6.61 to 152.14 per 100,000, and the varicella epidemics in 2018-2021 were relatively severe. The spatial and temporal distribution of varicella in Anhui was not random, with a positive spatial autocorrelation found at the county level ($I=0.412$). There were 11 districts or counties with high-high clusters, mainly distributed in the south of Anhui, and 3 districts or counties with high-low or low-high clusters. Space-time scan analysis identified five possible clusters of areas, and the most likely cluster was distributed in the southeastern region of Anhui.

Conclusions: This study comprehensively describes the epidemiology and changing trend of varicella in Anhui from 2012 to 2021. In the future, preventive and control measures should be strengthened for the key populations and regions of varicella.

(JMIR Public Health Surveill 2024;10:e50673) doi:10.2196/50673

KEYWORDS
varicella; incidence; epidemiology; spatial autocorrelation; contagious disease; chicken pox; varicella zoster virus; China

https://publichealth.jmir.org/2024/1/e50673
**Introduction**

Varicella, also known as chickenpox, is a highly contagious, self-limited disease caused by varicella-zoster virus (VZV) that commonly infects children and could become reactivated following primary infection to induce herpes zoster (HZ), or shingles, infection in the future [1]. As a human-specific pathogenic alphaherpesvirus, primary VZV infection first acts on the upper respiratory mucosa with an incubation period of 10-21 days, followed by a generalized vesicular rash, which is accompanied by fever and headache [2,3]. VZV will remain latent in the human ganglia after the primary infection, which may then become reactivated to induce HZ, a localized, painful vesicular rash that can lead to the development of postherpetic neuralgia in some individuals [4,5]. Although varicella usually presents with mild symptoms, the World Health Organization (WHO) conservatively estimated that 4.2 million hospitalizations were associated with severe varicella complications annually worldwide [6], and that use of the vaccine has reduced the disease burden of varicella to a certain extent.

During the 1960s and 1970s, Japanese scientists developed a live attenuated varicella vaccine with the Oka VZV strain after several cell cultures and passages, which was subsequently used for the active immunization of varicella in many countries [7,8]. Varicella vaccines are commonly used globally in monovalent form or combined with the measles, mumps, and rubella vaccine; varicella vaccines have an overall effectiveness of 81% and effectiveness against moderate or severe varicella of 98% [9]. The WHO recommended that varicella vaccination coverage should be maintained at more than 80% and considered including varicella vaccination in the routine childhood immunization program [6]. Currently, 50 WHO member countries or territories have introduced the varicella vaccine in the routine vaccination schedule [10]. In countries such as the United States, Germany, and Costa Rica, the incidence and hospitalization rates of varicella have significantly reduced compared to those of the prevaccination era [11-14]; thus, the varicella vaccination program has had a positive impact on the epidemiology and disease burden of varicella. However, varicella vaccination is voluntary and self-paying in most areas of China, with few areas providing the vaccine for free and the two-dose schedule has not yet been introduced nationwide. The varicella vaccination rate in China was estimated to be 61% [15], and the reported incidence of varicella increased from 55 per 100,000 people in 2016 to 70 per 100,000 people in 2019, with an average of 1000 outbreaks occurring annually [16]. In addition, reports from some regions of China have shown that the incidence of varicella has increased by approximately 40% annually [17-19]. However, related data have not yet been reported for Anhui Province.

Spatiotemporal investigations based on geographic information systems have been increasingly used in modern epidemiology research [20], which could help to visualize the incidence trends and spatial distributions of infectious diseases, providing a new framework for analyzing their spatiotemporal changes and transmission patterns. Previous studies demonstrated the spatiotemporal aggregation and spatial correlation patterns of varicella occurrence [19,21]. However, no study has yet focused on the epidemiological characteristics or spatial and temporal distributions of varicella in Anhui. Since Anhui is an inland province in the east of China with vast differences in population distribution and geographical characteristics, a systematic epidemiological analysis of the current varicella prevalence could provide a basis for future studies and strategies targeting varicella in Anhui.

Toward this end, the aim of this study was to analyze the epidemiological characteristics of varicella in Anhui using the data of varicella incidence from 2012 to 2021, combined with spatiotemporal epidemiological methods. These findings can provide a basis for the formulation of varicella prevention and control policies and immunization strategies.

**Methods**

**Setting**

Anhui Province covers a total area of 140,100 km², located at longitude 114°54'-119°37' E and latitude 29°41'-34°38' N. Anhui covers the jurisdiction of 16 cities and 104 county-level administrative regions and is part of the Yangtze River Delta region of China. Anhui lies in the transition zone between the warm temperate zone and subtropics, with an average annual temperature of 14-17 °C. The Yangtze River and the Huai River run through the province, dividing the province into three parts, with an overall high elevation in the southwest and low elevation in the northeast. The population of Anhui was estimated at approximately 61.13 million by the end of 2021, showing a greater density in the north than in the south of the province.

**Data Sources**

Data on reported varicella cases and population statistics were obtained from the China Information System for Disease Control and Prevention (CISDCP), which is the world’s largest internet-based disease reporting system that was established by the Chinese government after the 2003 SARS epidemic. The CISDCP is based on the real-time reporting of individual cases and the relevant information of each case can be obtained, such as age, sex, occupation, and the specific location of the onset [22]. Clinical- and laboratory-confirmed cases of varicella reported in the CISDCP were selected for analysis in this study.

**Descriptive Analysis**

The annual and monthly incidence rates of varicella from 2012 to 2021 were collected and epidemic curves were plotted to reveal the incidence peaks. The distribution characteristics of varicella according to variations in age, sex, and occupation were also analyzed. The incidence of varicella in each county or city was associated with geographical information according to the administrative district code, and annual incidence maps were plotted to describe the distribution of varicella in Anhui. All statistical analyses and data visualization were performed using R Studio software (version 4.1.1).

**Spatial Autocorrelation Analysis**

GeoDa (version 1.20) was used to perform the spatial autocorrelation analysis. The Moran index (Moran I) reflects the correlation of attribute values in adjacent areas and can be used to judge the overall spatial distribution [23]. The Moran I
value ranges from –1 to 1, indicating a stronger negative or positive spatial autocorrelation, respectively. In this study, the spatial weight matrix was constructed according to the first-order Queen adjacency rule and the significance test of spatial autocorrelation was carried out. The null hypothesis was that varicella is randomly distributed in Anhui. The hypothesis test of local indicators of spatial association (LISA) [24] was performed and the Moran scatter plot was used to explore the association pattern of hotspot areas. The local association pattern can be divided into four categories: high-high, high-low, low-high, and low-low. The LISA significance test was performed using 999 permutations at a probability level of .05.

**Space-Time Scan Analysis**

The space-time scan statistic is defined by a cylindrical window with a circular (or elliptic or network-based) geographic base and a height corresponding to time. The base is defined as a circular window with changing radius size centered on several possible grid points throughout the study area, while the height reflects the time period of potential clusters. The cylindrical window then moves in space and time so that every possible geographic location and size for each possible time period is accessed. In this process, an infinite number of overlapping cylinders of different sizes and shapes could be obtained, jointly covering the entire study area with each cylinder reflecting a possible cluster. For this study, we adopted the discrete Poisson model described by Kulldorff [25] to conduct a retrospective space-time scan analysis of the spatiotemporal aggregation of varicella. Maximum-likelihood estimation was performed for the locations and sizes of all cylindrical windows, and the alternative hypothesis was that the incidences were higher in the windows than in the outside regions. The $P$ value of the likelihood ratio test was obtained by 999 Monte Carlo simulations based on a probability level of .05. The cylindrical window with the maximum-likelihood ratio was considered to be the most likely cluster and the secondary cluster was sorted according to the size of likelihood ratio test statistics. The maximum spatial cluster size and temporal cluster size were set to 20% of the population at risk and 30% of the study period, respectively. The space-time scan analysis was performed using SaTScan (version 10.1).

**Ethical Considerations**

This study used monitoring data from the CISDCP and no sensitive information was retrieved about the cases or patients. Since this study did not involve any human subjects, it did not require notification to the Ethics Committee according to national ethical review regulations [26].

**Results**

**General Epidemiological Characteristics**

From 2012 to 2021, there were 276,115 cases of varicella and 4 deaths in Anhui, and the average annual incidence was 44.8 per 100,000. Overall, there was an increasing trend of varicella incidence between 2012 and 2019, followed by a decline in 2020 and a slight rebound in 2021. The proportion of male cases among the total number of new annual cases showed a decreasing trend over the decade, whereas the number of female cases gradually increased ($\chi^2 = 269.18$, $P_{\text{trend}} < .001$); the overall sex ratio of cases was approximately 1.26. Details are shown in Table 1 and Figure 1A.

The main age of varicella incidence was 5-14 years, accounting for more than 50% of all cases. The proportion of varicella cases decreased gradually in people aged 0-9 years, whereas there was an increasing trend among those 30 years or older (Figure 1B). In addition, students (ie, children and adolescents 6 years and older who are already enrolled in school) accounted for approximately 60% of reported cases, and the proportion of cases in home-reared children and kindergarten children (aged 3-6 years) increased by 5% and decreased by 9%, respectively (Table 1).

The epidemic curve showed the typical seasonal distribution of varicella with peak incidence occurring twice annually. The incidence was usually higher in the winter peak (November to January of the following year) than in the summer peak (May to June). Only the winter peak occurred in 2020 (Figure 1C).

In terms of the geographical distribution, the incidence of varicella was commonly higher in southern Anhui than in northern Anhui from 2012 to 2021. The high-prevalence areas were the cities of Huangshan, Wuhu, and Ma’anshan, and the lower-prevalence areas were the cities of Huainan, Suzhou, and Anqing (see Multimedia Appendix 1). The incidence in most areas was lower than 50 per 100,000 from 2012 to 2016, whereas between 2018 and 2021, there were some relatively severe varicella epidemics along with clear regional differences (Figure 2). The average incidence at the county level ranged from 6.61 to 152.14 per 100,000.
<table>
<thead>
<tr>
<th>Variables</th>
<th>2012-2013 (n=25,934), n (%)</th>
<th>2014-2015 (n=35,410), n (%)</th>
<th>2016-2017 (n=50,848), n (%)</th>
<th>2018-2019 (n=91,994), n (%)</th>
<th>2020-2021 (n=71,929), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>15,224 (58.7)</td>
<td>20,464 (57.8)</td>
<td>28,672 (56.4)</td>
<td>51,105 (55.6)</td>
<td>38,675 (53.8)</td>
</tr>
<tr>
<td>Female</td>
<td>10,710 (41.3)</td>
<td>14,946 (42.2)</td>
<td>22,176 (43.6)</td>
<td>40,889 (44.4)</td>
<td>33,254 (46.2)</td>
</tr>
<tr>
<td>Male/female</td>
<td>1.42</td>
<td>1.37</td>
<td>1.29</td>
<td>1.25</td>
<td>1.16</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>1036 (4.0)</td>
<td>1177 (3.3)</td>
<td>1693 (3.3)</td>
<td>2290 (2.5)</td>
<td>1520 (2.1)</td>
</tr>
<tr>
<td>1-4</td>
<td>3366 (13.0)</td>
<td>4012 (11.3)</td>
<td>5771 (11.3)</td>
<td>10,261 (11.2)</td>
<td>7871 (10.9)</td>
</tr>
<tr>
<td>5-9</td>
<td>7641 (29.5)</td>
<td>9577 (27.0)</td>
<td>13,555 (26.7)</td>
<td>25,357 (27.6)</td>
<td>16,951 (23.6)</td>
</tr>
<tr>
<td>10-14</td>
<td>5952 (23.0)</td>
<td>9415 (26.6)</td>
<td>13,225 (26.0)</td>
<td>23,805 (25.9)</td>
<td>16,975 (23.6)</td>
</tr>
<tr>
<td>15-19</td>
<td>3850 (14.8)</td>
<td>5545 (15.7)</td>
<td>8059 (15.8)</td>
<td>14,060 (15.3)</td>
<td>13,857 (19.3)</td>
</tr>
<tr>
<td>20-29</td>
<td>3114 (12.0)</td>
<td>4111 (11.6)</td>
<td>5757 (11.3)</td>
<td>9811 (10.7)</td>
<td>7856 (10.9)</td>
</tr>
<tr>
<td>≥30</td>
<td>975 (3.7)</td>
<td>1573 (4.5)</td>
<td>2788 (5.6)</td>
<td>6410 (6.9)</td>
<td>6899 (9.6)</td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home-reared children&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2397 (9.2)</td>
<td>3367 (9.5)</td>
<td>5035 (9.9)</td>
<td>11,653 (12.7)</td>
<td>9583 (13.3)</td>
</tr>
<tr>
<td>Nursery/kindergarten children (3-6 years old)</td>
<td>4394 (16.9)</td>
<td>5240 (14.8)</td>
<td>6579 (12.9)</td>
<td>8312 (9.0)</td>
<td>5678 (7.9)</td>
</tr>
<tr>
<td>Student</td>
<td>15,249 (58.8)</td>
<td>21,458 (60.6)</td>
<td>30,849 (60.7)</td>
<td>56,425 (61.3)</td>
<td>43327 (60.2)</td>
</tr>
<tr>
<td>Farmer</td>
<td>1107 (4.3)</td>
<td>1654 (4.7)</td>
<td>2799 (5.5)</td>
<td>4494 (4.9)</td>
<td>3447 (4.8)</td>
</tr>
<tr>
<td>Others</td>
<td>2787 (10.8)</td>
<td>3691 (10.4)</td>
<td>5586 (11.0)</td>
<td>7110 (12.1)</td>
<td>9894 (13.8)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Home-reared children refer to infants and young children (from birth to 7 years old) who are not sent to nursery or kindergarten and are raised and educated at home.
Figure 1. (A) Incidence of varicella and number of cases by sex, 2012-2021. (B) Age (years) distribution of varicella, 2012-2021. (C) Monthly incidence of varicella in Anhui Province, 2012-2021.
Spatial Autocorrelation

Globally, a positive spatial autocorrelation of varicella incidence was found at the county level from 2012 to 2021, with Moran I of 0.412; this positive correlation was significant throughout the observation period, with Moran I ranging from 0.186 to 0.451 (Table 2). All of the hotspot clusters (high-high) were distributed in the southern region, including six districts and five counties in Huangshan, Wuhu, and Chizhou. The low-low clusters were mainly identified in Huainan, Anqing, and Suzhou. The high-low and low-high clusters included Tianjia’an District, Taihu, and Dangtu County (Figure 3).
Table 2. Global spatial autocorrelation analysis of varicella in Anhui Province, 2012-2021.

<table>
<thead>
<tr>
<th>Year</th>
<th>Moran index</th>
<th>Mean (SD)</th>
<th>Z value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>0.313</td>
<td>-0.0109 (0.0644)</td>
<td>5.029</td>
<td>.001</td>
</tr>
<tr>
<td>2013</td>
<td>0.253</td>
<td>-0.0095 (0.0621)</td>
<td>4.226</td>
<td>.001</td>
</tr>
<tr>
<td>2014</td>
<td>0.186</td>
<td>-0.0086 (0.0649)</td>
<td>2.993</td>
<td>.003</td>
</tr>
<tr>
<td>2015</td>
<td>0.204</td>
<td>-0.0121 (0.062)</td>
<td>3.485</td>
<td>.003</td>
</tr>
<tr>
<td>2016</td>
<td>0.347</td>
<td>-0.0107 (0.0622)</td>
<td>5.762</td>
<td>.001</td>
</tr>
<tr>
<td>2017</td>
<td>0.319</td>
<td>-0.0098 (0.0625)</td>
<td>5.264</td>
<td>.001</td>
</tr>
<tr>
<td>2018</td>
<td>0.282</td>
<td>-0.0097 (0.0593)</td>
<td>4.913</td>
<td>.001</td>
</tr>
<tr>
<td>2019</td>
<td>0.409</td>
<td>-0.0118 (0.0612)</td>
<td>6.877</td>
<td>.001</td>
</tr>
<tr>
<td>2020</td>
<td>0.282</td>
<td>-0.0084 (0.0621)</td>
<td>4.681</td>
<td>.001</td>
</tr>
<tr>
<td>2021</td>
<td>0.451</td>
<td>-0.0106 (0.0622)</td>
<td>7.424</td>
<td>.001</td>
</tr>
<tr>
<td>Overall</td>
<td>0.412</td>
<td>-0.0132 (0.0626)</td>
<td>6.795</td>
<td>.001</td>
</tr>
</tbody>
</table>
Spatiotemporal Clusters

Space-time scan analysis identified five possible clusters of areas where varicella incidence was higher than that in other areas in the specific time period. The most likely cluster was located in southeastern Anhui, including parts of Ma’anshan, Wuhu, Huangshan, Tongling, and Chizhou and the whole territory of Xuancheng. The four secondary clusters were mainly distributed in Hefei, Bengbu, Fuyang, and the Xiangshan district of Huaihe (Figure 4).

Figure 3. Spatial autocorrelation local indicators of spatial association (LISA) cluster maps of varicella in Anhui Province, 2021-2021.
Discussion

Principal Findings

In this study, we comprehensively described the epidemiology of varicella in Anhui over the last decade. To our knowledge, this is the first study to analyze the changing epidemiological characteristics and identify the incidence hotspot areas of varicella from a province perspective. The reported incidence of varicella in Anhui increased by approximately 20% annually from 2012 to 2019, with a 10-year average incidence of 44.8 per 100,000, which is largely consistent with the rising trend in other regions of China [16,27,28]. The incidence peaked in 2019 at 81.2 per 100,000 people, which is higher than the national average but is lower than the averages for Guizhou, Hangzhou, and Dalian for the same period [28-30]. The higher incidence of varicella in Anhui may be related to the increased sensitivity of infectious disease surveillance in recent years but also may be related to the insufficient coverage of vaccination. In the United States, the routine varicella vaccination program began in 1995, and the Advisory Committee on Immunization Practices recommended adding a second dose for children 4-6 years of age in 2007 [31]; there has been a 97% reduction in varicella incidence in 2019 compared to that prior to the implementation of routine vaccination [32]. Japan introduced routine vaccination with two doses of varicella vaccine in 2014, and the number of varicella hospitalizations decreased by 88% in 2017, including an 88% decrease in hospitalization of children aged 1-4 years [33]. By 2019, the incidence of varicella in Japan had decreased by 78% from that reported in 2000-2011 [34]. In South Korea and Turkey, in which one-dose routine immunization was implemented, the incidence, hospitalization, and complication rates of varicella were significantly reduced due to high vaccination rates [14,35]. The varicella vaccine is not currently included in the routine immunization program in China and there is no universal two-dose varicella vaccination, which may explain the high disease burden of varicella in Anhui.

Similar to previous studies, there were more cases of varicella in males than females [16,19]. However, the sex difference has narrowed gradually in recent years, which was also found in the study in Chongqing, China [19]. This may be caused by the widespread transmission of VZV in susceptible populations as the community has not yet established immune barriers. The proportion of varicella cases among children and adolescents under 14 years of age decreased slightly, whereas the proportion among adults 30 years and older tended to increase in Anhui. More children are now vaccinated compared to the historical proportion, and the vaccine is demonstrated to be effective in reducing the risk of infection in younger age groups. This study found a shift in the age of onset to older age groups, which is consistent with previous studies [36,37]. However, this outcome differs from other studies that did not observe an increase in the average age of varicella incidence, likely due to herd immunity obtained from high vaccine coverage [14,38,39]. Nevertheless, enhancing surveillance of changes in varicella prevalence among adults in Anhui, especially those over 30 years of age, is warranted in the future. In addition, the proportion of varicella cases among kindergarten children gradually decreased, whereas the proportion increased among home-reared children, which slightly differs from the results of a previous study in Chongqing [19]. More children were
vaccinated before entering kindergarten according to confirmation of vaccination certificates, which also provides considerable protection despite the risk of clustered outbreaks. Suspected cases of varicella can also be detected in a timely manner through routine morning and noon physical examinations, thereby reducing the impact on other children. In general, home-reared children are less likely to be vaccinated because there is not a preadmission vaccination certificate requirement as is the case for enrolling in kindergarten. Moreover, since home-reared children typically accompany their guardians in their living and socializing activities, they are more likely to contact a wider variety of people compared to kindergarten children and do not have the immunity from vaccination; thus, home-reared children are more likely to come into contact with VZV-infected people, making them more susceptible to VZV infection. This suggests that more attention should be paid to home-reared children to improve their vaccination rate and protection, along with further education of their guardians regarding varicella in parallel with strengthening school outbreak surveillance.

Varicella showed a seasonal rise in winter and summer, which was consistent with previous studies [19,21]. In 2020, the incidence of varicella significantly declined and only a predominate winter peak occurred; this phenomenon has also been observed for other respiratory infectious diseases. The incidence of notifiable respiratory infectious diseases in China decreased significantly in 2020, with a 65% decrease in both measles and influenza [40]. Following the COVID-19 pandemic, many public health measures, including wider lockdowns, wearing masks, and social distancing, were taken to reduce the transmission of SARS-CoV-2 via droplets and aerosols [41]. These measures may have reduced the exposure to other respiratory viruses. In addition, concerns about COVID-19 infection and increased awareness of self-protection have led to fewer reports from health care facilities. According to the seasonal pattern of varicella (not considering 2020), health education, vaccination, and the sensitivity of surveillance should be strengthened during the low-incidence period to help reduce outbreaks and more severe prevalence during the high-incidence season.

The spatial and temporal distributions of varicella were not random in Anhui. Although spatiotemporal clusters of varicella have occurred in the cities of Bengbu, Huaiabei, and Fuyang in the northern region of Anhui from 2017 to 2021, on the whole, the reported incidence in southern Anhui was higher than that in northern Anhui. Several studies found that respiratory infectious diseases showed patterns of high-risk clusters in remote rural or urban-rural transition zones where economic, educational, and medical resources are often lacking [21,42,43]. However, the most likely clusters of varicella in Anhui were located in the more economically developed areas of Hefei and in the southeast region of the province. Geographically, Anhui is connected to Jiangsu Province in the east and Zhejiang Province in the southeast, and the prevalence of varicella was higher in these two adjacent provinces than in Anhui, whereas the incidence was lower in the Henan and Shandong provinces north of Anhui [16,44]. There has been more frequent movement of the population in Anhui, mainly to Jiangsu and Zhejiang, in the past few years than in other provinces [45]. This may be one of the reasons for the inconsistency between the characteristics of high-risk clusters in this study and in previous studies. For example, Huangshan and Wuhu, which are located in the southeast of Anhui, were not only the high-incidence areas of varicella but were also high-high clusters, and the most likely spatiotemporal clusters were also distributed in these cities and their surroundings. Alternatively, there may be differences in surveillance sensitivity, vaccination rates, and geographic and climatic characteristics in different areas of Anhui, resulting in the observed variable spatial distribution pattern of varicella. By contrast, northern Anhui is located in the southern part of the North China Plain with a relatively higher population density and mobility than other areas of the province, which would theoretically facilitate the transmission of respiratory infectious diseases. Considering the high underestimation rate of varicella reporting in China [44], the actual disease burden in these apparently low-incidence areas may actually be far greater than indicated in our analysis. Therefore, the true reasons for the overall difference in incidence between northern and southern Anhui need to be thoroughly explored.

Moreover, the high-low and low-high clusters identified by spatial autocorrelation analysis are perhaps of greater concern than clusters of hot or cold spots. In other words, more attention needs to be paid to the surrounding varicella incidence to avoid the inflow of the epidemic from the high-incidence areas (Dangtu County) or the increased incidence in the surrounding areas due to the high local incidence (Tianjia’an District and Taihu County), providing evidence for the future focus of varicella prevention and control strategies.

Although vaccination can have a positive impact on the epidemiology of varicella, some studies found that universal routine immunization with varicella vaccine increased the incidence of HZ [46,47]. Although caused by the same virus, these two diseases were not analyzed together in this study due to the lack of data on HZ. Humoral immunity persists in individuals with prior varicella infection; however, cell-mediated immunity, which plays an important role in the development of HZ, typically wanes with age [47,48]. Exposure to wild-type varicella cases could enhance cell-mediated immunity against VZV and maintain the immune effect for a longer time, whereas the opportunity for such exogenous enhancement decreases when varicella vaccination rates increase, which may account for the increased incidence of HZ [49]. Varicella is closely related to the epidemiology of HZ; thus, attention should be paid to enhancing the surveillance of HZ. Although the vaccine reduces the disease burden of varicella, the disease burden resulting from the increased incidence of HZ may offset some of the economic benefits of varicella vaccination, which should be taken into account when including varicella into routine immunization programs. According to the varicella vaccine immunization strategy in Anhui, since June 2021, the first dose of varicella vaccine is recommended at the age of 1 year and the second dose is recommended at the age of 4 years. Therefore, the single- and double-dose vaccination rates and the effects of administering two doses of the vaccine also warrant further study.
Limitations

There are some limitations of this study. First, although substantial disease information has been provided through the CISDCP surveillance, there are still some shortcomings in the surveillance of varicella in Anhui Province. The proportion of cases underreported after mandatory reporting is likely to be small [50,51] and is unlikely to affect the epidemiology of the disease. Some regions in China already require the reporting of varicella cases within 24 hours [19,29], although this is not mandatory in Anhui and the actual data on underreporting are currently unavailable, which may have led to an underestimation of the incidence and possible alteration of the epidemiological characteristics of varicella. Therefore, more comprehensive varicella surveillance data are needed to assess the effectiveness of vaccination and to guide future varicella control and immunization strategies. Second, scan statistics based on circular windows often produce clusters that are larger than they actually are and can fail to identify irregular clusters; therefore, varicella clusters identified using flexible scan statistics may be more practical [52]. Finally, information on factors affecting the spread of infectious diseases, such as economy, population density, vaccination rates, and climate, was not considered, as this work only described the aggregation phenomenon of varicella; thus, the potential underlying causes need to be further studied.

Conclusion

This study provides detailed epidemiological characteristics and changes of varicella from the perspectives of population, time, and space based on the reported incidence data in Anhui from 2012 to 2021. Measures should be taken in the future to strengthen varicella prevention and control in the key populations and regions identified in this study.

Acknowledgments

This research was supported by the Health Research Program of Anhui (AHWJ2023A20369). We appreciate the staff at all levels from the medical institutions of Anhui Province for reporting and reviewing the surveillance data. We declare that no artificial intelligence was used in the preparation of this manuscript.

Data Availability

The data sets generated and/or analyzed during the present study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to this paper. NZ provided the study data. KX, XP, and QL analyzed the data. KX, NZ, and TL wrote the manuscript. BW and ZZ made the initial revision. JT designed the study and revised the manuscript. TZ touched up the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Total incidence of varicella in each city of Anhui Province, 2012-2021. [PNG File, 127 KB - publichealth_v10i1e50673_app1.png]

References


Abbreviations

CISDCP: China Information System for Disease Control and Prevention
HZ: herpes zoster
LISA: local indicators of spatial association
VZV: varicella-zoster virus
WHO: World Health Organization

https://publichealth.jmir.org/2024/1/e50673
Original Paper

Epidemic Characteristics and Meteorological Risk Factors of Hemorrhagic Fever With Renal Syndrome in 151 Cities in China From 2015 to 2021: Retrospective Analysis

Yizhe Luo1,2*, DrPH; Longyao Zhang3*, MPH; Yameng Xu1,2, MPH; Qiyuan Kui2, PhD; Wenhao Li2, PhD; Yifan Wu2, BMed; Licheng Liu4, PhD; Jiarong Ren5, MD; Lingling Zhang6, PhD; Qiufang Shi3*, MPH; Xiaobo Liu5,7,8*, PhD; Weilong Tan1,2*, MD

1Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing, China
2Nanjing Bioengineering (Gene) Technology Center for Medicines, Nanjing, China
3Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China
4Jiangsu Macro and Micro Test Med-tech Co, Ltd, Nantong, China
5National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, Beijing, China
6College of Life Science, Fujian Agriculture and Forestry University, Fuzhou, China
7Department of Vector Control, School of Public Health, Shandong University, Jinan, China
8Xinjiang Key Laboratory of Vector-borne Infectious Diseases, Urumqi, China

*these authors contributed equally

Corresponding Author:
Weilong Tan, MD
Department of Epidemiology
School of Public Health
Nanjing Medical University
Meiyuan Xincun Street
Nanjing, 210002
China
Phone: 86 17384408593
Email: njdc@163.com

Abstract

Background: Hemorrhagic fever with renal syndrome (HFRS) continues to pose a significant public health threat to the population in China. Previous epidemiological evidence indicates that HFRS is climate sensitive and influenced by meteorological factors. However, past studies either focused on too-narrow geographical regions or investigated time periods that were too early. There is an urgent need for a comprehensive analysis to interpret the epidemiological patterns of meteorological factors affecting the incidence of HFRS across diverse climate zones.

Objective: In this study, we aimed to describe the overall epidemic characteristics of HFRS and explore the linkage between monthly HFRS cases and meteorological factors at different climate levels in China.

Methods: The reported HFRS cases and meteorological data were collected from 151 cities in China during the period from 2015 to 2021. We conducted a 3-stage analysis, adopting a distributed lag nonlinear model and a generalized additive model to estimate the interactions and marginal effects of meteorological factors on HFRS.

Results: This study included a total of 63,180 cases of HFRS; the epidemic trends showed seasonal fluctuations, with patterns varying across different climate zones. Temperature had the greatest impact on the incidence of HFRS, with the maximum hysteresis effects being at 1 month (−19 °C; relative risk [RR] 1.64, 95% CI 1.24-2.15) in the midtemperate zone, 0 months (28 °C; RR 3.15, 95% CI 2.13-4.65) in the warm-temperate zone, and 0 months (4 °C; RR 1.72, 95% CI 1.31-2.25) in the subtropical zone. Interactions were discovered between the average temperature, relative humidity, and precipitation in different temperature zones. Moreover, the influence of precipitation and relative humidity on the incidence of HFRS had different characteristics under different temperature layers. The hysteresis effect of meteorological factors did not end after an epidemic season, but gradually weakened in the following 1 or 2 seasons.

https://publichealth.jmir.org/2024/1/e52221
Conclusions: Weather variability, especially low temperature, plays an important role in epidemics of HFRS in China. A long hysteresis effect indicates the necessity of continuous intervention following an HFRS epidemic. This finding can help public health departments guide the prevention and control of HFRS and develop strategies to cope with the impacts of climate change in specific regions.

(JMIR Public Health Surveill 2024;10:e52221) doi: 10.2196/52221

KEYWORDS
China; hemorrhagic fever with renal syndrome; HFRS; climate change; meteorological factors; distributed lag nonlinear model

Introduction
Hemorrhagic fever with renal syndrome (HFRS) is a rodent-borne zoonotic disease caused by Hantavirus (HTNV), causing symptoms such as fever, headache, and renal dysfunction [1]. Globally, China is one of the countries most affected by HFRS, with cases covering 31 provinces, municipalities, and autonomous regions and numbering nearly 10,000 per year in the past decade [2]. HFRS is listed as a class B infectious disease due to its potential threat to public health in China [3]. Despite the implementation of vaccination programs, HFRS remains a serious public health problem in China. More than 10 HTNV-potential hosts have been identified with population levels sufficient to sustain virus viability and reproduction in nature [4]. Moreover, the enormous geographic differences and the variety of climate types in China make it difficult or impossible to eliminate HFRS, where 9187 new cases were reported in 2021, including 64 deaths [5].

Known risk factors for HFRS incidence include climate, host population, and viral dynamics [6]. Climate is widely recognized as a key factor in HTNV transmission, mainly affecting the prevalence of the virus and the risk of human infection by affecting rodent population dynamics [7,8]. For example, temperature, humidity, and precipitation can affect crop yields, which are a food source for rodents [7]. Earlier studies showed that climate-related HFRS outbreaks had a hysteresis effect, usually delayed by 1 to 6 months, and seasonal patterns of HFRS epidemics also showed sensitivity to climate [9-11]. Using a distributed lag nonlinear model (DLNM), Luo et al [12] investigated temperature (lag=6 months, relative risk [RR] 3.05) and precipitation (lag=0 months, RR 2.08), which had the greatest impact on the incidence of HFRS. Sun et al [7] also identified extremely high or low temperature as being strongly associated with HFRS. Concerns have been raised in recent years about the expansion of HFRS-affected areas and the reemergence of HFRS in regions where it had been eliminated. Under global warming, cyclic dynamics of rodent populations are changing, and new endemic areas are forming [13]. There is an urgent need to explore the propagation of HFRS under different climate conditions. Cao et al [14] explored the interactions and marginal effects of meteorological factors on HFRS in different climate zones in 254 cities in China. However, the time span of their study was 2006 to 2016, which is too far in the past; few or no studies have discussed the relationship between HFRS and meteorological factors in different regions in recent years [15].

Based on HFRS monitoring data from 151 cities from 2015 to 2021, we estimated the hysteresis effects and interactions of meteorological factors in different climate zones on HFRS in China. We also sought to identify associations between temperature and HFRS under different environmental conditions and examine whether these associations varied geographically. Compared with previous studies, our research covers a wider range of subjects, is more recent, and has more convincing results.

Methods

Ethical Considerations
The study was approved by the ethical review board of the Nanjing Bioengineering (Gene) Technology Center for Medicines (2022005). Consent to participate was not applicable because this study used HFRS surveillance data. All participant data were anonymized and kept confidential to protect the privacy of participants.

Study Sites
This study was based on a national database of meteorological factors and confirmed HFRS case counts in 151 Chinese prefecture-level cities from January 1, 2015, to December 31, 2021. According to China’s national reporting system for infectious diseases, the total number of HFRS cases in these cities during the period they were included in the study exceeded 50.

China can be divided into 6 climate zones (Multimedia Appendix 1, Figure S1) [16]. In this study, 3 climatic zones were chosen as the research zones: the midtemperate zone, the warm-temperate zone, and the subtropical zone. The cold-temperate zone, the plateau-temperate zone, and the tropical zone were excluded due to having too few cities (<5 cities). A final total of 151 Chinese prefecture-level cities were included in our study (Multimedia Appendix 1, Table S1). When the city boundary spanned multiple climatic zones, the urban climatic zones were divided according to the location of the city center.

Collection of Data
Monthly data on HFRS from the 151 prefecture-level cities across China from January 1, 2015, to December 31, 2021, including the number of cases and incidence, were provided by the Chinese Center for Disease Control and Prevention. All notified HFRS cases were confirmed according to the united diagnostic criteria issued by the Ministry of Health of China in 1998 [17].

Monthly meteorological data including average temperature (Celsius), average relative humidity (%), and average precipitation (mm) in the selected cities were obtained from
839 meteorological monitoring stations [18] (Multimedia Appendix 1, Figure S2).

**Statistical Analysis**

In the descriptive analysis, the mean (SD), median (IQR), and range were used to describe the distribution of cases of HFRS and weather variables in the 3 selected climatic zones. In 2018, there were some missing values for meteorological variables in 18 cities, including mean temperature (n=216), relative humidity (n=216) and precipitation (n=216). Thus, we imputed the values of each city using their last year’s value. The descriptive statistics before and after imputing are summarized in Table 1.
Table 1. Descriptive analysis of monthly mean temperature, precipitation, and average relative humidity in different climate zones in China from 2015 to 2021. There were 12 missing values in the midtemperate zone, 60 in the warm-temperate zone, and 144 in the subtropical zone.

<table>
<thead>
<tr>
<th>Zones and factors</th>
<th>Values</th>
<th>Imputed values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Midtemperate zone (n=25 cities)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HFRS cases (n)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>6 (7)</td>
<td>b</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>4 (2 to 8)</td>
<td>—</td>
</tr>
<tr>
<td>Range</td>
<td>0 to 79</td>
<td>—</td>
</tr>
<tr>
<td><strong>Temperature (°C)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>5.42 (13.68)</td>
<td>5.42 (13.68)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>7.49 (–6.93 to 17.36)</td>
<td>7.47 (6.96 to 17.36)</td>
</tr>
<tr>
<td><strong>Precipitation (mm)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>55.22 (67.32)</td>
<td>55.22 (67.32)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>27.70 (8.90 to 79.55)</td>
<td>27.70 (8.90 to 79.55)</td>
</tr>
<tr>
<td>Range</td>
<td>0.00 to 677.20</td>
<td>0.00 to 677.20</td>
</tr>
<tr>
<td><strong>Relative humidity (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>64.55 (11.82)</td>
<td>64.55 (11.82)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>65 (56.86 to 73)</td>
<td>25 (56.93 to 65.02)</td>
</tr>
<tr>
<td>Range</td>
<td>25 to 93</td>
<td>25 to 93</td>
</tr>
<tr>
<td><strong>Warm-temperate zone (n=54 cities)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HFRS cases (n)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>6 (22)</td>
<td>—</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2 (0 to 5)</td>
<td>—</td>
</tr>
<tr>
<td>Range</td>
<td>0 to 878</td>
<td>—</td>
</tr>
<tr>
<td><strong>Temperature (°C)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>13.43 (10.08)</td>
<td>13.43 (10.08)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>14.25 (4.32 to 22.59)</td>
<td>14.25 (4.32 to 22.59)</td>
</tr>
<tr>
<td>Range</td>
<td>–10.96 to 30.48</td>
<td>–10.96 to 30.48</td>
</tr>
<tr>
<td><strong>Precipitation (mm)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>63.14 (136.24)</td>
<td>63.65 (136.24)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>33.00 (9.50 to 78.30)</td>
<td>32.90 (9.30 to 3450.80)</td>
</tr>
<tr>
<td>Range</td>
<td>0.00 to 3450.80</td>
<td>32.90 (9.30 to 3450.80)</td>
</tr>
<tr>
<td><strong>Relative humidity (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>63.63 (12.38)</td>
<td>63.52 (12.42)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>63.87 (55 to 73)</td>
<td>63.87 (54.69 to 73)</td>
</tr>
<tr>
<td>Range</td>
<td>23.32 to 94.42</td>
<td>23.32 to 94.42</td>
</tr>
<tr>
<td><strong>Subtropical zone (n=72 cities)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HFRS cases (n)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>4 (6)</td>
<td>—</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2 (1 to 5)</td>
<td>—</td>
</tr>
<tr>
<td>Range</td>
<td>0 to 100</td>
<td>—</td>
</tr>
<tr>
<td><strong>Temperature (°C)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>18.68 (7.94)</td>
<td>18.70 (7.94)</td>
</tr>
</tbody>
</table>
First Stage Analysis

We first captured the association between weather conditions and HFRS incidence in different climate zones with a DLNM, which can flexibly describe relationships and explore underlying lag nonlinear effects [19]. The climate-specific model with adjustment for potential confounders used equation (1):

\[
E(Y_{it}) = \alpha + LIM_{it} + cb(\text{Temp mean}) + cb(\text{Rh mean}) + cb(\text{Prec mean}) + \text{Season}_{it} + \text{Province}_{it} + \text{Time}_{it} + \text{offset} + \epsilon
\]

Here, \( E(Y_{it}) \) denotes the monthly expected number of cases in the city and month \( t \), and \( \alpha \) is the intercept. \( LIM_{it} \) represents the previous month’s incidence, used to reduce autocorrelation between values. \( cb(\text{Temp mean}) \), \( cb(\text{Rh mean}) \), and \( cb(\text{Prec mean}) \) are cross-basis matrices of selected monthly meteorological factors, constructed using B-splines for exposure and natural cubic splines for lag (maximum lag of 6 month) dimensions, respectively. A natural cubic spline function of 1 degree of freedom per year was used to control the long-term trend of incidence. The \( \text{Time} \) variable indicates the sequence (from 1 to 84) with month as the unit during the study period, 2015 to 2021. \( \text{Season}_{it} \) and \( \text{Province}_{it} \) are categorical variables to control seasonal patterns and different provinces, respectively. The offset term is the population of each city.

Second Stage Analysis

In the second stage, the correlation of HFRS and climate under different hysteresis conditions is examined. Exposure-effect curves for climate variables with different lag times in the 3 selected temperature zones were drawn to illustrate the hysteresis effects and their duration under different meteorological conditions. Furthermore, the HFRS-climate association at temperature extremes was explored. Taking the median of different meteorological conditions as the reference, the relationship between the corrected meteorological conditions at the 2.5th and 97.5th percentiles and the incidence of HFRS were calculated, respectively. Finally, a random-effect meta-analysis of city-specific relationships for different climate zones estimated in the first stage was performed using the restricted maximum likelihood estimation method to provide more accurate estimates.

Third Stage Analysis

To explore the interaction and stratification effects between the 3 weather condition and HFRS epidemics in different climate zones, we constructed a generalized additive model (GAM) in the third stage. The model can be written as equation (2):

\[
E(Y_{it}) = \alpha + s(K, X) + s_1(K, X) + s_2(K) + s_3(X) + \text{offset} + \epsilon
\]

Here, \( \alpha \) is the intercept; \( K \) denotes 1 of the weather conditions (mean temperature, relative humidity, and precipitation), and \( X \) and \( Z \) denote the other 2 indicated penalized spline functions. \( s_1(K, X) \) is the spline function of the interaction between variables \( K \) and \( X \).

Then, the meteorological stratification between mean temperature and the incidence of HFRS was determined for relative humidity and precipitation. We split relative humidity and precipitation into category variables, including the medians for “high” and “low.”

Sensitivity Analysis

To test whether our main conclusions were robust, we performed a sensitivity analysis relying on the quasi-Akaike information criterion (QAIC) and quasi-Bayesian information criterion (QBIC). We used QAIC and QBIC to identify the optimal number and location of knots for the natural spline and the optimal number of lag months from 1 to 6 (Multimedia Appendix 1, Table S2).

R (version 4.2.1; R Foundation for Statistical Computing) and the R packages dlnm, metafor, and mcgv were used for constructing the DLNM and GAM based on the 3 meteorological variables. All maps were created using ArcGIS (version 10.2; Esri Inc). The confidence level of all 2-sided statistical tests in this study was set at 95%, and the significance level was set at .05.

Results

Descriptive Analysis

A total of 63,180 HFRS cases were involved in our study (Figure 1). Between 2015 and 2021, 12,481, 28,353, and 22,346 cases...
of HFRS occurred in the midtemperate zone, warm-temperate zone, and subtropical zone. Multimedia Appendix 1, Figure S3 shows bar charts of different seasonal prevalence patterns of HFRS in the 3 climate zones. The peaks for HFRS cases in the midtemperate, warm-temperate, and subtropical zones occurred in autumn; winter and spring; and winter and spring, respectively. The mean values for temperature, precipitation, and relative humidity in the 3 zones were 5.43 °C, 55.28 mm, and 64.53%; 13.44 °C, 63.14 mm, and 63.63%; and 18.68 °C, 128.8 mm, and 77.03%, respectively. Thus, there was a gradual increase from north to south.

Figure 1. Distribution of cases of hemorrhagic fever with renal syndrome in China from 2015 to 2021.

DLNM Analysis

The cumulative risk between meteorological factors and HFRS incidence in the DLNM for different climate zones in China after controlling for seasonal and long-term trends is shown in Figure 2. In the midtemperate zone, the meteorological conditions that were positively correlated with HFRS risk were mean temperature <–7 °C and precipitation 28 to 134 mm; in the warm-temperate zone, the meteorological conditions were a mean temperature <–7 °C or 14 °C to 24 °C and precipitation 143 mm to 274 mm; in the subtropical zone, the meteorological conditions were mean temperature 9 °C to 19 °C, precipitation 11 mm to 22 mm or 95 mm to 299 mm, and relative humidity 78% to 84%. Figure 3 and Multimedia Appendix 1, Figure S4 show the impact of different lag months on climate-related HFRS risk. In the midtemperate zone, significant RRs were observed at lag of 1 month when mean temperature was –19 °C (RR 1.64, 95% CI 1.24-2.15). In the warm-temperate zone, a temperature of 28 °C (0-month lag; RR 3.15, 95% CI 2.13-4.65), precipitation of 239 mm (1-month lag; RR 1.22, 95% CI 1.06-1.40), and relative humidity of 83% (6-month lag; RR 1.21, 95% CI 1.07-1.36) resulted in significantly higher RR. In addition, in the subtropical zone, temperature, precipitation, and relative humidity with lags of 0, 6, and 4 months had high RRs at 4 °C (RR 1.72, 95% CI 1.31-2.25), 360 mm (RR 1.16, 95% CI 1.06-1.26), and 90% (RR 1.11, 95% CI 1.05-1.17), respectively. Multimedia Appendix 1, Figure S5 shows the RR between climate and HFRS with different lag months for extreme weather. In the midtemperate zone and subtropical zone, HFRS was sensitive to low temperature, while in the warm-temperate zone, HFRS was more sensitive to high temperature. Higher precipitation and humidity were associated with the incidence of HFRS in the warm temperate zone and subtropical zone. Moreover, the results of the meta-analysis showed that low temperature, relatively high precipitation, and high relative humidity were risk factors for the onset of HFRS, but the subtropical zone showed a different trend (Figure 4).
Figure 2. Summary of cumulative exposure-response curves of hemorrhagic fever with renal syndrome incidence for meteorological factors with a lag of 0-6 months in 3 selected temperature zones from 2015 to 2021. The y-axis represents the relative risk of each variable. The x-axis represents the range of observations for each variable. The blue lines represent means estimated by the distributed lag nonlinear model, and the shaded areas represent the 95% CI. MTZ: midtemperate zone; SZ: subtropical zone; WTZ: warm temperate zone.
Figure 3. Lag-specific effects of meteorological factors on hemorrhagic fever with renal syndrome infection in different climate zones from 2015 to 2021. The y-axis represents the relative risk of each variable. The x-axis represents the range of observations for each variable. The purple lines represent means estimated by the distributed lag nonlinear model, and the shaded areas represent the 95% CI. MTZ: midtemperate zone; SZ: subtropical zone; WTZ: warm temperate zone.
**Interaction and Stratified Analysis**

The results for 25 cities in the midtemperate zone showed that in environments with low temperature and low relative humidity and in environments with low temperature and high relative humidity, each decrease of 1 °C led to an increase in the risk of HFRS of 3% (95% CI 2.6%-3.4%) and 4.6% (95% CI 3.8%-5.4%), respectively. In environments with low temperature and low precipitation and in environments with low temperature and high precipitation, each 1 °C decrease led to an increase in the risk of HFRS of 3.3% (95% CI 2.8%-3.9%) and 5.3% (95% CI 4.4%-6.1%), respectively. The results from 54 cities in the warm-temperate zone showed that in environments with low temperature and low relative humidity and in environments with low temperature and high relative humidity, each decrease of 1 °C led to an increase in the risk of HFRS of 1.8% (95% CI 1.1%-2.5%).
1.5%-2.2%) and 4.7% (95% CI 3.8%-5.7%), respectively. In environments with low temperature and low precipitation and in environments with low temperature and high precipitation, each decrease of 1 °C led to an increase in the risk of HFRS of 2.3% (95% CI 2%-2.7%) and 2.4% (95% CI 1.6%-3.3%), respectively (Table 2 and Multimedia Appendix 1, Table S3-S12).

<table>
<thead>
<tr>
<th>Temperature zones</th>
<th>Relative humidity, relative risk (95% CI)</th>
<th>Precipitation, relative risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Midtemperate zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low temperature (−19 °C)a</td>
<td>1.030 (1.034-1.026)</td>
<td>1.046 (1.054-1.038)</td>
</tr>
<tr>
<td>Middle temperature (7 °C)</td>
<td>0.971 (0.976-0.965)</td>
<td>0.956 (0.952-0.961)</td>
</tr>
<tr>
<td>High temperature (25 °C)</td>
<td>0.971 (0.978-0.963)</td>
<td>0.956 (0.958-0.955)</td>
</tr>
<tr>
<td>Warm-temperate zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low temperature (−5 °C)b</td>
<td>1.018 (1.022-1.015)</td>
<td>1.047 (1.057-1.038)</td>
</tr>
<tr>
<td>Middle temperature (14 °C)</td>
<td>0.982 (0.986-0.978)</td>
<td>0.955 (0.951-0.959)</td>
</tr>
<tr>
<td>High temperature (28 °C)</td>
<td>0.982 (0.990-0.975)</td>
<td>0.955 (0.956-0.953)</td>
</tr>
<tr>
<td>Subtropical zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low temperature (4 °C)c</td>
<td>0.969 (0.967-0.971)</td>
<td>0.973 (0.965-0.980)</td>
</tr>
<tr>
<td>Middle temperature (19 °C)</td>
<td>0.969 (0.974-0.964)</td>
<td>0.973 (0.970-0.975)</td>
</tr>
<tr>
<td>High temperature (30 °C)</td>
<td>0.969 (0.977-0.961)</td>
<td>0.973 (0.976-0.969)</td>
</tr>
</tbody>
</table>

These values represent the effect of a decrease of 1 °C, while unmarked values represent the effect of an increase of 1 °C.

**Discussion**

**Principal Findings**

In this study, we quantified the HFRS-weather association in 151 selected cities using a DLNM. Previous studies have investigated the HFRS-weather associations at single- or multiple-location levels [14,15]. However, few of them explored the interaction and hysteresis effects of meteorological factors at different levels on the incidence of HFRS in different climate zones. To the best of our knowledge, this is the first study using national data to explore the interaction and hysteresis effects of meteorological factors at different levels on the incidence of HFRS in China. Our pooled analysis results showed that meteorological factors, especially temperature, had a significant impact on the incidence of HFRS in China, although the association between HFRS and different meteorological factors, the lag time with the greatest impact, and the duration of the lagged effect varied by location. Moreover, we found low temperature was a contributing factor to the pathogenesis of HFRS, and HFRS was relatively more sensitive to temperature changes in the warm- and midtemperate zones than in the subtropical zone. This finding can help public health departments guide the prevention and control of HFRS and develop strategies to cope with the impacts of climate change in specific regions.

We found that low temperature was the most important meteorological factor for the onset of HFRS. It is worth noting that the relationship between HFRS and weather varies by location. There were different impact patterns in the 3 climatic zones. The midtemperate zone peaked at low temperatures (<−7 °C); the warm-temperate zone peaked at both low temperatures (<−4 °C) and high temperatures (14-24 °C); the subtropical zone peaked at 9-19 °C. The differences may be related to the different host animal species, reproduction, and activity cycles. *Apodemus agrarius* and *Rattus norvegicus* are the main species involved in HFRS transmission in China. HFRS infection caused by *A agrarius* mainly occurs in autumn and winter, while infection caused by *R norvegicus* occurs in spring [14]. Combined with the seasonal patterns, it can be inferred that *A agrarius* is the main vector in the midtemperate zone, while *A agrarius* and *R norvegicus* are the main vectors in the warm-temperate and subtropical zones. The main hosts of HTNV and Seoul virus (SEOV) are *A agrarius* and *R norvegicus*, respectively. It can be inferred that HTNV is epidemic in the midtemperate zone, while HTNV and SEOV are mainly epidemic in the warm-temperate and subtropical zones. Prior studies investigated sites and identified HTNV in the midtemperate Heilongjiang Province, the warm-temperate Qingdao City, and the subtropical Jiangxi Province, which is consistent with our inferences [20-22]. Moreover, low temperatures prolong the survival of the virus outside the host, allowing the virus to remain infectious even in the absence of direct rodent contact or rodent-to-human contact [23].

The results of the DLNM showed that the risk of climate-related HFRS varied from place to place with different lag months. The lagged effects of climate variables in different temperature zones may be related to the dominant rodent population, breeding and living conditions, HTNV-positive rate, and human contact frequency [24,25]. We found that the maximum lag effects of temperature on HFRS incidence were 1 month, 0 months, and
0 months from northern to southern China. However, inconsistent findings on the lag time with maximum effects have also been reported by Cao et al [14]. The lag effects of temperature were 1 month (midtemperate zone), 2 months (warm-temperate zone), and 3 months (subtropical zone), respectively. Our results showed that the hysteresis effect of meteorological factors did not end after one epidemic season, but gradually weakened in the following 1 to 2 epidemic seasons, and the duration of the hysteresis effect varied by region, indicating the necessity of continuous intervention after the HFRS epidemic. At the same time, it also provides theoretical support for the important role of weather variability, especially temperature, in the propagation of HFRS.

Consistent with previous studies, we found that high precipitation and relative humidity are risk conditions for HFRS (Multimedia Appendix 1, Figure S6) [15,26]. Wet conditions and high relative humidity are good for rodents to survive or breed. Adequate rainfall provides a suitable environment and sufficient food for rodents, which ultimately increases the risk of virus transmission [27]. Higher relative humidity affects the spread of HFRS by affecting the infectivity and stability of HTNV in vitro, which is consistent with the fact that HFRS epidemic areas are mostly located in humid or semihumid mountainous areas [28,29]. The interaction and stratification analysis showed that in low-temperature environments, more precipitation and higher relative humidity were climate risk factors for HFRS occurrence, which is consistent with previous studies. Zhang et al [9] found that average temperature, relative humidity, and precipitation interacted with HFRS through stratified analysis; the risk of HFRS was inversely proportional to average temperature and directly proportional to relative humidity.

There are several key limitations to our study that should be acknowledged. First, in addition to meteorological factors, other factors may also affect the occurrence of HFRS, including vaccination programs, economic factors, health care level, and host animal diversity. Second, HFRS case data comes from the notifiable infectious disease detection system, and there are cases of underreporting. For example, patients with mild symptoms may self-isolate at home, which would lead us to underestimate the impact of meteorological factors on HFRS. Third, our data cannot distinguish which viruses caused the HFRS cases, nor can it be targeted to study the relationship between different viruses and climate. Therefore, future research should explore the relationship between HFRS incidence and the different viruses that cause HFRS and explore the intersection between the COVID-19 pandemic and HFRS to fully understand the broader implications for public health.

**Conclusions**

Using data for HFRS cases from 151 cities, we provide first-hand evidence of the interaction and stratification effects of meteorological factors on HFRS in different regions of China. Furthermore, the magnitude and timing of hysteresis effects varied across climate zones. Our findings indicate that low temperature positively influences the long-term incidence of HFRS. The results of this study can provide a valuable scientific basis for public health departments to formulate targeted HFRS interventions, understand the relationship between weather and HFRS, and use low temperature as an early warning signal to carry out HFRS control and outbreak response.

**Acknowledgments**

We thank the Chinese Center for Disease Control and Prevention staff for their contributions to data collection. This study was supported by fundamental research projects (JK2023GK001) and Jiangsu social development project (M2020087, BE2022682, BK20221196).

**Data Availability**

The meteorological data are available from the National Meteorological Information Center [18] and HFRS case data are available from the Chinese Center for Disease Control and Prevention [30]. The data set analyzed in this study is available from the corresponding author on reasonable request.

**Authors' Contributions**

YL, LZ, YX, and WL collected the data and performed the analysis of the data; YW, LL, and JR conceived the study; YL, LZ, and QK wrote the manuscript; and QS, XL and WT reviewed and finalized the manuscript. All authors contributed to the article and approved the submitted version.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Supplementary tables and figures.

[DOC File, 1797 KB - publichealth_v10i1e52221_app1.doc ]

**References**


Abbreviations
- DLNM: distributed lag nonlinear model
- HFRS: hemorrhagic fever with renal syndrome
- HTNV: Hantavirus
- QAIC: quasi-Akaike information criterion
- QBIC: quasi-Bayesian information criterion
- RR: relative risk
- SEOV: Seoul virus

©Yizhe Luo, Longyao Zhang, Yameng Xu, Qiuyuan Kuai, Wenhao Li, Yifan Wu, Licheng Liu, Jiarong Ren, Lingling Zhang, Qifang Shi, Xiaobo Liu, Weilong Tan. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 05.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Population Behavior Changes Underlying Phasic Shifts of SARS-CoV-2 Exposure Settings Across 3 Omicron Epidemic Waves in Hong Kong: Prospective Cohort Study

Chin Pok Chan¹, BSc; Shui Shan Lee²,³, MD; Tsz Ho Kwan¹,²,³, PhD; Samuel Yeung Shan Wong¹, MD; Eng-Kiong Yeoh¹,⁴, MBBS; Ngai Sze Wong¹,²,³, PhD

¹JC School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China
²Stanley Ho Centre for Emerging Infectious Diseases, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China
³S.H. Ho Research Centre for Infectious Diseases, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China
⁴Centre for Health Systems and Policy Research, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

Corresponding Author:
Ngai Sze Wong, PhD
JC School of Public Health and Primary Care
Faculty of Medicine
The Chinese University of Hong Kong
Rm 204, School of Public Health Building
Prince of Wales Hospital, New Territories
Hong Kong
China
Phone: 86 22528862
Email: candy_wong@cuhk.edu.hk

Abstract

Background: Exposure risk was shown to have affected individual susceptibility and the epidemic spread of COVID-19. The dynamics of risk by and across exposure settings alongside the variations following the implementation of social distancing interventions are understudied.

Objective: This study aims to examine the population’s trajectory of exposure risk in different settings and its association with SARS-CoV-2 infection across 3 consecutive Omicron epidemic waves in Hong Kong.

Methods: From March to June 2022, invitation letters were posted to 41,132 randomly selected residential addresses for the recruitment of households into a prospective population cohort. Through web-based monthly surveys coupled with email reminders, a representative from each enrolled household self-reported incidents of SARS-CoV-2 infections, COVID-19 vaccination uptake, their activity pattern in the workplace, and daily and social settings in the preceding month. As a proxy of their exposure risk, the reported activity trend in each setting was differentiated into trajectories based on latent class growth analyses. The associations of different trajectories of SARS-CoV-2 infection overall and by Omicron wave (wave 1: February-April; wave 2: May-September; wave 3: October-December) in 2022 were evaluated by using Cox proportional hazards models and Kaplan-Meier analysis.

Results: In total, 33,501 monthly responses in the observation period of February-December 2022 were collected from 5321 individuals, with 41.7% (2221/5321) being male and a median age of 46 (IQR 34-57) years. Against an expanding COVID-19 vaccination coverage from 81.9% to 95.9% for 2 doses and 20% to 77.7% for 3 doses, the cumulative incidence of SARS-CoV-2 infection escalated from <0.2% to 25.3%, 32.4%, and 43.8% by the end of waves 1, 2, and 3, respectively. Throughout February-December 2022, 52.2% (647/1240) of participants had worked regularly on-site, 28.7% (356/1240) worked remotely, and 19.1% (237/1240) showed an assorted pattern. For daily and social settings, 4 and 5 trajectories were identified, respectively, with 11.5% (142/1240) and 14.6% (181/1240) of the participants gauged to have a high exposure risk. Compared to remote working, working regularly on-site (adjusted hazard ratio [aHR] 1.47, 95% CI 1.19-1.80) and living in a larger household (aHR 1.12, 95% CI 1.06-1.18) were associated with a higher risk of SARS-CoV-2 infection in wave 1. Those from the highest daily exposure risk trajectory (aHR 1.46, 95% CI 1.07-2.00) and the second highest social exposure risk trajectory (aHR 1.52, 95% CI 1.18-1.97) were also at an increased risk of infection in waves 2 and 3, respectively, relative to the lowest risk trajectory.

Conclusions: In an infection-naive population, SARS-CoV-2 transmission was predominantly initiated at the workplace, accelerated in the household, and perpetuated in the daily and social environments, as stringent restrictions were scaled down.
Introduction

In 2020-2021, Hong Kong’s COVID-19 burden in the population was among the lowest in the Asia Pacific region (<1%). Ranging from border control and mandatory mask wearing to social activity restrictions, the implementation of various nonpharmaceutical interventions (NPIs) contributed to the successful control initially. It was not until February 2022 when the SARS-CoV-2 Omicron outbreak swept through the city and caused over 2.6 million cases (35%) in a 7.4 million population by the end of the year [1]. Albeit the epidemic becoming more controllable today, it is still unclear what factors had contributed to the initial surge of Omicron and subsequent waves of infection. Since the effectiveness of vaccinations against emerging variants is waning, investigations into the impact of the population’s behavior on virus transmission would therefore be crucial [2].

Exposure risk and its association with SARS-CoV-2 infection have been examined from various perspectives. Previous studies have analyzed the social contact and mobility pattern to model the change in the transmission dynamics [3,4]. However, the application of these metrics as a proxy for aggregated exposure risks has ignored the heterogeneity of the exposure profiles across the population [5]. The distribution of population activities in various community settings is often overlooked. Former research has presented a remarkable difference in the infection rate following SARS-CoV-2 exposure to a household and health care source [6,7]. Thus, the nature of social contacts in different settings could also play an important role in affecting the level of virus exposure and infection risk. Based on contact-tracing data, the transmission dynamics of SARS-CoV-2 can also be observed to be largely variable among clusters of infection emerging from different environments [8-10]. Yet, evidence formally appraising the association between exposure and outbreaks in a broader community context remains scarce.

As an integral part of NPIs, different combinations of social distancing interventions were implemented worldwide to tackle the COVID-19 pandemic. The use of regulatory restriction was shown to have effectively blocked transmission chains associated with social events [11], while prohibiting dine-in services and imposing a work from home (WFH) order were deemed to have rapidly brought down the scale of virus spread [12]. A multi-country analysis in 2020-2022 also revealed the fluctuation in social contact patterns in different settings during the pandemic and highlighted its connection with different containment measures [13]. The above evidence conjectured that the whereabouts and dynamics of virus exposure could be influenced by the coverage and stringency of social distancing measures [14]. Yet, their interplay was still vaguely understood as the epidemic progressed through waves driven by different SARS-CoV-2 variants in the real world.

In recent decades, participatory surveillance has been increasingly used as a complementary tool to detect exposure and disease patterns in community outbreaks [15,16]. Without imposing a complete lockdown, the comparison between infected and uninfected individuals in their activity patterns by setting would enable the identification of the changing hot spots for SARS-CoV-2 exposure as the epidemic unfolds. Against these backgrounds, this study aims to examine the dynamics of population exposure risks in different settings and to evaluate its association with SARS-CoV-2 infection across the 3 Omicron epidemic waves in 2022 in Hong Kong.

Methods

Participant Recruitment and Study Design

In the form of a prospective cohort, an ongoing web-based participatory surveillance platform was set up in Hong Kong with territory-wide recruitment in March-June 2022. Taking individual households as a sampling unit, cluster sampling was performed based on a sampling frame called “building groups”—a demarcation system codeveloped by the Census and Statistics Department of the Hong Kong Government and the Centamap Company Limited for grouping local households with a similar socioeconomic background for census data analysis. The demarcation system, updated as of 2016, with 3083 defined building groups (strata) covering 165,965 residential buildings, was used as the sampling frame of this study. In each stratum resided by at least 1000 individuals, 12-14 households were randomly selected to receive an invitation letter bearing a unique code mapped to their residential address. A representative household member aged 18 years or older, residing in Hong Kong, and who understood Chinese or English was invited to participate. Following registration on the platform by using the unique household code, participants who consented were asked to fill in a baseline questionnaire followed by monthly updates during the follow-up period.

Ethics Approval

Ethics approval was sought from the Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong (SBRE-21-0048). Informed e-consent was obtained from each participant at the time of registration. The survey data set was deidentified by removing participants’ residential addresses. Separately, residential addresses used for recruitment and incentive delivery were stored in another data set with matched study IDs to preserve anonymity. All data sets were password-protected and only accessed by designated research staff and investigators. Upon completion of the baseline questionnaire, participants were entitled to a HK $50 cash
voucher (HK $7.8=US $1). During follow-up, additional HK $50 vouchers would be offered to participants who had completed at least 3 monthly updates every 6 months.

**Survey Instrument**

Available in both Chinese and English, the baseline questionnaire could be completed in 20 minutes and covered the participants’ particulars on (1) sociodemographic characteristics; (2) household size; (3) chronic illness status; (4) baseline history of SARS-CoV-2 infection and COVID-19 vaccination; (5) frequency of rapid antigen/nucleic acid testing for SARS-CoV-2 in the previous month; (6) monthly pattern of work, daily, and social activities; (7) nature of workplace (health care/non–health care) and the number of workplace contacts; (8) perceived risk of SARS-CoV-2 exposure (on an 11-point Likert scale from 0 to 10); and (9) suspected presence of infected individuals among daily contacts (yes/no). For the ensuing monthly updates, participants were asked to self-report any COVID-19 vaccine uptake and incident of SARS-CoV-2 infection, defined by a positive result from a polymerase chain reaction or rapid antigen test regardless of the presence of symptoms. Updates on sections 5-9 and household infection status based on the situation in the previous month were also requested. A personalized survey link was sent to each participant by email on the first of each month and made accessible for 14 days. Paper questionnaires were supplied upon request (n=23).

**Definition of Exposure Risk and Tiers of Social Distancing Policy**

In this study, exposure risk was defined as the intensity of behavior leading to potential virus exposure, while its dynamics were examined in 4 contexts: household, workplace, social setting, and daily setting. Persons living in larger households and working entirely outside the home in a month were expected to have a greater exposure risk. For daily and social settings, the exposure risk was measured by their frequency of engagement in different types of daily, leisure, and social activities. Details of the exposure risk definition are illustrated in Table 1. Based on the stringency of the regulations imposed on the operations of catering businesses, bars/pubs, and a range of high-risk premises and activities; restrictions on social gatherings; and government recommendations on WFH practices, the evolvement of the local social distancing policies was divided into 4 tiers. The details of the regulations together with other NPIs are delineated in Figure S1 in Multimedia Appendix 1.

### Table 1. Study definitions of exposure risk in different settings.

<table>
<thead>
<tr>
<th>Exposure setting</th>
<th>Proxy question for exposure risk</th>
<th>Measurements used in analyses</th>
<th>Assumption on exposure risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household</td>
<td>Number of coliving members in the household at baseline</td>
<td>Continuous variable: household size/ordinal variable with 4 levels: 1 person (living alone); 2 persons; 3-4 persons; ≥5 persons</td>
<td>Constant throughout due to repeated contact with family members; higher exposure risk in larger families</td>
</tr>
<tr>
<td>Workplace</td>
<td>Reports on whether one had practiced WFH completely, intermittently, or worked entirely outside home in each month</td>
<td>Ordinal variable with 3 levels: Completely WFH/not working; Intermittently WFH; Work entirely outside home</td>
<td>Varied; higher exposure risk in those who worked outside the home for the entire month</td>
</tr>
<tr>
<td>Daily</td>
<td>Daily activity*: reported number of days one had (1) visited an eatery, (2) visited a shopping mall/market, and (3) used public transportation last week in each month (scale of 1: never; 2: 1 day; 3: 2-3 days; 4: 4-6 days; 5: everyday) Leisure activity: reports on whether one had ever visited sauna/bathhouse, massage/beauty parlor, fitness center, beach/pool, sport facility, or entertainment venue such as a cinema in each month</td>
<td>Ordinal variable with 4 levels: Inactive and without leisure activity; Inactive and with leisure activity; Active and without leisure activity; Active and with leisure activity</td>
<td>Varied; higher exposure risk in those who had a more active daily lifestyle and participated in leisure activity</td>
</tr>
<tr>
<td>Social</td>
<td>Reported number of times one had paid visit to relatives/friends’ home or vice versa, dined out for the purpose of gathering, gone to bar/club, karaoke room, party room, attended banquet, went on a local hotel vacation, and outdoor travel (eg, hiking/camping) in each month</td>
<td>Ordinal variable with 4 levels: None; 1-3 times; 4-7 times; ≥8 times</td>
<td>Varied; higher exposure risk in those who had attended more social events</td>
</tr>
</tbody>
</table>

*WFH: work from home.

*The frequencies for each of the daily activities were standardized and averaged for each participant. Those with a computed average higher than the population median were classified as active or otherwise inactive for a particular month.
Statistical Analysis
Participants’ baseline characteristics were summarized using descriptive statistics. The geographical representativeness of the recruited households was then assessed by comparison with the census distribution [17]. The cumulative incidence of SARS-CoV-2 infection and monthly proportion of participants reporting an infection were profiled against the number documented by the government, while age and sex adjustments were performed in reference to the population demographics [17]. The patterns of age-specific COVID-19 burden and vaccination coverage were also illustrated.

Based on the reported work pattern and computed exposure risk level in daily and social settings, the temporal dynamics of each were differentiated using a latent class growth model—a mixture model for identifying underlying subgroups exhibiting similar growth trajectories [18]. Only participants with a missing response in at most 1 month during each epidemic wave were included in the analysis. Individual responses of specific months, during which SARS-CoV-2 infection was reported in themselves or their household, were also eliminated so that the biased activity level resulting from isolation/quarantine was not considered. A 2-stage optimization of maximum likelihood was adopted to model the intercept, linear, and quadratic growth factor, with an increasing number of latent classes from 1 to 7. Goodness of fit was indicated by the lowest Bayesian information criterion (BIC) and an entropy of >0.8. Elbow plot was used in cases when the stated criteria were still met in the 7-class model. Based on the selected models, the class-specific exposure risk trajectories were plotted on a probability scale. Factors associated with each trajectory were determined by a multinomial mixed effects model.

The time patterns of SARS-CoV-2 infection among participants with different household sizes, work patterns, and daily and social exposure risk trajectories were examined using Kaplan-Meier analysis, while the difference in the infection rate in each epidemic wave was compared using relative risk (RR). Their relationships with SARS-CoV-2 infection risk were then parameterized using Cox proportional hazards models adjusted for sociodemographic characteristics and vaccination status before the Omicron outbreak. Subgroup analysis was also conducted between the subset of participants with different pre-Omicron vaccination statuses. R software (version 4.1.2; R Foundation for Statistical Computing) and Mplus (version 8.8; Muthen & Muthen) were used to perform statistical analyses. All tests were 2-tailed, and the significance was denoted by P<.05.

Results

Baseline Characteristics
Between March and June 2022, representative participants from 5321 households were recruited out of 41,132 invitations (response rate 12.9%). Concerning the observation period from February to December 2022, over 3000 survey entries were received each month except for the first 2 months of recruitment, amounting to a total of 33,501 responses. Two-thirds (3553/5321, 66.8%) of the recruited participants completed at least 50% of the monthly surveys, with 58.3% (3100/5321) being female, and the median age was 46 (IQR 34-57) years. A majority of the participants were in full-time employment or self-employed (3208/5321, 60.3%) and had attained tertiary education or above (3325/5321, 62.5%) (Table 2). The recruited households had a mean size of 3.1 (SD 1.5) persons, and 11.5% (607/5321) were single-person households. The geographical distribution of the recruited households was within a 1% difference compared to the census data for all 18 districts in the territory of Hong Kong (Table S1 in Multimedia Appendix 1).
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2221 (41.7)</td>
</tr>
<tr>
<td>Female</td>
<td>3100 (58.3)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>46 (34-57)</td>
</tr>
<tr>
<td>18-29, n (%)</td>
<td>732 (13.8)</td>
</tr>
<tr>
<td>30-39, n (%)</td>
<td>1188 (22.3)</td>
</tr>
<tr>
<td>40-49, n (%)</td>
<td>1180 (22.2)</td>
</tr>
<tr>
<td>50-64, n (%)</td>
<td>1614 (30.3)</td>
</tr>
<tr>
<td>≥65, n (%)</td>
<td>607 (11.4)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Local (Hong Kong) Chinese</td>
<td>5190 (97.5)</td>
</tr>
<tr>
<td>Nonlocal Chinese/other ethnicities</td>
<td>131 (2.5)</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Primary education or below</td>
<td>175 (3.3)</td>
</tr>
<tr>
<td>Secondary education</td>
<td>1821 (34.2)</td>
</tr>
<tr>
<td>Diploma/associate degree</td>
<td>803 (15.1)</td>
</tr>
<tr>
<td>Bachelor’s degree or above</td>
<td>2522 (47.4)</td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Full-time employment/self-employed</td>
<td>3208 (60.3)</td>
</tr>
<tr>
<td>Part-time/temporary employment</td>
<td>288 (5.4)</td>
</tr>
<tr>
<td>Student</td>
<td>251 (4.7)</td>
</tr>
<tr>
<td>Homemaker</td>
<td>406 (7.6)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>350 (6.6)</td>
</tr>
<tr>
<td>Retired</td>
<td>816 (15.3)</td>
</tr>
<tr>
<td><strong>Monthly income</strong></td>
<td></td>
</tr>
<tr>
<td>a range (n=3739; missing: n=8), n (%)</td>
<td></td>
</tr>
<tr>
<td>&lt;HK $10,000</td>
<td>434 (11.6)</td>
</tr>
<tr>
<td>HK $10,000-19,999</td>
<td>895 (23.9)</td>
</tr>
<tr>
<td>HK $20,000-29,999</td>
<td>759 (20.3)</td>
</tr>
<tr>
<td>HK $30,000-59,999</td>
<td>972 (26)</td>
</tr>
<tr>
<td>≥HK $60,000</td>
<td>410 (11)</td>
</tr>
<tr>
<td>Refuse to answer</td>
<td>269 (7.2)</td>
</tr>
<tr>
<td><strong>Reported chronic illness</strong> (unsure: n=246), n (%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3638 (71.7)</td>
</tr>
<tr>
<td>Yes</td>
<td>1437 (28.3)</td>
</tr>
<tr>
<td><strong>Household size (missing: n=62)</strong></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.1 (1.5)</td>
</tr>
<tr>
<td>1 (living alone), n (%)</td>
<td>607 (11.5)</td>
</tr>
<tr>
<td>2 persons, n (%)</td>
<td>1478 (28.1)</td>
</tr>
<tr>
<td>3-4 persons, n (%)</td>
<td>2264 (43.1)</td>
</tr>
<tr>
<td>5-6 persons, n (%)</td>
<td>805 (15.3)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Values, n (%)</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>≥7 persons, n (%)</td>
<td>105 (2)</td>
</tr>
</tbody>
</table>

*aHK $7.8=US $1.
*bIncluding any diagnosed conditions that require long-term clinical follow-up or medications, such as hypertension, diabetes mellitus, stroke, and asthma.

**SARS-CoV-2 Burden and COVID-19 Vaccination Uptake**

As of December 2022, a total of 2332 SARS-CoV-2 infections were reported, adding up to a cumulative incidence of 43.8% since 2020 (Figure 1). The age- and sex-adjusted estimate was 42.8%, which was 7% higher than the government figure. The number of incident infections peaked in March 2022 (273/1718, 15.9%), reached the trough in May (13/3505, 0.4%), rebounded in August (127/3208, 3.9%), and surged in December (410/3206, 13.5%), giving rise to 3 Omicron waves in 2022 (wave 1: February to April, wave 2: May to September, and wave 3: October to December). Overall, hospitalization was required in 20 of 2332 cases (0.9%), while reinfection with at least 3 months from the previous infection episode was reported in 45 (1.9%) participants. Stratified by age group, individuals aged 30-39 years had the highest cumulative incidence of 48.8% (580/1188), while the population aged ≥65 years was least affected by the epidemic (226/607, 37.2%). Over 80% (4231/5167) and 20% (1043/5208) of the participants had received at least 2 and 3 doses of COVID-19 vaccine before February 2022, respectively. Correspondingly, the 2 coverages escalated to 95.9% (4954/5167) and 77.7% (4047/5208) at the end of the year after the administration of vaccine passes requiring 2 and 3 doses by May and June (Figure S1 in Multimedia Appendix 1). A noticeably slower third dose uptake was observed throughout in younger age groups.
Activity Pattern in Workplace, Daily Settings, and Social Settings

When the Omicron outbreak first broke out in February 2022, the most stringent tier 1 social distancing measures were in place. In reference to the government advice, 26.8% (656/2444) and 49% (1198/2444) reported intermittent and complete WFH practice (or not doing any work), while one-quarter of the participants (590/2444, 24.1%) continued to work on-site in their workplace. Due to the mass closure of community facilities, almost none had reported participation in leisure activities in daily settings. Half of the participants (1202/2444, 49.2%) refrained from taking part in any social activity, whereas 22.7% (555/2444), 16% (392/2444), and 12.1% (295/2444) still went out for 1-3 times, 4-7 times, and ≥8 times, respectively, despite the restrictions, with a mean gathering size of 3.4 persons.
Following the tier 2 relaxation in May 2022, the proportion working on-site returned to 54.1% (1899/3505), with 11% (384/3505) maintaining their intermittent WFH practice. Based on their reported frequency of visiting eateries and shopping malls/markets, and using public transportation, living an active lifestyle with engagement in leisure activity was acknowledged in 21.8% (763/3505) of the participants, compared to 31% (1087/3505) of the participants who were considered to be inactive. Social activity was also restored, with 24% (839/3502) of the participants reporting engagement for ≥ 8 times per month in May. The overall pattern of work, daily, and social activities remained stable, as social distancing measures continued to ease, while the average gathering size increased from 5.1 in May to 9.6 persons in December 2022.

**Trajectories of Work Pattern, Daily, and Social Exposure Risk and the Associated Characteristics**

Excluding those with inadequate follow-up responses amid the 3 waves, a total of 1240 participants were included in latent class growth analyses. This subgroup was similar to the rest, except that slightly more of those being excluded were local (Hong Kong) Chinese (1222/1240, 98.5% vs 3968/4081, 97.2%; \( P=0.01 \)) with a higher median age (48 in those included vs 45 years in those excluded from the latent growth class analyses; \( P<0.001 \); Table S2 in Multimedia Appendix 1).

To differentiate between work patterns, the best fit was demonstrated in the 3-class latent class growth model (BIC 12,308.19, entropy 0.957; Table S3 and Figure S2 in Multimedia Appendix 1). Approximately 28.7% (356/1240) of the participants were not working or had worked remotely throughout the year (class 1), while about half (class 3: 647/1240, 52.2%) regularly worked outside the home, with 19.1% (237/1240) following an assorted work pattern (class 2; Figure 2).

**Figure 2.** Trajectories of the work pattern and exposure risk in daily and social settings by latent class growth models. Compartments divided by the dashed line indicate the 3 periods of Omicron waves in 2022: wave 1 (February to April), wave 2 (May to September), and wave 3 (October to December).
were significantly more likely to be in full-time employment (class 2: adjusted odds ratio [aOR] 2.05, 95% CI 1.13-3.73; class 3: aOR 3.22, 95% CI 1.76-5.89; class 4: aOR 2.81, 95% CI 1.48-5.37) and self-perceive a higher chance of virus exposure (class 2: aOR 1.18, 95% CI 1.06-1.30; class 3: aOR 1.21, 95% CI 1.09-1.34; class 4: aOR 1.31, 95% CI 1.16-1.47; Table 3). Significantly more class 4 participants had received tertiary education or above (aOR 2.28, 95% CI 1.18-4.39) and undergone SARS-CoV-2 testing more frequently (aOR 1.03, 95% CI 1.00-1.06).

Table 3. Multinomial mixed effects model on factors associated with daily exposure risk trajectories.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Class 1 (minimal and nonadaptive), aOR (95% CI)</th>
<th>Class 2 (low and nonadaptive), aOR (95% CI)</th>
<th>Class 3 (moderate and nonadaptive), aOR (95% CI)</th>
<th>Class 4 (high and adaptive), aOR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Reference 1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Reference 0.93 (0.53-1.61)</td>
<td>1.81 (1.03-3.17)*</td>
<td>1.94 (1.08-3.46)*</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>Reference 1.01 (0.99-1.04)</td>
<td>1.03 (1.01-1.06)**</td>
<td>1.06 (1.03-1.09)**</td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary education or below</td>
<td>Reference 1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Tertiary education or above</td>
<td>Reference 1.09 (0.59-2.02)</td>
<td>0.95 (0.51-1.77)</td>
<td>2.28 (1.18-4.39)*</td>
<td></td>
</tr>
<tr>
<td>Full-time worker</td>
<td>Reference 2.05 (1.13-3.73)*</td>
<td>3.22 (1.76-5.89)**</td>
<td>2.81 (1.48-5.37)**</td>
<td></td>
</tr>
<tr>
<td>Household size</td>
<td>Reference 1.07 (0.88-1.31)</td>
<td>1.11 (0.91-1.35)</td>
<td>1.08 (0.88-1.33)</td>
<td></td>
</tr>
<tr>
<td>Reported chronic illness</td>
<td>Reference 1.08 (0.56-2.07)</td>
<td>0.93 (0.48-1.79)</td>
<td>1.47 (0.76-2.87)</td>
<td></td>
</tr>
<tr>
<td>Monthly frequency of SARS-CoV-2 testing</td>
<td>Reference 1.01 (0.99-1.04)</td>
<td>1.03 (1.00-1.05)</td>
<td>1.03 (1.00-1.06)*</td>
<td></td>
</tr>
<tr>
<td>Suspected SARS-CoV-2 infection among daily contacts</td>
<td>Reference 1.12 (0.77-1.62)</td>
<td>1.17 (0.81-1.70)</td>
<td>1.38 (0.91-2.08)</td>
<td></td>
</tr>
<tr>
<td>Perceived risk of SARS-CoV-2 exposure (range 0-10)</td>
<td>Reference 1.18 (1.06-1.30)**</td>
<td>1.21 (1.09-1.34)**</td>
<td>1.31 (1.16-1.47)**</td>
<td></td>
</tr>
</tbody>
</table>

aOR: adjusted odds ratio.

For social exposure risk, the best model fit was observed in the 5-class model (BIC 25,503.3, entropy 0.816). Most participants (411/1240, 33.1%) belonged to the “class 2: low and nonadaptive” group, followed by “class 1: minimal and nonadaptive” (260/1240, 21%), “class 3: moderate and adaptive” (181/1240, 14.6%), “class 4: moderate and nonadaptive” (207/1240, 16.7%), and “class 5: high and adaptive” (181/1240, 14.6%). Participants with a higher social exposure risk were in general older (class 2: aOR 1.04, 95% CI 1.01-1.06; class 3: aOR 1.05, 95% CI 1.02-1.07; class 4: aOR 1.07, 95% CI 1.04-1.09; class 5: aOR 1.06, 95% CI 1.04-1.09) and more educated (class 3: aOR 3.70, 95% CI 1.95-7.01; class 4: aOR 4.57, 95% CI 2.35-8.73; class 5: aOR 5.42, 95% CI 2.81-10.46; Table 4). Female behaviors appeared to be more responsive to adjustments in social distancing policy (class 3: aOR 2.30, 95% CI 1.30-4.09; class 5: aOR 2.09, 95% CI 1.17-3.74).
JMIR PUBLIC HEALTH AND SURVEILLANCE

Chan et al

Table 4. Multinomial mixed effects model on factors associated with social exposure risk trajectories.
Variables

Class 1 (minimal
and nonadaptive)

Class 2 (low and nonadap- Class 3 (moder- Class 4 (moderate and
ate and adaptive), nonadaptive), aOR
tive), aORa (95% CI)
aOR (95% CI)
(95% CI)

Class 5 (high and adaptive),
aOR (95% CI)

Male

Reference

1.00

1.00

1.00

1.00

Female

Reference

1.69 (0.96-2.99)

2.30 (1.304.09)**

1.67 (0.94-2.98)

2.09 (1.17-3.74)*

Age (years)

Reference

1.04 (1.01-1.06)**

1.05 (1.021.07)***

1.07 (1.04-1.09)***

1.06 (1.04-1.09)***

Secondary educa- Reference
tion or below

1.00

1.00

1.00

1.00

Tertiary education or above

Reference

1.69 (0.90-3.19)

3.70 (1.957.01)***

4.57 (2.39-8.73)***

5.42 (2.81-10.46)***

Full-time worker

Reference

1.75 (0.94-3.25)

0.99 (0.54-1.83)

1.41 (0.75-2.65)

0.67 (0.36-1.25)

Household size

Reference

1.04 (0.85-1.28)

1.13 (0.92-1.38)

0.95 (0.78-1.17)

0.92 (0.75-1.13)

Reported chronic illness

Reference

0.77 (0.39-1.52)

1.20 (0.62-2.34)

1.06 (0.54-2.07)

0.74 (0.38-1.44)

Monthly frequency of Reference

1.00 (0.97-1.02)

1.00 (0.97-1.03)

0.99 (0.97-1.02)

1.00 (0.97-1.03)

Reference

1.23 (0.85-1.78)

1.41 (0.94-2.10)

1.38 (0.93-2.04)

1.59 (1.06-2.37)*

Reference

1.02 (0.92-1.14)

1.07 (0.96-1.20)

1.07 (0.96-1.20)

1.08 (0.96-1.21)

Sex

Education level

b

SARS-CoV-2 testing
Suspected SARSCoV-2 infection

among daily contactsb
Perceived risk of
SARS-CoV-2 exposure (range 0-10)b
a

aOR: adjusted odds ratio.

b

Time-varying predictor with the participant set to be the random effect.
*P<.05; **P<.01; ***P<.001.

Timing and Risk Factors of SARS-CoV-2 Infection
Overall, there was a significant heterogeneity in SARS-CoV-2
infection risk among groups of varying household size
(χ24=28.6; P<.001), work pattern (χ23=25.6; P<.001), and daily
exposure risk trajectories (χ23=11.6; P=.009; Figure 3). When
the results were adjusted for sex, age, education, work pattern,
and pre-Omicron COVID-19 vaccination status, participants
living in a larger household (adjusted hazard ratio [aHR] 1.12,
95% CI 1.08-1.20) and being a health care worker (aHR 1.38,
95% CI 1.29-2.26) were linked to a significantly higher infection
risk (Table 5). Those who worked regularly on-site (aHR 1.47,
95% CI 1.19-1.80) and had more workplace contacts (hazard
ratio [HR] 1.01, 95% CI 1.00-1.02) with a higher daily (class
2: aHR 1.38, 95% CI 1.07-1.77; class 3: aHR 1.49, 95% CI
1.16-1.91; class 4: aHR 1.46, 95% CI 1.07-2.00) and social
exposure risk (class 4: aHR 1.52, 95% CI 1.18-1.97) were also
more prone to SARS-CoV-2 infection. Stratified by the epidemic
wave, individuals living in a household with size of 3 (RR 1.77;
P=.01), 4 (RR 2.12; P<.001), and ≥5 persons (RR 2.05; P=.002)

https://publichealth.jmir.org/2024/1/e51498

XSL• FO
RenderX

had a higher risk of infection in wave 1 compared to those living
alone. A similar association was also identified in those who
regularly worked outside the home (RR 1.73; P<.001) and daily
exposure risk trajectory class 3 participants (RR 1.52; P=.008).
Significantly more daily exposure risk trajectory class 4
participants had acquired SARS-CoV-2 infection in wave 2 (RR
2.41; P=.007). In contrast, an excess of social exposure risk
trajectory class 3 (RR 1.60; P=.03) and class 5 (RR 1.64; P=.02)
participants became infected in wave 3.
Stratified by the COVID-19 vaccination status, a higher
cumulative incidence was reported in participants who had
received 0-2 doses before the Omicron outbreak (515/946,
54.4% vs 112/262, 42.7%). In this subgroup, household,
workplace, daily, and social exposure risk remained significantly
associated with an episode of SARS-CoV-2 infection. However,
for the rest who had received 3 doses or more, only larger
household size (HR 1.13, 95% CI 1.01-1.28) and regular on-site
work pattern (HR 1.56, 95% CI 1.03-1.36) were significant risk
factors in the unadjusted model. The effect of daily exposure
risk was, however, absent.

JMIR Public Health Surveill 2024 | vol. 10 | e51498 | p.2171
(page number not for citation purposes)


Figure 3. Results of Kaplan-Meier analysis on the timing of the SARS-CoV-2 infection among groups of varying household size, work pattern, daily, and social exposure risk trajectories. An event was defined by an incident of SARS-CoV-2 infection. Relative risks based on the comparison of SARS-CoV-2 infection rates were computed among different groups at each Omicron wave (wave 1: February to April; wave 2: May to September; and wave 3: October to December) in 2022. The significance value of the log rank test is also displayed in each panel. *P<.05, **P<.01, ***P<.001.
Table 5. Cox proportional-hazards model on the risk of SARS-CoV-2 infection and subgroup analyses.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall (n=1240)(^a)</th>
<th>Pre-Omicron vaccination status: 0-2 doses (n=946)(^b)</th>
<th>Pre-Omicron vaccination status: ≥3 doses (n=262)(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>aHR (95% CI)</td>
<td>HR (95% CI)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Female</td>
<td>1.10 (0.94-1.29)</td>
<td>1.19 (1.00-1.43)</td>
<td>0.81 (0.56-1.18)</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>0.99 (0.98-0.99)(^***)</td>
<td>0.99 (0.99-1.00)**</td>
<td>0.98 (0.97-1.00)*</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary education or below</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Tertiary education or above</td>
<td>1.06 (0.90-1.25)</td>
<td>1.10 (0.91-1.32)</td>
<td>0.90 (0.62-1.31)</td>
</tr>
<tr>
<td>Health care worker</td>
<td>1.70 (1.29-2.26)(^***)</td>
<td>1.74 (1.26-2.40)(^***)</td>
<td>1.46 (1.05-2.03)*</td>
</tr>
<tr>
<td>Number of workplace contacts in February 2022 (n=577)</td>
<td>1.01 (1.00-1.02)(^**)</td>
<td>1.01 (1.00-1.02)*</td>
<td>1.01 (0.96-1.07)</td>
</tr>
<tr>
<td>Household size (missing=22)</td>
<td>1.14 (1.08-1.20)(^***)</td>
<td>1.13 (1.07-1.20)(^***)</td>
<td>1.13 (1.01-1.28)*</td>
</tr>
<tr>
<td>COVID-19 vaccination status prior to Omicron outbreak (missing=32)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unvaccinated/1 dose</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2 doses</td>
<td>0.91 (0.74-1.13)</td>
<td>1.06 (0.84-1.31)</td>
<td>1.04 (0.78-1.31)</td>
</tr>
<tr>
<td>≥3 doses</td>
<td>0.65 (0.49-0.84)(^**)</td>
<td>1.02 (0.80-1.30)</td>
<td>1.02 (0.80-1.30)</td>
</tr>
<tr>
<td>Work pattern</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No work/remote</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Flexible</td>
<td>1.13 (0.88-1.44)</td>
<td>1.10 (0.81-1.41)</td>
<td>1.03 (0.58-1.84)</td>
</tr>
<tr>
<td>Regularly on-site</td>
<td>1.57 (1.30-1.89)(^***)</td>
<td>1.50 (1.21-1.86)(^***)</td>
<td>1.56 (1.03-2.36)*</td>
</tr>
<tr>
<td>Daily exposure risk trajectory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class 2</td>
<td>1.38 (1.07-1.77)*</td>
<td>1.41 (1.07-1.84)*</td>
<td>1.33 (1.02-1.76)*</td>
</tr>
<tr>
<td>Class 3</td>
<td>1.50 (1.18-1.90)(^**)</td>
<td>1.57 (1.12-2.04)(^***)</td>
<td>1.46 (1.11-1.92)(^**)</td>
</tr>
<tr>
<td>Class 4</td>
<td>1.47 (1.09-1.99)*</td>
<td>1.70 (1.23-2.36)*</td>
<td>0.92 (0.39-2.18)</td>
</tr>
<tr>
<td>Social exposure risk trajectory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Class 2</td>
<td>0.99 (0.79-1.23)</td>
<td>1.04 (0.83-1.31)</td>
<td>1.02 (0.80-1.30)</td>
</tr>
<tr>
<td>Class 3</td>
<td>1.07 (0.82-1.40)</td>
<td>1.10 (0.84-1.53)</td>
<td>1.11 (0.82-1.49)</td>
</tr>
<tr>
<td>Class 4</td>
<td>1.35 (1.05-1.73)(^*)</td>
<td>1.40 (1.06-1.85)*</td>
<td>1.41 (1.06-1.87)*</td>
</tr>
<tr>
<td>Class 5</td>
<td>1.03 (0.79-1.34)</td>
<td>1.01 (0.79-1.45)</td>
<td>1.13 (0.59-2.21)</td>
</tr>
</tbody>
</table>

\(^a\)Adjusted for sex, age, education level, COVID-19 vaccination status, and work pattern. Cumulative incidence (as of December 2022)=51.6%.

\(^b\)Adjusted for sex, age, education level, and work pattern. Cumulative incidence for pre-Omicron vaccination status: 0-2 doses=54.4%. Cumulative incidence of pre-Omicron vaccination status ≥3 doses=42.7%.

\(^c\)HR: hazard ratio.

\(^d\)aHR: adjusted hazard ratio.
Discussion

Principal Findings

Based on the heterogeneous profiles of behavior changes over 3 epidemic waves, our results demonstrated the evolution of the SARS-CoV-2 Omicron outbreak through different exposure settings unfolding in the workplace, accelerating in the household, and sustained in daily and social environments in an 11-month period. With similar travel restrictions and mandatory mask wearing in place during the whole study period, these patterns illuminate the apparently stereotyped evolution of exposure settings brought on by changes in social distancing interventions and population susceptibility. The enhanced understanding of the contextual variations along an epidemic trajectory could inform more targeted control strategies against future respiratory epidemics.

When the Omicron BA.2 lineage predominated in wave 1, individuals working outside the home, having more workplace contacts, and living in a larger household were at a higher risk of infection [19]. It can be deduced that, against a low baseline burden (<0.2%), the high population susceptibility had rendered a majority of virus transmissions initiated at the workplace and scattered to workers’ residences afterward. The office is a ubiquitous work setting in Hong Kong. Apart from health care facilities where outbreaks were commonly reported, the close proximity and prolonged contact between coworkers in the office could also create a favorable environment for virus transmission and contribute to the initial spread [20]. This could help explain the higher infection risk identified not merely in health care workers but also in the population not sheltered by remote work in our study [21]. In Hong Kong, WFH practice had never been mandated over the entire epidemic. The absence of linkage between workplace exposure risk and SARS-CoV-2 infection in later epidemic waves also suggested the effectiveness of alternative WFH interventions aside from a hard mandate (ie, forcing all employees to work from home). Alternatives such as assigning employees into cohorts to return to the workplace on alternate days and limiting the number of employees at any one time should be considered in future outbreaks to minimize societal costs [22].

Stemming from workplace outbreaks, household exposure to SARS-CoV-2 had probably culminated in early February 2022 when WFH interventions were introduced [23]. A similar transition of exposure setting was observed upon the emergence of COVID-19 elsewhere, proving that such a pattern was unlikely to be coincidental [24]. With extensive facility closures at tier 1, the halting of community activities had likely further compelled SARS-CoV-2 transmissions within the residence [25]. Such an epidemiological transition was mirrored in the outbreak dynamics, in which the precipitous rise mimicked the multiple infections established in fully susceptible households, whereas the exponential decline resulted from the limited onward transmission preempted by isolation [10]. In hindsight, discouraging multihousehold gatherings could be an effective means to confine the epidemic spread when new infections surge in the community [26]. The observed higher risk in larger households, however, did not corroborate other study findings [27,28]. A plausible explanation could be that the practice of home isolation, as a makeshift for the shortage of isolation facilities, had facilitated secondary transmission in larger households [29]. Being one of the most densely populated cities worldwide, the case in Hong Kong called for a strategic focus on the pragmatic approach of home quarantine and isolation in preparation for the emerging pandemics [30].

During wave 2 when Omicron BA.4/5 lineages prevailed, the outbreaks had most likely affected individuals with a higher daily exposure risk [31]. This pattern concurred with the partial reopening of community facilities at tier 2 and the results of another study [32]. A meta-analysis estimated a low attack rate for SARS-CoV-2 transmission derived from public and casual contacts [33]. The inefficient mode of spread, coupled with the accelerated third-dose vaccination uptake in previous months, had probably contributed to the milder outbreaks in wave 2 even when >70% remained uninfected in the population [34]. The findings also revealed that, although class 4 participants had a wider source of exposure through participation in more leisure activities than class 3 participants, the former was not significantly more vulnerable to infection. It is possible that the heightened perceived risk and precautionary behaviors compensated the actual risk [35]. Based on the subgroup analysis, a higher daily exposure risk did not result in an increased likelihood of SARS-CoV-2 acquisition among individuals having received ≥3 vaccine doses before the outbreak. This has somewhat suggested the effectiveness of booster vaccination in preventing infections arising from a lower infecting dose through casual exposure. Likewise, a previous study found that vaccinated health care workers before infection were more likely to report probable exposure in households than in other settings [36]. These findings supported in part the administration of vaccine passes as a means to allow the safe resumption of day-to-day activities in protected individuals in the face of similar epidemics [37].

With presumably >30% of the population conferred with hybrid immunity resulting from the 2 epidemic waves, the number of incident cases in wave 3 had somewhat exceeded expectations in the lack of reinfection cases [38]. It is hypothesized that following tier 4 relaxation, the reinstatement of a larger gathering had rendered an increased virus exposure to socially active subgroups such as social exposure risk trajectory class 3 and class 5 participants, who might have played a part in extending the outbreaks. Although we could not rule out that the imported traces of the XBB lineage had aided in the pervasive spread [39], the phenomenon resembled previous findings revealing a higher SARS-CoV-2 susceptibility in individuals who attended meetings of ≥10 people [40] and celebrative events [41]. The observation should, however, be...
interpreted with caution since exposure to a virus in a social context necessitates the presence of an infected seed in one’s social network [42]. Of note, older individuals presented a higher social exposure risk. This could partly be due to the uninterrupted practice of yum-cha (dining and socializing at a dim sum restaurant) when eateries remained open throughout the epidemic [43]. Furthermore, male participants were less likely to adjust their social behaviors along the progression of the epidemic. These findings convey the need to promote compliance by targeting specific community groups differentiated by age and gender in future epidemics [44].

**Limitations in This Study**

Several limitations existed in this study. The self-reported nature of SARS-CoV-2 infection could lead to an underestimation of the cumulative incidence if infected participants did not respond to the follow-up survey in that month. Such bias was minimized by excluding participants without adequate follow-up responses from analyses. The higher vaccination uptake in older individuals deviating from reality also informed the presence of a selection bias. This could be subject to the recruitment of older adults who have a higher computer literacy, which enabled them to acquire more information about COVID-19 vaccination online [45]. Recall bias could also be generated from the retrospective report of activities. This was minimized by restricting the accessibility of the follow-up survey to the first 14 days of each month. Social desirability bias could also be present when participants failed to disclose activities, which were against the quarantine and isolation orders. Furthermore, the role of virus transmission in school was not captured, despite its effect being anticipated to be small, since normal schooling was not resumed in most parts of the study period. The timing of survey initiation also coincided with the inception of the Omicron outbreak. The absence of pre-Omicron data impeded our inference on the change in the exposure risk pattern compared to the baseline. The results should also be interpreted with caution in light of the different transmissibility between the Omicron lineages [46]. The impact was, however, assumed to be small in a highly vaccinated population, as the effectiveness of mRNA vaccines against Omicron BA.2 did not differ significantly compared to BA.5 infection [47].

**Conclusions**

Monitoring the shift of exposure settings in future epidemics is important, as the pattern could inform the effective calibration of social distancing measures in a targeted manner as opposed to an exhaustive lockdown. When formulating preparedness and response plans for emerging epidemics of respiratory viruses with similar transmission dynamics, policy makers should also take such contextual variations and the accompanying effects into consideration. In a community-wide outbreak, where the workplace happens to be the first epicenter, health authorities should act in anticipation of the rise in intrahousehold transmission when WFH interventions are expected. Later strategies could focus on the prevention of explosive outbreaks in pursuit of maximum resumption of daily and social activities in protected individuals. In light of the identified patterns, future research may also explore the broader implications of exposure settings on other NPIs such as the effectiveness of masking. Methodologically, this study demonstrates the capability of a participatory surveillance cohort in examining the effect of social distancing measures on population exposure and disease patterns. This has introduced opportunities for its wider applications in detecting early epidemiological patterns and retrospectively evaluating epidemic responses for future outbreaks.

**Acknowledgments**

The authors thank Ms Mandy Li and Ms Priscilla Wong for their logistic support and assistance in data collation. We also appreciate the continuous input from the participants of the EcSS2.0 cohort. This study is supported by the Health and Medical Research Fund Commissioned Research on the Novel Coronavirus Diseases, Health Bureau of the Government of Hong Kong Special Administrative Region (COVID1903008 – Project A).

**Data Availability**

The questionnaire data are protected, as consent for sharing individual-level information was not obtained from research participants. The source data for each figure in aggregate form are available upon reasonable request from the corresponding author.

**Authors’ Contributions**

EKY, NSW, and SSL conceptualized this study. CPC, NSW, and THK administered the project. CPC, NSW, and SSL contributed to the methodology. CPC conducted the formal analysis and investigation and wrote the original draft. All authors contributed to reviewing and editing the manuscript. NSW, SSL, and SYSW contributed to funding acquisition and supervision. All authors have approved the final paper.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File, 101 KB, publichealth_v10i1e51498_app1.docx]
References


Abbreviations
- aHR: adjusted hazard ratio
- aOR: adjusted odds ratio
- BIC: Bayesian information criterion
- HR: hazard ratio
- NPI: nonpharmaceutical intervention
- RR: relative risk
- WFH: work from home

Edited by A Mavragani; submitted 02.08.23; peer-reviewed by D Nash, Y Oster; comments to author 07.10.23; revised version received 26.10.23; accepted 05.05.24; published 19.06.24.

Please cite as:
Chan CP, Lee SS, Kwan TH, Wong SYS, Yeoh EK, Wong NS
Population Behavior Changes Underlying Phasic Shifts of SARS-CoV-2 Exposure Settings Across 3 Omicron Epidemic Waves in Hong Kong: Prospective Cohort Study
JMIR Public Health Surveill 2024;10:e51498
URL: https://publichealth.jmir.org/2024/1/e51498
doi:10.2196/51498
PMID:38896447

©Chin Pok Chan, Shui Shan Lee, Tsz Ho Kwan, Samuel Yeung Shan Wong, Eng-Kiong Yeoh, Ngai Sze Wong. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 19.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
A Prediction Model for Identifying Seasonal Influenza Vaccination Uptake Among Children in Wuxi, China: Prospective Observational Study

Qiang Wang1,2*, PhD; Liuqing Yang2*, MM; Shixin Xiu3, MM; Yuan Shen3, MM; Hui Jin2, PhD; Leesa Lin4,5,6, PhD

1Department of Epidemiology, School of Public Health, Fudan University, Shanghai, China
2Department of Epidemiology and Health Statistics, School of Public Health, Southeast University, Nanjing, China
3Department of Immunization, Affiliated Wuxi Center for Disease Control and Prevention of Nanjing Medical University, Wuxi Center for Disease Control and Prevention, Wuxi, China
4Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom
5Laboratory of Data Discovery for Health (D24H), Hong Kong Science Park, Hong Kong Special Administrative Region, China
6WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China

*these authors contributed equally

Corresponding Author:
Leesa Lin, PhD
Department of Infectious Disease Epidemiology
London School of Hygiene & Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 617 632 6142
Email: Leesa.Lin@lshtm.ac.uk

Abstract

Background: Predicting vaccination behaviors accurately could provide insights for health care professionals to develop targeted interventions.

Objective: The aim of this study was to develop predictive models for influenza vaccination behavior among children in China.

Methods: We obtained data from a prospective observational study in Wuxi, eastern China. The predicted outcome was individual-level vaccine uptake and covariates included sociodemographics of the child and parent, parental vaccine hesitancy, perceptions of convenience to the clinic, satisfaction with clinic services, and willingness to vaccinate. Bayesian networks, logistic regression, least absolute shrinkage and selection operator (LASSO) regression, support vector machine (SVM), naive Bayes (NB), random forest (RF), and decision tree classifiers were used to construct prediction models. Various performance metrics, including area under the receiver operating characteristic curve (AUC), were used to evaluate the predictive performance of the different models. Receiver operating characteristic curves and calibration plots were used to assess model performance.

Results: A total of 2383 participants were included in the study; 83.2% of these children (n=1982) were <5 years old and 6.6% (n=158) had previously received an influenza vaccine. More than half (1356/2383, 56.9%) the parents indicated a willingness to vaccinate their child against influenza. Among the 2383 children, 26.3% (n=627) received influenza vaccination during the 2020-2021 season. Within the training set, the RF model showed the best performance across all metrics. In the validation set, the logistic regression model and NB model had the highest AUC values; the SVM model had the highest precision; the NB model had the highest recall; and the logistic regression model had the highest accuracy, F1 score, and Cohen $\kappa$ value. The LASSO and logistic regression models were well-calibrated.

Conclusions: The developed prediction model can be used to quantify the uptake of seasonal influenza vaccination for children in China. The stepwise logistic regression model may be better suited for prediction purposes.

(JMIR Public Health Surveill 2024;10:e56064) doi:10.2196/56064

https://publichealth.jmir.org/2024/1/e56064
Introduction

The population of China comprises over 200 million children under the age of 15 years, accounting for approximately 10% of the global population in that age band [1]. The incidence rate of influenza infections among children aged ≤14 years was estimated to be 15.86 per 1000 person-seasons, a notably elevated figure in comparison to those of younger adults (5.17) and older adults (2.37) in China [2]. In the realm of public health interventions preventing influenza transmission, vaccination stands as an important strategy. A study conducted in the United States revealed that influenza vaccination averted 2.36 million influenza-associated illnesses among children aged 6 months to 17 years in the 2018-2019 season [3]. The technical influenza vaccination guideline in China recommends the inoculation of children aged 6 to 59 months as well as school-age children against influenza [4]. Such recommendations were underpinned by the rationale that children aged 6 to 59 months are at heightened risk of developing severe symptoms subsequent to infection, whereas school-age children, owing to their extensive social interactions, might propagate the transmission of the influenza virus [5,6]. However, a meta-analysis showed that only one-quarter of children aged 6 months to 17 years in mainland China were vaccinated against influenza [7].

Identifying individuals who are unlikely to receive the influenza vaccine in the upcoming season could provide valuable insights for public health strategies. For example, this information would be essential for health care professionals and policy makers to allocate resources effectively and develop targeted interventions to increase vaccination uptake. Several efforts have been made to develop a prediction model to identify vaccination uptake behavior. Loiacono et al [8] developed and validated a logistic regression model to predict adults’ influenza vaccine uptake in England. Another logistic regression model was constructed by Oster et al [9] using birth hospitalization records to predict undervaccination status in the United States. An expanding array of methodologies has been adopted for constructing prediction models and their subsequent application within clinical contexts, including Bayesian networks (BNs) and machine learning (ML) models [10,11]. Models derived from these diverse, adaptable, and intricate techniques have demonstrated potential for enhanced predictive accuracy. However, there are limited studies examining the performance of BN and ML models in predicting vaccination behaviors.

Methods

Study Design and Participants

We performed a cross-section survey in Wuxi city of eastern China between September 21 and October 17, 2020. Wuxi (31.5704° N latitude and 120.3055° E longitude) has a densely populated demographic (approximately 7.46 million residents) and a well-developed transportation infrastructure, which provide conducive conditions for the spread of influenza virus. Thus, determining the factors that can accurately predict influenza vaccination behaviors has practical implications, potentially alleviating the influenza-related burden in the city. Participants were recruited from six immunization clinics, including Anzhen, Dongting, Huazhuang, Jiangxi, Meicun, and Taihu, in Wuxi between September 21 and October 17, 2020 [15]. Parents bringing their child to the immunization clinic for vaccination were encouraged to participate in our study. We excluded children lacking an immunization record number (which is used to track vaccination administration records) and children aged >14 years. A total of 3009 participants were recruited. Individuals aged ≥6 months are eligible to receive the influenza vaccine in China [4,5]. To ensure that children would be eligible for influenza vaccination (ie, aged ≥6 months) at cohort entry, we excluded children born after March 2020 (n=626) (Figure 1).
Ethical Considerations

The Ethics Committee of Wuxi Center for Disease Control and Prevention granted approval for this study (2020No10). All participants who consented to complete the survey were required to sign informed consent forms. We assured participants that their survey responses would remain anonymous. Participants received a gift valued at 5 RMB (US $1=6.8148 RMB) upon completing the survey.

Outcomes and Candidate Predictors

One year after the survey, we extracted the influenza vaccination records for each child, including the vaccination time and site. The immunization records were retrieved from the Jiangsu Information Management System of Vaccination Cases [16]. The predicted outcome variable was influenza vaccination status during the 2020-2021 season, specifically between October 18, 2020, and June 30, 2021 (the final date for the supply of influenza vaccines in Wuxi during the 2020-2021 season).

Data on children’s characteristics such as age, sex, and firstborn status were obtained at cohort entry. Since parents are mainly responsible for children’s vaccinations, we also collected the parents’ sociodemographic information, including age, relationship to the child, education level, annual family income, and occupation. Parental vaccine hesitancy was measured using the 10-item Vaccine Hesitancy Scale based on a 5-point Likert scale [17]. We reversed answers to items L5, L9, and L10 as they are phrased negatively. The total score for all 10 items was summed and the total maximum score was 50 points. We established a cut-off value of 40, where scores ≤40 indicated “high hesitancy” and scores >40 indicated “low hesitancy” [15]. A previous study indicated that the distance to the immunization site was associated with vaccination behavior [18]. Therefore, we investigated parents’ perceptions of convenience to the clinic and satisfaction with the clinic services using one question for each aspect rated on a 5-point Likert scale. The ratings of “strong agreement” and “agreement” were coded together into “agree” and otherwise the responses were coded as “disagree.” Parents were asked about their intention to have their children vaccinated against influenza in the 2020-2021 season, with answers of “yes,” “not sure,” or “no” possible. The responses “no” and “not sure” were both included in the category “no” in the analysis. We also developed the predictive models under the conditions with and without combining “no” and “not sure” for comparison. The prior influenza vaccination history (prior to the 2020-2021 influenza season) was extracted as a predictor from the electronic health record system. All covariates are provided in Table S1 of Multimedia Appendix 1.

Model Building

The BN-based model was constructed using an iterative process in which the network structure was gradually constructed through a combination of manual construction and data-driven methods. Initially, a blocklist of arcs (ie, arcs that cannot be present in the network structure) was determined based on expert knowledge and logical relationships (order of event occurrences) using the variables collected. Using a scoring algorithm (the hill-climbing algorithm), 200 network structures were learned and constructed using bootstrap sampling based on all variables. The frequency of each directed arc in the 200 network structures was calculated, and arcs with frequencies higher than 60% were retained to derive an average network structure [10,19]. The variables not included in the network would then be excluded. The obtained average network structure was further optimized by removing illogical arcs (such as child’s sex to firstborn status).
and adjusting the direction of arcs that did not align with the logic (such as firstborn status to parents’ education level). The resulting network structure served as the basis for Bayesian posterior estimation, enabling the computation of conditional probabilities for each node.

We considered both stepwise logistic regression and least absolute shrinkage and selection operator (LASSO) regression. In the initial construction of the logistic and LASSO regression models, all variables were included. The logistic regression model was progressively refined through backward stepwise algorithms that systematically reduced variables by minimizing the Akaike information criterion (AIC) [8]. The model with the lowest AIC would be chosen as the prediction model. For LASSO regression, variable selection was achieved by penalizing the absolute values of the coefficients of the variables. As the tuning parameter $\lambda$ increases, coefficients of more variables shrink toward 0, which could result in a sparse model where only a subset of predictors are retained in the final model [8]. Through determining the optimal value of $\lambda$ using 10-fold cross-validation on the training set, we could minimize the prediction error. We report the estimated coefficients of variables in the final logistic regression and LASSO regression to indicate the contribution of variables.

The support vector machine (SVM), naive Bayes (NB), random forest (RF), and decision tree classifier (DTC) algorithms were also used to construct prediction models. All variables collected were included in the SVM, NB, and RF models. Ranges were set for cost and in the SVM model and a grid search was performed within the specified ranges to optimize the hyperparameters. Initially, all variables were included in the DTC model. We calculated the cost-complexity parameter value that yielded the lowest cross-validated error and pruned the tree model accordingly (ie, some variables were excluded) to prevent overfitting in the DTC model, resulting in a new tree to predict.

### Statistical Analysis

The variables were characterized by the frequency of occurrence. The data set was randomly split into training and validation sets at a 7:3 ratio. The training set was used for constructing the prediction model, whereas the validation set was used to validate the model and assess the model’s performance. The children’s vaccination status was classified based on the predicted probability following the approach proposed by Loiacono et al [8]. Specifically, a predicted probability $\leq 0.50$ was assigned to indicate nonreceipt of vaccination, while a probability above 0.50 indicated receipt of vaccination. Receiver operating characteristic (ROC) curves were plotted using sensitivity and 1–specificity. We evaluated model performance using various metrics, including accuracy, precision, recall, F1 score, area under the ROC curve (AUC), and Cohen $\kappa$ (see Table S2 in Multimedia Appendix 1). Calibration plots were also generated to compare the predicted outcomes from the model with the observed outcomes, providing insights into the alignment between predicted and actual probabilities [20]. Additionally, we performed sensitivity analyses, excluding children who were too young to have an influenza vaccination history (aged less than 6 months in September 2019). We also developed a model for children aged 6 months to 5 years as this is highlighted as the high-risk demographic for influenza in the guidelines [4,5].

All analyses were performed using R packages (bnlearn 4.8.1 and gRain 1.3.9 for the BN and NB models, MASS 7.3-51.5 for the logistic regression, glmnet 4.1-7 for the LASSO regression, e1071 1.7-13 for the SVM model, randomForest 4.6-12 for the RF model, and part 4.1.19 for the DTC model). We used Netica software 6.09 [21] to perform BN inference. Statistical significance was determined at a $P$ value threshold of $<.05$ using a two-sided test.

### Results

#### Participant Characteristics

The analyses included a total of 2383 individuals. Overall, 56.9% (1356/2383) of parents expressed a willingness to vaccinate their children against influenza, whereas 26.3% (627/2383) of children had received the influenza vaccine during the 2020-2021 season. The vaccine uptake status was significantly associated with various variables, including the children’s age, firstborn status, and previous influenza vaccine history, as well as parental factors such as age, educational level, annual household income, vaccine hesitancy, and willingness to vaccinate (Table 1). Among parents expressing an unwillingness to vaccinate their child, a notable proportion of children (90/1027, 8.7%) received vaccination. Among parents indicating a willingness to vaccinate their child, a considerable proportion of children (819/1356, 60.4%) remained unvaccinated (Table S3 in Multimedia Appendix 1). These results suggest that within the subset of parents expressing a willingness to vaccinate their children, factors including younger age of children, firstborn status, and absence of prior influenza vaccination, as well as lower parental education level and household income, may pose barriers to eventual vaccine uptake.

The training data set comprised 1668 children and the validation data set included 715 children. There were no significant differences in the variables between the training and validation data sets (see Table S4 in Multimedia Appendix 1).
## Table 1. Characteristics of participants.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All (N=2383), n (%)</th>
<th>Not receiving influenza vaccine (n=1756), n (%)</th>
<th>Receiving influenza vaccine (n=627), n (%)</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Children</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 months-2 years</td>
<td>1229 (51.6)</td>
<td>1144 (65.1)</td>
<td>85 (13.6)</td>
<td>523.02</td>
<td>4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3-5 years</td>
<td>753 (31.6)</td>
<td>360 (20.5)</td>
<td>393 (62.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-8 years</td>
<td>349 (14.6)</td>
<td>219 (12.5)</td>
<td>130 (20.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-11 years</td>
<td>37 (1.6)</td>
<td>23 (1.3)</td>
<td>14 (2.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥12 years</td>
<td>15 (0.6)</td>
<td>10 (0.6)</td>
<td>5 (0.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1214 (50.9)</td>
<td>901 (51.3)</td>
<td>313 (49.9)</td>
<td>0.36</td>
<td>1</td>
<td>.55</td>
</tr>
<tr>
<td>Female</td>
<td>1169 (49.1)</td>
<td>855 (48.7)</td>
<td>314 (50.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Firstborn</strong></td>
<td></td>
<td></td>
<td></td>
<td>18.78</td>
<td>1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>1585 (66.5)</td>
<td>1124 (64.0)</td>
<td>461 (73.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>798 (33.5)</td>
<td>632 (36.0)</td>
<td>166 (26.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prior influenza vaccine uptake</strong></td>
<td></td>
<td></td>
<td></td>
<td>226.16</td>
<td>1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>2225 (93.4)</td>
<td>36 (2.1)</td>
<td>122 (19.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>158 (6.6)</td>
<td>1720 (97.9)</td>
<td>505 (80.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parents</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Relationship with child</strong></td>
<td></td>
<td></td>
<td></td>
<td>2.22</td>
<td>1</td>
<td>.14</td>
</tr>
<tr>
<td>Mother</td>
<td>1779 (74.7)</td>
<td>1297 (73.9)</td>
<td>482 (76.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Father</td>
<td>604 (25.3)</td>
<td>459 (26.1)</td>
<td>145 (23.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td>95.27</td>
<td>4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;26</td>
<td>163 (6.8)</td>
<td>155 (8.8)</td>
<td>8 (1.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26-30</td>
<td>799 (33.5)</td>
<td>649 (37)</td>
<td>150 (23.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31-35</td>
<td>959 (40.2)</td>
<td>648 (36.9)</td>
<td>311 (49.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-40</td>
<td>377 (15.8)</td>
<td>248 (14.1)</td>
<td>129 (20.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥41</td>
<td>85 (3.6)</td>
<td>56 (3.2)</td>
<td>29 (4.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
<td></td>
<td></td>
<td>40.98</td>
<td>3</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Junior high school or below</td>
<td>263 (11.0)</td>
<td>224 (12.8)</td>
<td>39 (6.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school graduate or equivalent</td>
<td>467 (19.6)</td>
<td>375 (21.4)</td>
<td>92 (14.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College or equivalent</td>
<td>1487 (62.4)</td>
<td>1046 (59.6)</td>
<td>441 (70.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>166 (7.0)</td>
<td>111 (6.3)</td>
<td>55 (8.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Annual household income (US $)</strong></td>
<td></td>
<td></td>
<td></td>
<td>66.87</td>
<td>3</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&lt;7669</td>
<td>167 (7.0)</td>
<td>140 (8.0)</td>
<td>27 (4.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7669 to &lt;15,337</td>
<td>677 (28.4)</td>
<td>556 (31.7)</td>
<td>121 (19.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15,337 to &lt;23,006</td>
<td>612 (25.7)</td>
<td>454 (25.9)</td>
<td>158 (25.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥23,006</td>
<td>927 (38.9)</td>
<td>606 (34.5)</td>
<td>321 (51.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Health care occupation</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
<td>1</td>
<td>.86</td>
</tr>
<tr>
<td>Yes</td>
<td>148 (6.2)</td>
<td>110 (6.3)</td>
<td>38 (6.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2235 (93.8)</td>
<td>1646 (93.7)</td>
<td>589 (93.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vaccine hesitancy</strong></td>
<td></td>
<td></td>
<td></td>
<td>13.64</td>
<td>1</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Model Training

The final BN framework consisted of 10 nodes and 11 arcs, including children’s age, firstborn status, previous influenza vaccine history, parental age, education level, annual household income, vaccine hesitancy, willingness to vaccinate, satisfaction with clinic services, and vaccination behavior (Figure S1 in Multimedia Appendix 1). Child’s sex, relationship with child, parent’s occupation, and convenience to the clinic were not included in the BN framework. Child’s age, influenza vaccine history, and willingness to vaccinate were directly linked to the influenza vaccination behavior. Additionally, parent’s age had an indirect association with the vaccination behavior. When the BN receives information from individual findings, probabilistic reasoning is used to determine the maximum a posteriori probability of receiving the vaccine. For instance, if parents expressed a willingness to vaccinate their child and the child had been vaccinated against influenza previously, the probability of receiving the influenza vaccine increased from 26.6% to 85.6% (Figure S1 in Multimedia Appendix 1).

A total of 6 variables were ultimately included in the logistic regression model (Table S5 in Multimedia Appendix 1). The LASSO regression model (Table S5 in Multimedia Appendix 1) included 10 variables, with the tuning parameter log(λ) set at −5.438 (Figure S2 in Multimedia Appendix 1). Child’s age, prior influenza vaccine uptake, and parents’ willingness to vaccinate had the greatest contributions to the logistic and LASSO regression models according to estimated coefficient values.

Model Performance

Model performance metrics are summarized in Table 2. Among the seven models constructed within the training set, the RF model demonstrated the best performance, with the highest values in AUC, accuracy, precision, recall, F1 score, and Cohen κ. Within the validation set, the logistic regression model and the NB model achieved the highest AUC values, the SVM model had the highest precision, and the NB model had the highest recall. The logistic regression model had the highest accuracy, F1 score, and Cohen κ values. The performance of the seven models was similar when using different methods of combining answers related to vaccination willingness (see Table S6 in Multimedia Appendix 1).
Within the training set, the prediction of receiving the vaccine was well-calibrated with the observed vaccination behavior using the DTC, LASSO, and logistic regression models (Figure S3 in Multimedia Appendix 1). Within the validation data set, the LASSO and logistic regression models consistently provided more reliable predicted probabilities. For the predicted probabilities between 0.25 and 0.50, the SVM, RF, and DTC models tended to overpredict, while the NB model tended to underpredict. For predicted probabilities over 0.50, all seven models displayed a bias toward underprediction.

Sensitivity Analysis
A total of 1699 participants aged <6 months in September 2019 were included for the sensitivity analysis (Table S7 in Multimedia Appendix 1). The RF model exhibited the best performance within the training set, achieving the highest AUC (0.944, 95% CI 0.932-0.956), accuracy (0.881), precision (0.832), recall (0.813), F1 score (0.822), and Cohen \( \kappa \) values (0.733) (see Figure S4 and Table S8 in Multimedia Appendix 1). Within the validation set, the LASSO regression model had the highest AUC (0.880, 95% CI 0.849-0.910), followed by the logistic regression model (0.878, 95% CI 0.848-0.909). The NB and RF models both had the highest accuracy (0.816), while the DTC model had the highest precision (0.774). The NB model achieved the highest recall, F1 score, and Cohen \( \kappa \) values at 0.698, 0.719, and 0.582, respectively.

Regarding the calibration of predictions, within the training set, the predictions of the DTC, LASSO regression, logistic regression, and BN models were generally well-calibrated with the observed behavior (Figure S5 in Multimedia Appendix 1). In the validation set, none of the seven models demonstrated good calibration. When the predicted probabilities were above 0.50, the DTC, BN, LASSO regression, logistic regression, and SVM models exhibited a bias toward overprediction, while the NB and RF models showed a bias toward underprediction.

A total of 1982 participants aged 6 months to 5 years were analyzed separately (Table S9 in Multimedia Appendix 1). Within the training set, the RF model also exhibited the highest values across all performance metrics (Figure S6 and Table S10 in Multimedia Appendix 1). Within the validation set, the NB model showed the highest AUC (0.884, 95% CI 0.851-0.916), accuracy (0.860), precision (0.727), and Cohen \( \kappa \) (0.607) values, whereas the BN model had the highest recall (0.713) and F1 score (0.701). Within the validation set, the SVM and logistic regression models showed better calibration than the other models (Figure S7 in Multimedia Appendix 1).

Discussions
Principal Findings
In the study, we developed models to predict the likelihood of influenza vaccination uptake among children aged 6 months to 14 years in the upcoming season. We conducted a survey in Wuxi city, eastern China, between September 21 and October 17, 2020, and extracted the participants’ influenza vaccination records after 1 year. Data from a total of 2383 participants were included in the analysis. The RF model demonstrated superior performance metrics within the training set, whereas the logistic regression model exhibited a narrow margin of performance superiority within the validation set. Both the logistic and LASSO regression models showed good calibration. Among these models, the stepwise logistic regression approach,
characterized by its simplicity and favorable interpretability, emerged as a viable candidate for predictive purposes.

The factors associated with influenza vaccination included the child’s age, child’s previous influenza vaccine uptake, parental age, educational level, annual household income, parental vaccine hesitancy, and influenza vaccination willingness. Among these factors, previous vaccine uptake emerged as a strong positive factor, as supported by previous studies [12,22,23]. Individuals who personally witnessed the positive impact due to vaccination, such as a decreased likelihood of infection, were more likely to continue receiving vaccinations. To ensure the accuracy of this predictor, we obtained this variable directly from the electronic record system rather than relying on participants’ recall, thereby enhancing its reliability. Compared to children aged between 6 months and 2 years, older children were more likely to be vaccinated against influenza. This may be attributed to their enrollment in kindergarten or school, where peer influence and social networks potentially foster vaccination uptake [24]. Parents were concerned about the adverse effects of vaccination on younger children, which may also decrease the likelihood of young children receiving the influenza vaccine. Additionally, children from households with higher levels of parental education or income were more prone to receiving the influenza vaccine in our study. However, evidence regarding the relationship between parental education, income, and children’s influenza vaccination is inconsistent [22,23]. We also analyzed the factors influencing parents who expressed a willingness to vaccinate their children but ultimately did not do so. In addition, participants might be more inclined to express willingness toward influenza vaccination due to social desirability bias [16].

Our results did not reveal superior predictive performance of the ML methods over the logistic regression model, which was consistent with the outcomes of a previous systematic review by Christodoulou et al [25]. The strength of ML lies in its ability to handle high-dimensional data [26]. In instances characterized by constrained sample sizes and a limited number of variables, ML may be susceptible to overfitting. The BN model often involves capturing conditional dependencies among modeled variables. If the data set is confined or the interrelationships between variables are not accurately captured, the predictive performance of a BN model might diminish. The number of predictors and the sample size of this study were limited, which might potentially constrain the purported advantages associated with some complex methodologies such as ML and BN models. In general, ML techniques require more than 10 instances for each predictor to mitigate the risk of overfitting [26]. Nonetheless, despite satisfying this criterion, challenges such as instability and high optimism may persist [27]. The minimum sample size required for predictive models constructed with ML and BN approaches is still being explored [28,29].

In the sensitivity analysis, RF still showed the best performance within the training set and the performance metrics of seven models did not show substantial discrepancies within the test set. Additionally, we did not find any model that showed a prominent advantage in terms of calibration capability. Logistic regression might still be the most suitable tool to predict vaccination behaviors owing to its simplicity and favorable interpretability.

Implications for Practice

Our findings offer valuable insight for practice, allowing for prediction of the likelihood of children receiving the influenza vaccine. These modeling results can enable health care professionals to tailor their strategies based on the predicted probabilities. When dealing with children who are highly likely to be vaccinated, health care workers could save time and simplify the decision-making process [8]. For example, they could leverage messaging systems or electronic letters to provide reminders [30,31] or they could inform parents about the accessibility of the influenza vaccine during medical consultations. These efforts could particularly benefit children with a low predicted probability of vaccination. Health care workers can also use various communication techniques such as motivational interviewing and improvisational theater to encourage vaccination, which are known for their persuasive and respectful features [32,33]. These approaches can help to identify barriers and increase vaccine confidence, thereby improving vaccine uptake.

Limitations

Our study has several limitations. First, in China, adherence to mandatory vaccination requirements prior to school enrollment is imperative [34]. Noncompliance with these vaccination mandates results in exclusion from school attendance. Consequently, parents were compelled to accompany their children to immunization clinics for vaccination administration. However, the recruitment of participants solely from immunization clinics may still introduce selection bias, potentially limiting the representativeness of the study population to the general population of children. In addition, the data collection was limited to Wuxi city in eastern China. Although we performed internal validation, the model has not been externally validated in large samples or diverse locations. Second, vaccination behavior is influenced by various factors, including parental perceptions of susceptibility to influenza, the child’s health status, and vaccine policies. To enhance the performance of the model, it is important to explore additional factors that contribute to vaccination behavior and incorporate them into the modeling. Third, influenza vaccination willingness may vary across seasons and our study focused on a specific time frame. To account for this variability, future research should consider conducting season-specific surveys to capture the changing dynamics of vaccination willingness.

Conclusion

Our findings indicate that the influenza vaccination behavior for children aged 6 months to 14 years could be identified using the established predictive model. We propose that the stepwise logistic regression model with high accuracy and a straightforward modeling methodology could be better suited for such prediction tasks. Further validation of the model in larger and more diverse samples is necessary. More critical predictors should be considered to increase the accuracy and reliability of predicting influenza vaccination behavior among children.
Acknowledgments
This work was supported by the Innovation and Technology Commission (grant AIR@InnoHK). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Availability
Data may be available from the corresponding author on reasonable request depending on the policy and procedures of the institutions that participate in the consortiums.

Authors' Contributions
QW was responsible for study conceptualization, data curation, formal analysis, writing the original draft, manuscript review and editing, methodology, and visualization. LY contributed to the methodology and writing of the original draft of the manuscript. SX was involved in study conceptualization and investigation. YS played a role in the investigation. HJ supervised the study. LL was involved in study conceptualization, supervision, manuscript review and editing, and funding acquisition. LL, HJ, and SX contributed equally as senior authors.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Table S1. Candidate variables for construction of the prediction model; Table S2. Explanation of metrics evaluating model performance; Table S3. Factors influencing the gap between willingness and behaviors; Table S4. Characteristics of participants in the training and validation data sets; Figure S1. Final BN for the prediction of receiving the influenza vaccine; Table S5. Estimated coefficients in the logistic and LASSO regressions; Figure S2. Binomial deviance versus log(λ) for 10-fold within-sample cross-validation of the LASSO model; Figure S3: Calibration plots for training and validation data sets; Table S7. Characteristics of participants (children aged less than 6 months on September, 2019); Figure S4. Receiver operating characteristic curves of various prediction models (children aged less than 6 months on September, 2019); Table S8. Model performance (children aged less than 6 months on September, 2019); Figure S5. Calibration plots for training and validation data sets (children aged less than 6 months on September, 2019); Table S9. Characteristics of participants (children aged 6 months to 5 years); Figure S6. Receiver operating characteristic curves of various prediction models (children aged 6 months to 5 years); Table S10. Model performance (children aged 6 months to 5 years); Figure S7. Calibration plots for training and validation data sets (children aged 6 months to 5 years).

References


Abbreviations

AIC: Akaike information criterion
AUC: area under the receiver operating characteristic curve
BN: Bayesian network
DTC: decision tree classifier
LASSO: least absolute shrinkage and selection operator
ML: machine learning
NB: naive Bayes
RF: random forest
ROC: receiver operating characteristic
SVM: support vector machine

©Qiang Wang, Liuqing Yang, Shixin Xiu, Yuan Shen, Hui Jin, Leesa Lin. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
The Migrant-Local Difference in the Relationship Between Social Support, Sleep Disturbance, and Loneliness Among Older Adults in China: Cross-Sectional Study

Mingli Pang1,2,3, MPH; Jieru Wang1,2,3, MPH; Mingyue Zhao4, MPH; Rui Chen1,2,3, MPH; Hui Liu1,2,3, MPH; Xixing Xu1,2,3, MPH; Shixue Li1,2,3, PhD; Fanlei Kong1,2,3, PhD

1Centre for Health Management and Policy Research, School of Public Health, Cheelu College of Medicine, Shandong University, Jinan, China
2NHC Key Lab of Health Economics and Policy Research, Shandong University, Jinan, China
3Institute of Health and Elderly Care, Shandong University, Jinan, China
4Human Resource Department, The Second Hospital of Shandong University, Jinan, China

Corresponding Author:
Fanlei Kong, PhD
Centre for Health Management and Policy Research
School of Public Health, Cheelu College of Medicine
Shandong University
44 West Wenhau Road
Jinan, 250012
China
Email: kongfanlei@sdu.edu.cn

Abstract

Background: Driven by the accelerated aging of the population of China, the number of older adults has increased rapidly in the country. Meanwhile, following children, migrant older adults (MOA) have emerged as a vulnerable group in the process of fast urbanization. Existed studies have illustrated the association between social support and loneliness and the relationship between sleep disturbance and loneliness; however, the underlying mechanisms and the migrant-local difference in the association between social support, sleep disturbance, and loneliness have not been identified.

Objective: This study aimed to clarify the migrant-local difference in the relationship between social support, sleep disturbance, and loneliness in older adults in China.

Methods: Multistage cluster random sampling was used to select participants: 1205 older adults (n=613, 50.9%, MOA and n=592, 49.1%, local older adults [LOA]) were selected in Weifang City, China, in August 2021. Loneliness was assessed with the 6-item short-form University of California, Los Angeles Loneliness Scale, social support was evaluated with the Social Support Rating Scale, and sleep disturbance was measured with the Pittsburgh Sleep Quality Index. The chi-square test, t test, and structural equation modeling (SEM) were adopted to explore the migrant-local difference between social support, sleep disturbance, and loneliness among the MOA and LOA.

Results: The mean score of loneliness was 8.58 (SD 3.03) for the MOA and 8.00 (SD 2.79) for the LOA. SEM analysis showed that social support exerts a direct negative effect on both sleep disturbance (standardized coefficient=–0.24 in the MOA and –0.20 in the LOA) and loneliness (standardized coefficient=–0.44 in the MOA and –0.40 in the LOA), while sleep disturbance generates a direct positive effect on loneliness (standardized coefficient=0.13 in the MOA and 0.22 in the LOA).

Conclusions: Both MOA and LOA have a low level of loneliness, but the MOA show higher loneliness than the LOA. There is a negative correlation between social support and loneliness as well as between social support and sleep disturbance among the MOA and LOA (MOA>LOA), while loneliness is positively associated with sleep disturbance in both populations (MOA<LOA). Measures should be taken by the government, society, and families to increase social support, decrease sleep disturbance, and further reduce the loneliness among older adults, especially the MOA.

(JMIR Public Health Surveill 2024;10:e49253) doi:10.2196/49253

KEYWORDS
loneliness; social support; sleep disturbance; older adults; migrant-local difference; structural equation modeling
Introduction

Loneliness is typically defined as the discrepancy between a person’s desired and actual social relationships [1], which would increase the risk of other mental health problems [2,3] and physical health [4]. All people face loneliness at some stage in life [5]. Loneliness varies at different ages, and those aged less than 25 years and those aged over 65 years demonstrate the highest levels of loneliness [6]; data shows that up to one-third of older adults feel lonely [7].

Migration is found to be a risk factor for loneliness [8]. A study among ethnic-migrant groups in Australia showed that migrants from non-English-speaking countries demonstrate higher levels of loneliness than native-born, nonindigenous Australians [8]. Meanwhile, a study in Canada also illustrated that immigrants report higher levels of loneliness than native-born Canadians [9]. Many studies have shown that compared with local older adults (LOA), migrant older adults (MOA) tend to report higher loneliness [10,11]. These findings make it worthwhile to focus on the migrant-local difference between older adults’ loneliness.

Social support is an important determinant of loneliness [12,13]. A study among older adults in South Korea showed that social support is negatively associated with loneliness [14]. Moreover, research among older persons in the United Kingdom revealed that lower social support during the pandemic increased the risk of experiencing loneliness [15]. Meanwhile, a study among women during the postpartum period also showed that poorer social support would generally indicate a higher potential risk of loneliness [16].

A previous study showed that poor sleep is an independent risk factor for poor mental health in older adults [17]. A previous study in Denmark among young adolescents reported that loneliness is correlated with poor sleep problems [18]. One research among the rural older adults of China found that poor sleep quality is associated with increased odds of loneliness [19]. A study among US adults illustrated that more severe insomnia symptoms are significantly related to more feelings of loneliness [20]. Meanwhile, many studies have shown that a disrupted circadian rhythm/sleep loss and sleep problems could cause loneliness [21-23]. These findings indicate that sleep is linked to loneliness.

Existing studies have shown that people with poor social support [24] have poor sleep quality [25], and social support is found to be significantly associated with sleep disturbance [26]. Those with instrumental support can reduce the risk of all sleep problems, and having emotional support can reduce the risk of poor sleep quality [27]. A previous study among college students in China also found that perceived social support is significantly and negatively associated with insomnia [28].

In summary, the relationship between social support and loneliness, between social support and sleep disturbance, and between sleep disturbance and loneliness has been separately explored in previous studies. However, few studies have investigated the association between social support, sleep disturbance, and loneliness simultaneously related to the migrant-local difference. Thus, this study aimed to identify the migrant-local difference in the relationship between social support, sleep disturbance, and loneliness among older adults in China.

Methods

Participants

A total of 1205 older adults were selected in Weifang City, China, in August 2021: 613 (50.9%) MOA and 592 (49.1%) LOA. Up to November 30, 2020, the total household population of Weifang City was 9.39 million, of which 2.04 million, accounting for 21.7% of the total population, were aged 60 years and above, and the migrant population in the city was 2.38 million [29].

Multistage cluster random sampling was used to select participants. In the first stage, 4 districts were selected as primary sampling units (PSUs) based on the economic development and geographical location of Weifang City. In the second stage, in total, 4 subdistricts from the PSUs were selected as secondary sampling units (SSUs). In the third stage, in total, 4 communities were selected from the SSUs. The population aged 60 years or above in the selected communities in Weifang City constituted the entire sample. The inclusion criteria for participants were (1) age 60 years and above and (2) clear awareness and cognition. The MOA’s household registration was beyond Weifang City, while the LOA’s household registration was in Weifang City.

After completing training on the research background, questionnaire content, and social survey techniques, a total of 25 university students were assigned as interviewers for this study. Approximately 20-minute face-to-face interviews were conducted with each participant. Initially, 1208 older adults were selected and interviewed. However, 3 (0.2%) participants were excluded as they answered their questionnaires incorrectly or incompletely. Ultimately, a total of 1205 (99.8%) participants were included in the sample.

Assessment and Measurements

Sociodemographic Characteristics

Sociodemographic characteristics included the following: gender (male, female), age (60-65, 66-70, 71-80, >80 years), marital status (married, single/divorced/separated/widowed), education level (primary school and below, junior high school, high school and above), job before retirement (agriculture, forestry, animal husbandry and fishery, others), and monthly household income (first quartile [Q1], second quartile [Q2], third quartile [Q3], fourth quartile [Q4]; Q1 was the poorest, and Q4 was the richest).

Loneliness

Loneliness is measured with the University of California at Los Angeles (UCLA) Loneliness Scale. This scale includes 20 items, which range from 1 (never) to 4 (often) [30]. In this study, we used the 6-item short-form UCLA Loneliness Scale (ULS-6) to assess loneliness among older adults. The ULS-6 includes 6 items, with higher scores indicating higher loneliness [31]. The ULS-6 is suitable to use among older adults, and its reliability

https://publichealth.jmir.org/2024/1/e49253
and validity in Chinese individuals has been verified [32,33], with a Cronbach α coefficient of .82 in this study.

**Social Support**

The Social Support Rating Scale (SSRS), developed by Xiao [34], was used to measure social support. The SSRS includes 10 items and 3 dimensions: objective social support, subjective social support, and social support use. The higher the score, the higher the social support level. The SSRS is appropriate to use among older adults [35], and the Cronbach α coefficient was .822 in this study.

**Sleep Disturbance**

The Pittsburgh Sleep Quality Index (PSQI) was used to assess the sleep disturbance of the participants in this study [36]. The PSQI consists 19 self-reported items divided into 7 components. The total score ranges from 0 to 21, and higher scores indicate severer sleep disturbance. The PSQI is appropriate to use among older adults [37,38], and its reliability and validity in Chinese individuals has been verified [39,40], with a Cronbach α coefficient of .731 in this study.

**Statistical Analysis**

Descriptive statistics were used to clarify the sociodemographic characteristics of the older adults. The migrant-local difference between the sociodemographic characteristics among older adults was investigated using the chi-square test, and the migrant-local difference between social support, sleep disturbance, and loneliness was explored using the t test. Statistical significance was defined as P<.05. All statistical analyses were performed using SPSS version 24.0 (SPSS Inc).

**Ethical Considerations**

The survey and data were obtained with written informed consent from all participants. The research program of the study was reviewed and approved by the Institutional Review Board (IRB) of Public Health and Preventive Medicine in Shandong University (#20180225) and was in accordance with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

### Results

**Characteristics of the Participants**

Table 1 demonstrates the sociodemographic characteristics of the participants. We included 1205 older adults in the analysis, of which 613 (50.9%) were MOA and 592 (49.1%) were LOA. The MOA were mostly women (n=448, 73.1%); more than half (n=342, 55.8%) were 60-65 years old; most of them were married (n=539, 87.9%); 346 (56.4%) had primary school education or less; 461 (75.2%) were occupied in agricultural, forestry, and animal husbandry and fishery industries; 217 (35.4%) belonged to the Q1 (the poorest quartile) monthly household income group; 351 (57.3%) had no chronic disease; 335 (54.6%) had physical pain; and 195 (31.8%) reported a good health status. The LOA were also mostly women (n=437, 73.8%); 192 (32.4%) were 60-65 years old; 432 (73%) were married; 287 (48.5%) had primary school education or less; 400 (67.6%) worked in occupations other than agricultural, forestry, and animal husbandry and fishery industries; 219 (37%) belonged to the Q4 (the richest quartile) monthly household income group; 355 (60%) had chronic disease; 351 (59.3%) had physical pain; and 169 (28.5%) reported a good health status.

The differences between the MOA and the LOA were statistically significant (P<.001, marital status (P<.001), education level (P<.02), job before retirement (P<.001), monthly household income (P<.001), chronic disease (P<.001), and self-reported health status (P<.001).

Table 2 shows the characteristics of social support, sleep quality, and loneliness among the MOA and LOA. The mean total score of loneliness was 8.58 (SD 3.03) for the MOA and 8.00 (SD 2.79) for the LOA. Specifically, the mean scores of individual items were as follows: often feel a lack of friends, 1.53 (SD 0.79) for the MOA and 1.36 (SD 0.71) for the LOA; often feel no one can be trusted, 1.47 (SD 0.73) for the MOA and 1.37 (SD 0.71) for the LOA; often feel left out, 1.37 (SD 0.62) for the MOA and 1.28 (SD 0.65) for the LOA; often feel separated from others, 1.37 (SD 0.67) for the MOA and 1.30 (SD 0.64) for the LOA; often feel shy, 1.36 (SD 0.64) for the MOA and 1.25 (SD 0.52) for the LOA; and often feel surrounded by people but not cared for, 1.49 (SD 0.72) for the MOA and 1.43 (SD 0.71) for the LOA. Statistically significant differences were found for the total score (t(1203)=−3.442, P<.001), often feel a lack of friends (t(1203)=−3.704, P<.001), often feel no one can be trusted (t(1203)=−2.326, P=.02), often feel left out (t(1203)=−2.672, P=.008), and often feel shy (t(1203)=−3.365, P=.001) between the MOA and the LOA.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N=1205), n (%)</th>
<th>MOA (n=613), n (%)</th>
<th>LOA (n=592), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex; ( \chi^2 = 0.083, P = .77 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>320 (26.6)</td>
<td>165 (26.9)</td>
<td>155 (26.2)</td>
</tr>
<tr>
<td>Female</td>
<td>885 (73.4)</td>
<td>448 (73.1)</td>
<td>437 (73.8)</td>
</tr>
<tr>
<td><strong>Age (years); ( \chi^2 = 131.429, P &lt; .001 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60-65</td>
<td>534 (44.3)</td>
<td>342 (55.8)</td>
<td>192 (32.4)</td>
</tr>
<tr>
<td>66-70</td>
<td>301 (25.0)</td>
<td>171 (27.9)</td>
<td>130 (22.0)</td>
</tr>
<tr>
<td>71-80</td>
<td>256 (21.2)</td>
<td>80 (13.1)</td>
<td>176 (29.7)</td>
</tr>
<tr>
<td>&gt;80</td>
<td>114 (9.5)</td>
<td>20 (3.2)</td>
<td>94 (15.9)</td>
</tr>
<tr>
<td><strong>Marital status; ( \chi^2 = 43.045, P &lt; .001 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>971 (80.6)</td>
<td>539 (87.9)</td>
<td>432 (73.0)</td>
</tr>
<tr>
<td>Single/divorced/separated/widowed</td>
<td>234 (19.4)</td>
<td>74 (12.1)</td>
<td>160 (27.0)</td>
</tr>
<tr>
<td><strong>Education level; ( \chi^2 = 7.662, P = .02 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school and below</td>
<td>633 (52.5)</td>
<td>346 (56.4)</td>
<td>287 (48.5)</td>
</tr>
<tr>
<td>Junior high school</td>
<td>338 (28.0)</td>
<td>158 (25.8)</td>
<td>180 (30.4)</td>
</tr>
<tr>
<td>High school and above</td>
<td>234 (19.5)</td>
<td>109 (17.8)</td>
<td>125 (21.1)</td>
</tr>
<tr>
<td><strong>Job before retirement; ( \chi^2 = 221.935, P &lt; .001 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agricultural, forestry, animal husbandry and fishery</td>
<td>653 (54.2)</td>
<td>461 (75.2)</td>
<td>192 (32.4)</td>
</tr>
<tr>
<td>Others</td>
<td>552 (45.8)</td>
<td>152 (24.8)</td>
<td>400 (67.6)</td>
</tr>
<tr>
<td><strong>Monthly household income; ( \chi^2 = 158.680, P &lt; .001 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>301 (25.0)</td>
<td>217 (35.4)</td>
<td>84 (14.2)</td>
</tr>
<tr>
<td>Q2&lt;sup&gt;d&lt;/sup&gt;</td>
<td>301 (25.0)</td>
<td>194 (31.6)</td>
<td>107 (18.1)</td>
</tr>
<tr>
<td>Q3&lt;sup&gt;e&lt;/sup&gt;</td>
<td>302 (25.0)</td>
<td>120 (19.6)</td>
<td>182 (30.7)</td>
</tr>
<tr>
<td>Q4&lt;sup&gt;f&lt;/sup&gt;</td>
<td>301 (25.0)</td>
<td>82 (13.4)</td>
<td>219 (37.0)</td>
</tr>
<tr>
<td><strong>Chronic disease; ( \chi^2 = 35.765, P &lt; .001 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>617 (51.2)</td>
<td>262 (42.7)</td>
<td>355 (60.0)</td>
</tr>
<tr>
<td>No</td>
<td>588 (48.8)</td>
<td>351 (57.3)</td>
<td>237 (40.0)</td>
</tr>
<tr>
<td><strong>Physical pain; ( \chi^2 = 2.646, P = .104 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>686 (56.9)</td>
<td>335 (54.6)</td>
<td>351 (59.3)</td>
</tr>
<tr>
<td>No</td>
<td>519 (43.1)</td>
<td>278 (45.4)</td>
<td>241 (40.7)</td>
</tr>
<tr>
<td><strong>Self-reported health status; ( \chi^2 = 35.289, P &lt; .001 )</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad</td>
<td>89 (7.4)</td>
<td>30 (4.9)</td>
<td>59 (10.0)</td>
</tr>
<tr>
<td>Relatively bad</td>
<td>29 (2.4)</td>
<td>15 (2.4)</td>
<td>14 (2.4)</td>
</tr>
<tr>
<td>Average</td>
<td>251 (20.8)</td>
<td>113 (18.4)</td>
<td>138 (23.2)</td>
</tr>
<tr>
<td>Relatively good</td>
<td>233 (19.3)</td>
<td>106 (17.3)</td>
<td>127 (21.5)</td>
</tr>
<tr>
<td>Good</td>
<td>364 (30.3)</td>
<td>195 (31.8)</td>
<td>169 (28.5)</td>
</tr>
<tr>
<td>Very good</td>
<td>239 (19.8)</td>
<td>154 (25.2)</td>
<td>85 (14.4)</td>
</tr>
</tbody>
</table>

<sup>a</sup>M OA: migrant older adults.
The mean total score of sleep disturbance was 4.29 (SD 3.57) for the MOA and 4.58 (SD 3.66) for the LOA. Specifically, the mean scores of individual items were as follows: subjective sleep quality, 0.88 (SD 0.83) for the MOA and 0.97 (SD 0.86) for the LOA; sleep latency, 1.11 (SD 1.21) for the MOA and 1.12 (SD 1.19) for the LOA; sleep continuity, 0.68 (SD 0.97) for the MOA and 0.68 (SD 0.97) for the LOA; habitual sleep efficiency, 0.18 (SD 0.56) for the MOA and 0.18 (SD 0.56) for the LOA; sleep disorder, 0.97 (SD 0.57) for the MOA and 0.97 (SD 0.57) for the LOA; use of sleep medicine, 0.13 (SD 0.55) for the MOA and 0.18 (SD 0.62) for the LOA; and daytime dysfunction, 0.52 (SD 0.78) for the MOA and 0.48 (SD 0.77) for the LOA. Statistically significant differences were found for subjective sleep quality (t_{1203}=1.968, P=0.05) and sleep continuity (t_{1203}=2.822, P=0.005) between the MOA and the LOA.

Table 2. General characteristics of social support, sleep disturbance, and loneliness among the MOA\textsuperscript{a} and LOA\textsuperscript{b} by the migrant-local difference using the \(t\) test in Weifang City, China, in August 2021.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total (N=1205), mean (SD)</th>
<th>Migrant-local difference</th>
<th>(t_{1203}) Test</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MOA (n=613), mean (SD)</td>
<td>LOA (n=592), mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support (SSRS\textsuperscript{c})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>38.94 (6.61)</td>
<td>38.89 (6.63)</td>
<td>39.51 (6.86)</td>
<td>1.612</td>
</tr>
<tr>
<td>Objective social support</td>
<td>8.22 (2.01)</td>
<td>8.47 (1.64)</td>
<td>7.95 (2.30)</td>
<td>-4.545</td>
</tr>
<tr>
<td>Subjective social support</td>
<td>23.68 (4.47)</td>
<td>23.47 (4.79)</td>
<td>24.43 (4.40)</td>
<td>3.608</td>
</tr>
<tr>
<td>Social support use</td>
<td>7.04 (2.35)</td>
<td>6.94 (2.26)</td>
<td>7.14 (2.44)</td>
<td>1.432</td>
</tr>
<tr>
<td>Sleep disturbance (PSQI\textsuperscript{d})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4.44 (3.66)</td>
<td>4.29 (3.57)</td>
<td>4.58 (3.66)</td>
<td>1.396</td>
</tr>
<tr>
<td>Subjective sleep quality</td>
<td>0.92 (0.84)</td>
<td>0.88 (0.83)</td>
<td>0.97 (0.86)</td>
<td>1.968</td>
</tr>
<tr>
<td>Sleep latency</td>
<td>1.12 (1.20)</td>
<td>1.11 (1.21)</td>
<td>1.12 (1.19)</td>
<td>0.108</td>
</tr>
<tr>
<td>Sleep continuity</td>
<td>0.61 (0.91)</td>
<td>0.54 (0.84)</td>
<td>0.68 (0.97)</td>
<td>2.822</td>
</tr>
<tr>
<td>Habitual sleep efficiency</td>
<td>0.17 (0.53)</td>
<td>0.16 (0.50)</td>
<td>0.18 (0.56)</td>
<td>0.624</td>
</tr>
<tr>
<td>Sleep disorder</td>
<td>0.97 (0.57)</td>
<td>0.97 (0.56)</td>
<td>0.97 (0.57)</td>
<td>0.170</td>
</tr>
<tr>
<td>Use of sleep medicine</td>
<td>0.15 (0.59)</td>
<td>0.13 (0.55)</td>
<td>0.18 (0.62)</td>
<td>1.482</td>
</tr>
<tr>
<td>Daytime dysfunction</td>
<td>0.50 (0.77)</td>
<td>0.52 (0.78)</td>
<td>0.48 (0.77)</td>
<td>-0.765</td>
</tr>
<tr>
<td>Loneliness (ULS-6\textsuperscript{e})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>8.29 (2.93)</td>
<td>8.58 (3.03)</td>
<td>8.00 (2.79)</td>
<td>-3.442</td>
</tr>
<tr>
<td>Often feel lack of friends</td>
<td>1.45 (0.76)</td>
<td>1.53 (0.79)</td>
<td>1.36 (0.71)</td>
<td>-3.704</td>
</tr>
<tr>
<td>Often feel no one can be trusted</td>
<td>1.42 (0.72)</td>
<td>1.47 (0.73)</td>
<td>1.37 (0.71)</td>
<td>-2.326</td>
</tr>
<tr>
<td>Often feel left out</td>
<td>1.32 (0.59)</td>
<td>1.37 (0.62)</td>
<td>1.28 (0.55)</td>
<td>-2.672</td>
</tr>
<tr>
<td>Often feel separated from others</td>
<td>1.34 (0.65)</td>
<td>1.37 (0.67)</td>
<td>1.30 (0.64)</td>
<td>-1.76</td>
</tr>
<tr>
<td>Often feel shy</td>
<td>1.31 (0.58)</td>
<td>1.36 (0.64)</td>
<td>1.25 (0.52)</td>
<td>-3.365</td>
</tr>
<tr>
<td>Often feel surrounded by people but not cared for</td>
<td>1.46 (0.71)</td>
<td>1.49 (0.72)</td>
<td>1.43 (0.71)</td>
<td>-1.265</td>
</tr>
</tbody>
</table>

\textsuperscript{a}MOA: migrant older adults.  
\textsuperscript{b}LOA: local older adults.  
\textsuperscript{c}SSRS: Social Support Rating Scale.  
\textsuperscript{d}PSQI: Pittsburgh Sleep Quality Index.  
\textsuperscript{e}ULS-6: 6-item short-form University of California, Los Angeles (UCLA) Loneliness Scale.
The mean total score of social support was 38.89 (SD 6.63) for the MOA and 39.51 (SD 6.86) for the LOA. Specifically, the mean scores of individual items were as follows: objective social support, 8.47 (SD 1.64) for the MOA and 7.95 (SD 2.30) for the LOA; subjective social support, 23.47 (SD 4.79) for the MOA and 24.43 (SD 4.40) for the LOA; and social support use, 6.94 (SD 2.26) for the MOA and 7.14 (SD 2.44) for the LOA. Statistical differences were found for objective social support ($t_{1203}=-4.545$, $P<.001$) and subjective social support ($t_{1203}=3.608$, $P<.001$) between the MOA and the LOA.

The Structural Model

Measurement Invariance Across the Migrant and Local Groups

Before the discussion of the migrant-local difference in the structural model, multigroup model invariance should be first tested. Partial measurement invariance was used to clarify the measurement invariance about the multigroup models in this study, which could assess the invariance with $\Delta$CFI (change in the CFI) and $\Delta$RMSEA (change in the RMSEA) for comparing the less restricted model with the more constrained model. The basic test strategy was to outline correspondence to model trimming, where an initial unconstrained model was gradually restricted by adding constrains [42]. Table 3 illustrates the related fit statistics of the measures' invariance across the migrant-local difference and the fitness index for the 7 chosen models. Fitness indices of the MOA (M1) and LOA (M2) models were first compared to test whether the variable migrant-local difference was suitable for group comparison. As shown in Table 3, the model fitness indices between the M1 and M2 were $\Delta$CFI=0 and $\Delta$RMSEA=0, indicating that the different effects between the MOA and LOA could be compared. Next, the measurement invariance was evaluated with $\Delta$CFI and $\Delta$RMSEA between the unconstrained model (M3), the measurement weights model (M4), the structural weights model (M5), the structural covariances model (M6), and the structural residuals model (M7). M3 did not restrict any coefficients, M4 supposed that the indicator loadings for the corresponding construct in each group were the same, M5 constrained both indicator loadings for the corresponding construct and the structural coefficients across the groups, M6 assumed that the indicator loadings for the corresponding construct and the structural coefficients across the groups were equal and also that the covariance of the endogenous variables across the groups were equal, and M7 assumed that the indicator loadings, structural coefficients, covariance of the endogenous variables, and variance of the exogenous variable were equal across the groups.

Table 3. Multigroup model invariance test results using AMOS software. Variables in 7 models were social support, sleep disturbance, and loneliness among the MOA$^a$ and LOA$^b$ in Weifang City, China, in August 2021 (N=1205).

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$ (df)</th>
<th>P value</th>
<th>$\chi^2$/df</th>
<th>GFI$^c$</th>
<th>AGFI$^d$</th>
<th>CFI$^e$</th>
<th>RMSEA$^f$</th>
<th>$\Delta$CFI$^g$</th>
<th>$\Delta$RMSEA$^h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1$^i$</td>
<td>579.952 (202)</td>
<td>&lt;.001</td>
<td>2.871</td>
<td>0.941</td>
<td>0.920</td>
<td>0.925</td>
<td>0.039</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>M2$^k$</td>
<td>579.952 (202)</td>
<td>&lt;.001</td>
<td>2.871</td>
<td>0.941</td>
<td>0.920</td>
<td>0.925</td>
<td>0.039</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M3$^l$</td>
<td>579.952 (202)</td>
<td>&lt;.001</td>
<td>2.871</td>
<td>0.941</td>
<td>0.920</td>
<td>0.925</td>
<td>0.039</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M4$^m$</td>
<td>593.973 (215)</td>
<td>&lt;.001</td>
<td>2.763</td>
<td>0.939</td>
<td>0.923</td>
<td>0.925</td>
<td>0.038</td>
<td>0</td>
<td>-0.001</td>
</tr>
<tr>
<td>M5$^n$</td>
<td>597.464 (218)</td>
<td>&lt;.001</td>
<td>2.741</td>
<td>0.939</td>
<td>0.924</td>
<td>0.924</td>
<td>0.038</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>M6$^o$</td>
<td>598.632 (219)</td>
<td>&lt;.001</td>
<td>2.733</td>
<td>0.939</td>
<td>0.924</td>
<td>0.924</td>
<td>0.038</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>M7$^p$</td>
<td>601.83 (221)</td>
<td>&lt;.001</td>
<td>2.723</td>
<td>0.939</td>
<td>0.924</td>
<td>0.924</td>
<td>0.038</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
</tbody>
</table>

$^a$MOA: migrant older adults.
$^b$LOA: local older adults.
$^c$GFI: goodness-of-fit index.
$^d$AGFI: adjusted goodness-of-fit index.
$^e$CFI: comparative fitness index.
$^f$RMSEA: root mean square error of approximation.
$^g$ΔCFI: change in the CFI.
$^h$ΔRMSEA: change in the RMSEA.
$^i$M1: MOA model.
$^j$Not applicable.
$^k$M2: LOA model.
$^l$M3: unconstrained model.
$^m$M4: measurement weights model.
$^n$M5: structural weights model.
$^o$M6: structural covariances model.
$^p$M7: structural residuals model.
According to Table 3, ΔCFI was 0 between M4 and M3 and –0.001 between M5 and M4, between M6 and M5, and between M7 and M6, while ΔRMSEA was –0.001 between M4 and M3, between M5 and M4, between M6 and M5, and between M7 and M6. The fact that all the ΔCFI values were <0.010 and all the ΔRMSEA values were <0.015 indicated that the M1, M2, M3, M4, M5, M6, and M7 models had established measurement invariance across the migrant-local difference [43]. The comparison of the values between different variables across the MOA and LOA was thus established.

**Model Fitness Indices**

The proposed unconstrained model (M3) for the MOA and LOA is illustrated in Figures 1 and 2, respectively, and 3 latent variables (social support, sleep disturbance, and loneliness) were included in the model; the model fitness indices are also presented in Table 3. The estimates of model fitness were similar for both groups: GFI=0.941>0.90, AGFI=0.920>0.90, CFI=0.925>0.90, and RMSEA=0.039<0.05, implying that the proposed model well fitted the empirical data for both the MOA and the LOA in this study.

**Figure 1.** SEM analysis of the association between social support, sleep disturbance, and loneliness of MOA (n=613) in Weifang City, China, in August 2021. e: residual variables; MOA: migrant older adults; S1: subjective sleep quality; S2: sleep latency; S3: sleep continuity; S4: habitual sleep efficiency; S5: sleep disorder; S6: use of sleep medicine; S7: daytime dysfunction; SEM: structural equation modeling.

**Figure 2.** SEM analysis of the association between social support, sleep disturbance, and loneliness of LOA (n=592) in Weifang City, China, in August 2021. e: residual variables; LOA: local older adults; S1: subjective sleep quality; S2: sleep latency; S3: sleep continuity; S4: habitual sleep efficiency; S5: sleep disorder; S6: use of sleep medicine; S7: daytime dysfunction; SEM: structural equation modeling.
Relationship Between Social Support, Sleep Disturbance, and Loneliness

The relationship between social support, sleep disturbance, and loneliness is illustrated in Table 4 and in Figures 1 and 2 among the MOA and LOA, respectively.

Table 4. Standardized effects between social support, sleep disturbance, and loneliness among the MOA\(^a\) and LOA\(^b\) by the migrant-local difference in Weifang, China, in August 2021 (N=1205).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Direct effects</th>
<th>Indirect effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MOA</td>
<td>LOA</td>
</tr>
<tr>
<td>Social support → loneliness</td>
<td>-0.44 ((P&lt;.001))</td>
<td>-0.40 ((P&lt;.001))</td>
</tr>
<tr>
<td>Sleep disturbance → loneliness</td>
<td>0.13 ((P=.03))</td>
<td>0.22 ((P&lt;.001))</td>
</tr>
<tr>
<td>Social support → sleep disturbance</td>
<td>-0.24 ((P&lt;.001))</td>
<td>-0.20 ((P&lt;.001))</td>
</tr>
<tr>
<td>Social support → sleep disturbance → loneliness</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

\(^a\)MOA: migrant old adults.

\(^b\)LOA: local old adults.

\(^c\)Not applicable.

Association Between Social Support and Loneliness

Social support and its indicators exerted a negative effect on loneliness. The direct negative effect of social support on loneliness was found among both MOA (standardized coefficient=-0.44) and LOA (standardized coefficient=-0.40), although the effect was stronger in the MOA than in the LOA. Additionally, the standardized effect of social support on loneliness was statistically significant in both groups. The results indicated that higher social support would result in less loneliness among both MOA and LOA.

Association Between Sleep Disturbance and Loneliness

Sleep disturbance exerted a direct positive effect on loneliness among both MOA (standardized coefficient=0.13) and LOA (standardized coefficient=0.22), although the effect was stronger in the LOA than in the MOA. Additionally, the standardized effect of sleep disturbance on loneliness was statistically significant in both groups. The results implied severer that sleep disturbance among both MOA and LOA would result in higher loneliness.

Association Between Social Support and Sleep Disturbance

Social support had a direct negative effect on sleep disturbance among both MOA (standardized coefficient=-0.24) and LOA (standardized coefficient=-0.20), although the effect was stronger in the LOA than in the MOA. A statistically significant relationship between social support and sleep quality was found in both groups. The results indicated that higher social support would result in less sleep disturbance.

Discussion

Principal Findings

The mean score of loneliness among the MOA (mean 8.58, SD 3.03) and the LOA (mean 8.00, SD 2.79) indicated a fairly low level of loneliness in both groups in this research, also lower than the mean score of loneliness reported in a previous study among the rural empty-nest older adults in China (mean 16.19, SD 3.90) [44]. Moreover, the MOA showed a mildly higher loneliness than the LOA in this study, which is similar to a previous study that showed that the mean score of loneliness is higher among older Turkish German adults than their native-born German counterparts [10].

Using SEM, this study investigated the migrant-local difference in the association between social support, sleep disturbance, and loneliness among MOA and LOA. First, the results illustrated that social support affected loneliness among both MOA and LOA, although the effect was stronger in the MOA than in the LOA. Second, sleep disturbance exerted a direct positive effect on loneliness among both MOA and LOA, although the effect was stronger in the LOA than in the MOA. Third, social support affected sleep disturbance negatively and directly among both MOA and LOA, although the effect was stronger in the MOA than in the LOA.

Relationship Between Social Support and Loneliness

A negative relationship between social support and loneliness was found among both MOA and LOA, which is consistent with previous research [45,46]. The association between social support and loneliness can be explained by the main effect and buffer effect models [47-49]. The main effect model holds that social support has a universal gain effect and that any increase in social support is bound to result in an increase in health status (eg, a lower level of loneliness), regardless of an individual’s social support level. In the buffer effect model, social support is only associated with physical and mental health in stressful situations, which cushions the negative effects of stressful events on physical and mental health and maintains or improves an individual’s physical and mental health [50]. Through these 2 models, social support can increase positive emotional experiences, control negative emotional experiences, and finally reduce loneliness.

Relationship Between Sleep Disturbance and Loneliness

Sleep disturbance was found could exert a positive effect on loneliness among both MOA and LOA, which is similar to a previous study that showed that individuals with poor sleep quality report higher levels of loneliness [51]. A study among
adolescents also showed that loneliness is significantly positively associated with sleep disturbance [52]. The positive association between sleep disturbance and loneliness may be due to the result of sleep disturbance–related metabolic, neural, and hormonal processes [51]; moreover, sleep disturbance may result in exhaustion in older adults, which further reduces their communication with others, leading to loneliness.

**Relationship Between Social Support and Sleep Disturbance**

A negative association between social support and sleep disturbance was found among both MOA and LOA, which is similar to previous studies that have reported that higher social support would generally predict less sleep disturbance [53-55]. Previous studies have also shown that sleep disturbance is significantly associated with social support among Chinese medical staff [56], as well as young Chinese rural residents [57] in China. Moreover, a meta-analysis showed a robust association between social support and favorable sleep outcomes [58]; that is, higher social support would result in less sleep disruption and less sleep disturbance.

**Migrant-Local Difference in the Relationship Between Social Support, Sleep Disturbance, and Loneliness Among MOA and LOA**

Regarding the migrant-local difference, a negative effect of social support on loneliness was found among both MOA and LOA, although the effect was higher in the MOA than in the LOA. The difference between these 2 groups may be because the MOA migrated to a new place and faced many issues with adaptation, such as acculturation stress [59,60], a low sense of psychological identity [61], and less use of outpatient mental health care services [62]; all these factors would increase their loneliness. In addition, after moving to the city where their children lived, the support network of the MOA in their hometowns broke but a new urban support network was not yet formed; consequently, the MOA faced a situation in which they lacked social support [63]. Thus, compared with the LOA, the MOA were more lonely with less social support, so the effect of social support on loneliness was higher among the MOA than the LOA.

Regarding the migrant-local difference, a direct positive effect of sleep disturbance on loneliness was found among both MOA and LOA, although the effect was higher in the MOA than in the LOA. This may be due to the result of sleep disturbance–related metabolic, neural, and hormonal processes [51]; moreover, sleep disturbance may result in exhaustion in older adults, which further reduces their communication with others, leading to loneliness.

**Implications**

Due to the migrant-local difference, different supportive measures need to be targeted for MOA and LOA. First, since the negative relationship between social support and loneliness in this study was stronger in the MOA, family members should provide more social support for older adults, especially for MOA, and provide them with more knowledge about the importance and prevention of loneliness, while providing social support. Second, the community should design more programs and create activities for MOA, which could increase the interpersonal communication between the MOA, as well as between the MOA and the LOA, by enhancing their peer social support, thus decreasing sleep disturbance and loneliness among both groups, especially the MOA. Third, since the positive relationship between sleep disturbance and loneliness in this study was stronger among the LOA, it is suggested that the community pay more attention to LOA (eg, no loud noises at night), while the family members should create a better sleeping environment (eg, adequate sleeping temperature and well-ventilated bedroom). Fourth, community health care providers should provide more mental health monitoring for both MOA and LOA, especially for older adults with sleep disturbance. Fifth, the findings of this study could attract more Chinese scholars to focus on the level of loneliness among MOA and LOA, especially sociocultural and physical health status contributors, and conduct comparative research between different areas among MOA and LOA in the future.

**Limitations**

There are several limitations of this study. First, we used cross-sectional data, which could not be well explored for causal relationships. Second, socioeconomic status plays a fundamental role in the study of population health, which was not explored in this study; we will examine this more in future studies. Third, some other factors that may also affect loneliness among MOA and LOA (eg, stress and depression) were not included in the study, which will be considered in future studies. Fourth, many other factors (eg, medication usage) might influence the sleep quality in older adults, and information related to health status is also a potential confounder of the effect of the quality of sleep, so future studies should examine whether other mediators can explain the internal mechanism underlying the relationship between sleep disturbance and social support.

**Conclusion**

This study illustrated the migrant-local difference in the association between social support, sleep quality, and loneliness between MOA and LOA in Weifang City, China. The results...
showed that (1) social support exerts a negative effect on loneliness among both MOA and LOA, although the effect is higher in MOA; (2) loneliness is significantly and positively associated with sleep disturbance among both MOA and LOA, although the effect is higher in LOA; and (3) there is a negative association between social support and sleep disturbance among both MOA and LOA, yet it is slightly higher in the MOA. This study may provide helpful information on how to decrease and prevent loneliness among older adults by enhancing their social support and decreasing their sleep disturbance.

Acknowledgments
The research team greatly appreciates the funding support obtained and thanks the research participants for their cooperation and support. This study was supported and funded by the National Natural Science Foundation of China (#71804094), the China Postdoctoral Science Foundation (#2016MS92161), the Natural Science Foundation of Shandong Province (#ZR2016GB02), the Postdoctoral Science Foundation of Shandong Province (#201603021), and the Fundamental Research Funds for the Central Universities (#2022KJIGL01, #2018JC055).

Data Availability
The data sets generated and analyzed during this study are not publicly available due to privacy restrictions but are available from the corresponding author upon reasonable request.

Authors’ Contributions
Conceptualization, validation, and writing—review and editing were managed by FK and MP; methodology, software, formal analysis, writing—original draft preparation, and visualization, MP; investigation and data curation, FK, MP, JW, RC, HL, and XX; resources, project administration, and funding acquisition, FK; and supervision, FK, MZ, and SL. No generative AI was used in any portion of the manuscript writing. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

References


https://publichealth.jmir.org/2024/1/e49253


61. Meng Y. Exploring the feasible path of community integration for the elderly who have moved with the family--a case study of Y community in Z city. Rural Econ Sci-Technol 2022;33(02):252-253.


Abbreviations
AGFI: adjusted goodness-of-fit index
CFI: comparative fit index
GFI: goodness-of-fit index
LOA: local older adults
MOA: migrant older adults
PSQI: Pittsburgh Sleep Quality Index
PSU: primary sampling unit
Q1: first quartile
Q2: second quartile
Q3: third quartile
Q4: fourth quartile
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SSRS: Social Support Rating Scale
SSU: secondary sampling unit
UCLA: University of California at Los Angels
ULS-6: 6-item short-form University of California, Los Angeles (UCLA) Loneliness Scale
Impact of the Narcotics Information Management System on Opioid Use Among Outpatients With Musculoskeletal and Connective Tissue Disorders: Quasi-Experimental Study Using Interrupted Time Series

Iyn-Hyang Lee1*, PhD; So Young Kim2*, PharmD; Susin Park3, PhD; Jae Gon Ryu4, PharmD; Nam Kyung Je5,6, PharmD, PhD

1College of Pharmacy, Yeungnam University, Gyeongsan, Republic of Korea
2Department of Pharmacy, Kosin University Gospel Hospital, Busan, Republic of Korea
3College of Pharmacy, Woosuk University, Wanju, Republic of Korea
4Department of Pharmacy, Sungkyunkwan University Samsung Changwon Hospital, Changwon, Republic of Korea
5College of Pharmacy, Pusan National University, Busan, Republic of Korea
6Research Institute for Drug Development, Pusan National University, Busan, Republic of Korea

*these authors contributed equally

Corresponding Author:
Nam Kyung Je, PharmD, PhD
College of Pharmacy
Pusan National University
2 Busandaehak-ro, 63 Beon-gil, Geumjeong-gu
Busan, 46241
Republic of Korea
Phone: 82 +82515102802
Email: jenk@pusan.ac.kr

Abstract

Background: Opioids have traditionally been used to manage acute or terminal pain. However, their prolonged use has the potential for abuse, misuse, and addiction. South Korea introduced a new health care IT system named the Narcotics Information Management System (NIMS) with the objective of managing all aspects of opioid use, including manufacturing, distribution, sales, disposal, etc.

Objective: This study aimed to assess the impact of NIMS on opioid use.

Methods: We conducted an analysis using national claims data from 45,582 patients diagnosed with musculoskeletal and connective tissue disorders between 2016 and 2020. Our approach included using an interrupted time-series analysis and constructing segmented regression models. Within these models, we considered the primary intervention to be the implementation of NIMS, while we treated the COVID-19 outbreak as the secondary event. To comprehensively assess inappropriate opioid use, we examined 4 key indicators, as established in previous studies: (1) the proportion of patients on high-dose opioid treatment, (2) the proportion of patients receiving opioid prescriptions from multiple providers, (3) the overlap rate of opioid prescriptions per patient, and (4) the naloxone use rate among opioid users.

Results: During the study period, there was a general trend of increasing opioid use. After the implementation of NIMS, significant increases were observed in the trend of the proportion of patients on high-dose opioid treatment (coefficient=0.0271; P=.01) and in the level of the proportion of patients receiving opioid prescriptions from multiple providers (coefficient=0.6252; P=.004). An abrupt decline was seen in the level of the naloxone use rate among opioid users (coefficient=−0.2968; P=.04). While these changes were statistically significant, their clinical significance appears to be minor. No significant changes were observed after both the implementation of NIMS and the COVID-19 outbreak.

Conclusions: This study suggests that, in its current form, the NIMS may not have brought significant improvements to the identified indicators of opioid overuse and misuse. Additionally, the COVID-19 outbreak exhibited no significant influence on opioid use patterns. The absence of real-time monitoring feature within the NIMS could be a key contributing factor. Further exploration and enhancements are needed to maximize the NIMS’ impact on curbing inappropriate opioid use.
Opioids have traditionally been used to manage acute or terminal pain [1-3]. However, their prolonged use has generated controversy due to the potential risk for abuse, overdose, and addiction [4], leading to negative outcomes and excessive mortality rates [3,5]. Previously confined to cancer-related pain management before the 1980s [6], opioids have gradually expanded their application to noncancer pain management, resulting in an increased incidence of opioid use disorders [7]. The overprescription of opioids has even steered individuals toward illicit substances such as heroin [8]. In 2018, opioid overdoses were responsible for over 60,000 deaths in the United States, making them the leading cause of drug-related deaths [9,10]. This worrisome trend has transcended national borders, manifesting as a significant global public health challenge [11-15]. An ecological study conducted in Spain corroborated this concern, demonstrating a direct correlation between greater prescribed opioid drug availability and opioid-related mortality [16].

The use of opioids in South Korea was lower compared to the United States and Europe, according to the International Narcotics Control Board’s report [17]. However, in recent years, there has been a rapid increase in the prescription rate of opioids [18]. For instance, from 2002 to 2015, the number of patients using opioids chronically for more than 90 days increased 6-9 times [19]. To address this issue, the Ministry of Food and Drug Safety (MFDS) implemented the Narcotics Information Management System (NIMS) on May 18, 2018 [20,21]. Before the NIMS, health care institutions relied on manual record-keeping to manage the use and inventory of opioids [22]. The NIMS is a web-based system that tracks patient information and inventory using a serial number assigned to each unit (bottle and carton) of opioid [22]. The system stores and monitors information at all times, making it easier to track and manage narcotics. The introduction of NIMS has transformed the monitoring of drug handling from a random management process to a selective on-site monitoring system by analyzing drug handling information. The MFDS aims to establish a safer narcotics management network through real-time tracking of narcotic handling history [21,22].

To date, there are no published studies that have directly evaluated the impact of NIMS on opioid use by comparing prescription patterns before and after the installation of the system. Therefore, we aimed to examine whether the NIMS led to any improvement in indicators related to inappropriate opioid use using an interrupted time-series (ITS) analysis. The study period coincided with the outbreak of COVID-19, and the analysis also sought to assess its effect on opioid use.

**Introduction**

Opioids have traditionally been used to manage acute or terminal pain [1-3]. However, their prolonged use has generated controversy due to the potential risk for abuse, overdose, and addiction [4], leading to negative outcomes and excessive mortality rates [3,5]. Previously confined to cancer-related pain management before the 1980s [6], opioids have gradually expanded their application to noncancer pain management, resulting in an increased incidence of opioid use disorders [7]. The overprescription of opioids has even steered individuals toward illicit substances such as heroin [8]. In 2018, opioid overdoses were responsible for over 60,000 deaths in the United States, making them the leading cause of drug-related deaths [9,10]. This worrisome trend has transcended national borders, manifesting as a significant global public health challenge [11-15]. An ecological study conducted in Spain corroborated this concern, demonstrating a direct correlation between greater prescribed opioid drug availability and opioid-related mortality [16].

The use of opioids in South Korea was lower compared to the United States and Europe, according to the International Narcotics Control Board’s report [17]. However, in recent years, there has been a rapid increase in the prescription rate of opioids [18]. For instance, from 2002 to 2015, the number of patients using opioids chronically for more than 90 days increased 6-9 times [19]. To address this issue, the Ministry of Food and Drug Safety (MFDS) implemented the Narcotics Information Management System (NIMS) on May 18, 2018 [20,21]. Before the NIMS, health care institutions relied on manual record-keeping to manage the use and inventory of opioids [22]. The NIMS is a web-based system that tracks patient information and inventory using a serial number assigned to each unit (bottle and carton) of opioid [22]. The system stores and monitors information at all times, making it easier to track and manage narcotics. The introduction of NIMS has transformed the monitoring of drug handling from a random management process to a selective on-site monitoring system by analyzing drug handling information. The MFDS aims to establish a safer narcotics management network through real-time tracking of narcotic handling history [21,22].

To date, there are no published studies that have directly evaluated the impact of NIMS on opioid use by comparing prescription patterns before and after the installation of the system. Therefore, we aimed to examine whether the NIMS led to any improvement in indicators related to inappropriate opioid use using an interrupted time-series (ITS) analysis. The study period coincided with the outbreak of COVID-19, and the analysis also sought to assess its effect on opioid use.

**Methods**

**Data Sources**

The analysis was performed using National Health Insurance (NHI) claims data between 2016 and 2020. For research purposes, the Health Insurance Review and Assessment Service (HIRA) annually provides patient sample data (HIRA–National Patient Sample [HIRA-NPS]). HIRA-NPS is derived from health insurance claim data submitted by medical institutions for reimbursement [23]. HIRA provides a sample representing 3% of the Korean population for 2016, 2017, and 2018 and a sample representing 2% of the Korean population for 2019 and 2020.

**Study Design**

To assess the impact of NIMS on opioid use, a segmented ITS analysis was performed. The data were aggregated monthly over the 5-year study period. As per the recommended practice, the ITS analysis required observations of at least 12 months before and after the intervention [24]. The NIMS was introduced on May 18, 2018, and the analysis period covered 28 months before the intervention and 32 months after. The study period also included the COVID-19 outbreak, which was treated as a second interruption to control its influence on opioid use. The second interruption occurred in February 2020 in South Korea.

**Study Population**

The study included patients with disorders of the musculoskeletal system and connective tissue [25], identified using the International Classification of Diseases, Tenth Revision (ICD-10) codes. The subject diseases are arthropathies (eg, osteoarthritis), dorsopathies (eg, spondylitis), soft tissue disorders (eg, fibromyalgia), and others (eg, postprocedural musculoskeletal disorders). Patients younger than 20 years and those with a cancer diagnosis were excluded. Outpatients who received opioid prescriptions were selected as target patients (Figure 1). The patient population was further categorized by age groups (20-39, 40-59, 60-74, and over 75 years) and the types of insurance (NHI and Medical Aid [MedAid] or Patriots and Veterans Insurance Plan [PVI]). NHI covers most of the study population, while MedAid includes individuals with low income. PVI is a special health program for patriots and veterans who have served in the military or government agencies. While PVI is distinct from NHI and MedAid in terms of its target population, it is grouped together with patients with MedAid insurance in our study due to their similar patterns of drug use [26].
**Study Drugs**

The study aimed to analyze the use of outpatient opioids in oral and patch formulations. The following opioid medications were included in the analysis: buprenorphine transdermal, codeine oral, dihydrocodeine oral, fentanyl transdermal, hydrocodone oral, hydromorphone oral, morphine oral, oxycodone oral, and tapentadol oral formulations (Multimedia Appendix 1). However, sublingual tablets, buccal tablets, and nasal sprays were excluded from the analysis as they are primarily used for breakthrough pain in patients with cancer, in accordance with the Korean reimbursement criteria [27].

**Ethical Considerations**

The study protocol was approved by the institutional review board of Pusan National University, Busan, South Korea (PNU IRB/2022_27_HR).
Outcome Measures

Overview
Throughout the study period, we assessed inappropriate opioid use monthly using 4 key indicators, each carefully selected to encompass different dimensions of inappropriate opioid use based on previous research in this area [28-33]. These indicators included (1) the proportion of patients on high-dose opioid treatment, (2) the proportion of patients receiving opioid prescriptions from multiple providers, (3) the overlap rate of opioid prescriptions per patient, and (4) the naloxone use rate among opioid users.

The Proportion of Patients on High-Dose Opioid Treatment
The proportion of patients on high-dose opioid treatment refers to the percentage of patients prescribed high-dose opioid treatments out of all patients included. High-dose opioid treatment was determined by assessing their daily morphine milligram equivalents (MME), a standardized measure of opioid potency that allows for comparing and aggregating each patient’s opioid consumption. For our study’s purposes, a daily dose of 100 MME or higher was considered high-dose opioid treatment [28-30]. Such treatment has been associated with elevated rates of mental health issues, including substance use disorders, and increased health care service use when compared to cases of no opioid use or low-dose opioid treatment [30].

The Proportion of Patients Receiving Opioid Prescriptions From Multiple Providers
The proportion of patients receiving opioid prescriptions from multiple providers was calculated to evaluate the dispersion of patient care within the outpatient setting [31]. This computation encompassed patients obtaining prescriptions from 2 or more providers, irrespective of the provider’s institutional affiliation. This was accomplished by summing up the number of providers who issued opioid prescriptions to each patient monthly. The practice of opioid prescribing by multiple providers is widespread and has been linked to elevated rates of hospital admissions directly associated with opioid use [31].

The Overlap Rate of Opioid Prescriptions Per Patient
The ratio of overlapping narcotic prescriptions is a measure of the extent to which the opioid prescriptions of study drugs overlap for each patient. It is calculated as the ratio of the overlapping prescription period to the total prescription period of the study drug per patient. Overlapping opioid prescriptions stand as one of the indicators signifying potential opioid misuse [32].

The Naloxone Use Rate Among Opioid Users
The percentage of patients prescribed and administered naloxone among opioid users each month is calculated as a proxy for opioid overdose incidents. Naloxone is a widely used medication for reversing the effects of opioid overdose [33], and the proportion of patients receiving naloxone reflects the frequency of opioid overdose or toxicity incidents.

Statistical Analysis
Descriptive statistics for all variables were presented. The time-series data of outcomes were examined graphically and used to establish a segmented regression model to assess the statistical significance of the effect of NIMS, which was investigated by analyzing the change in opioid use using 4 outcomes [34]. The time-series data of these outcomes were analyzed over 60 months, from January 2016 to December 2020, with the NIMS intervention occurring in May 2018 (29th month). The model to be tested in this study is as follows:

\[ Y_t = \beta_0 + \beta_1 \times \text{time} + \beta_2 \times \text{NIMS} + \beta_3 \times \text{time after NIMS} + \beta_4 \times \text{COVID19} + \beta_5 \times \text{time after COVID19} + \epsilon_t \]

where \( Y_t \) is the outcome of interest in month \( t \); time is a continuous variable, indicating time in months at time \( t \) from the start of the observation period; \( \text{NIMS} \) is an indicator for time \( t \) occurring before (\( \text{NIMS}=0 \)) or after (\( \text{NIMS}=1 \)) the NIMS, which was implemented at month 29 in the series; and \( \text{time after NIMS} \) is a continuous variable counting the number of months after the NIMS at time \( t \), coded 0 before the NIMS and (time-28) after the NIMS. \( \beta_0 \) estimates the baseline level of the outcome at zero; \( \beta_1 \) estimates the baseline trend before the NIMS; \( \beta_2 \) estimates the level change at the month of NIMS introduction; \( \beta_4 \) estimates the change in the trend after the NIMS, compared with the monthly trend before the NIMS; \( \beta_5 \) estimates the level change at the month of the COVID-19 outbreak at month 50 in the series; and \( \beta_5 \) estimates the change in the trend after the COVID-19, compared with the monthly trend before the COVID-19 outbreak.

The sum of \( \beta_1 \) and \( \beta_2 \) is the postintervention slope. As can be seen from the model, each interval is represented by a level and a trend. The change in the level of the outcome after the intervention indicates the immediate effect of the intervention. The change in trend is defined as the change in slope in the interval after the intervention compared to the interval before the intervention [34]. The Durbin-Watson test was used to assess the serial correlation of error terms and estimate the regression coefficients with either an ordinary least squares or a first-order autocorrelation maximum likelihood estimate, depending on the significance of serial correlations [35]. Residual analyses based on autocorrelation plots and partial autocorrelation plots were carried out to review the goodness-of-fit of the model. Statistical analysis was performed using R software (version 4.1.1; R Foundation for Statistical Computing), and when the \( P \) value was <.05, it was considered statistically significant.

Results
Characteristics of Study Population
A total of 45,582 patients received outpatient opioid prescriptions, and the most common subject disorders were dorsopathies, followed by arthropathies and soft tissue disorders. The study patients were predominantly female (27,756/45,582, 60.9%), with the largest age group being 60-74 years (15,752/45,582, 34.6%). The percentage of MedAid beneficiaries among the study patients ranged from 8.8%...
(2189/24,924) to 10.7% (539/5016), which was higher than the general population (which is approximately 3%). Some fluctuations were observed in sex, age distribution, and insurance type between the 3 time periods (Table 1).

Table 1. Characteristics of the study population and overall changes in 4 outcome variables from January 2016 to December 2020 (N=45,582).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (60 months; N=45,582)</th>
<th>Pre-NIMS&lt;sup&gt;a&lt;/sup&gt; (28 months; January 2016-April 2018; n=24,924)</th>
<th>Post-NIMS Pre–COVID-19 (21 months; May 2018-January 2020; n=15,642)</th>
<th>Post–COVID-19 (11 months; February 2020-December 2020; n=5016)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17,826 (39.1)</td>
<td>9591 (38.5)</td>
<td>6125 (39.2)</td>
<td>2110 (42.1)</td>
</tr>
<tr>
<td>Female</td>
<td>27,756 (60.9)</td>
<td>15,333 (61.5)</td>
<td>9517 (60.8)</td>
<td>2906 (57.9)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-39</td>
<td>5706 (12.5)</td>
<td>3236 (13)</td>
<td>1951 (12.5)</td>
<td>519 (10.3)</td>
</tr>
<tr>
<td>40-59</td>
<td>15,079 (33.1)</td>
<td>8401 (33.7)</td>
<td>5148 (32.9)</td>
<td>1530 (30.5)</td>
</tr>
<tr>
<td>60-74</td>
<td>15,752 (34.6)</td>
<td>8518 (34.2)</td>
<td>5397 (34.5)</td>
<td>1837 (36.6)</td>
</tr>
<tr>
<td>≥75</td>
<td>9045 (19.8)</td>
<td>4769 (19.1)</td>
<td>3146 (20.1)</td>
<td>1130 (22.5)</td>
</tr>
<tr>
<td><strong>Insurance type, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHI&lt;sup&gt;b&lt;/sup&gt;</td>
<td>41,452 (90.9)</td>
<td>22,735 (91.2)</td>
<td>14,240 (91)</td>
<td>4477 (89.3)</td>
</tr>
<tr>
<td>MedAid&lt;sup&gt;c&lt;/sup&gt; or PVI&lt;sup&gt;d&lt;/sup&gt;</td>
<td>4130 (9.1)</td>
<td>2189 (8.8)</td>
<td>1402 (9)</td>
<td>539 (10.7)</td>
</tr>
<tr>
<td><strong>Morbidity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arthropathies</td>
<td>11,902 (26.1)</td>
<td>6533 (26.2)</td>
<td>4092 (26.2)</td>
<td>1277 (25.5)</td>
</tr>
<tr>
<td>Dorsopathies</td>
<td>21,234 (46.6)</td>
<td>11,588 (46.5)</td>
<td>7283 (46.6)</td>
<td>2363 (47.1)</td>
</tr>
<tr>
<td>Soft tissue disorders</td>
<td>11,908 (26.1)</td>
<td>6534 (26.2)</td>
<td>4086 (26.1)</td>
<td>1288 (25.7)</td>
</tr>
<tr>
<td>Others&lt;sup&gt;e&lt;/sup&gt;</td>
<td>538 (1.2)</td>
<td>269 (1.1)</td>
<td>181 (1.2)</td>
<td>88 (1.8)</td>
</tr>
<tr>
<td><strong>Outcome variables, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of patients on high-dose opioid treatment</td>
<td>N/A&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.75 (0.22)</td>
<td>1.25 (0.27)</td>
<td>1.89 (0.30)</td>
</tr>
<tr>
<td>Proportion of patients receiving opioid prescriptions from multiple providers</td>
<td>N/A&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.69 (0.27)</td>
<td>1.02 (0.27)</td>
<td>1.37 (0.37)</td>
</tr>
<tr>
<td>Overlap rate of opioid prescriptions per patient</td>
<td>N/A&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1.95 (0.42)</td>
<td>2.17 (0.20)</td>
<td>2.30 (0.30)</td>
</tr>
<tr>
<td>Naloxone use rate among opioid users</td>
<td>N/A&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.57 (0.28)</td>
<td>0.51 (0.16)</td>
<td>0.45 (0.26)</td>
</tr>
</tbody>
</table>

<sup>a</sup>NIMS: Narcotics Information Management System.
<sup>b</sup>NHI: National Health Insurance.
<sup>c</sup>MedAid: Medical Aid.
<sup>d</sup>PVI: Patriots and Veterans Insurance.
<sup>e</sup>Other disorders of the musculoskeletal system and connective tissue.
<sup>f</sup>N/A: not applicable.

The Proportion of Patients on High-Dose Opioid Treatment

The proportion of patients on high-dose opioid treatment was the lowest in the pre-NIMS period (187/24,924, 0.75%), followed by the post-NIMS pre–COVID-19 period (196/15,642, 1.25%), and then the highest in the post–COVID-19 period (95/5016, 1.89%; Table 1). The trend showed a significant increase of 0.0271 from 0.0040 at baseline to 0.0311 (P=.01) after the implementation of NIMS, with no significant change in the level. The COVID-19 outbreak had little impact on both the level and trend of the percentage of patients receiving high-dose opioid treatment. The regression results are shown in Table 2 and Figure 2A.
Table 2. Modeling interrupted time series to evaluate the impact of the Narcotics Information Management System (NIMS) and COVID-19 on opioid use among outpatients with musculoskeletal and connective tissue disorders in South Korea.

<table>
<thead>
<tr>
<th>Outcome variables and estimator (AR(^a))</th>
<th>Segmented regression coefficient</th>
<th>(P) value</th>
<th>D-W(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proportion of patients on high-dose opioid treatment</strong></td>
<td></td>
<td></td>
<td>1.9426</td>
</tr>
<tr>
<td>Intercept ((\beta_0))</td>
<td>0.7947</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Baseline trend ((\beta_1))</td>
<td>0.0040</td>
<td>.48</td>
<td></td>
</tr>
<tr>
<td>Level change after NIMS ((\beta_2))</td>
<td>0.1743</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>Trend change after NIMS ((\beta_3))</td>
<td>0.0271</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Level change after COVID-19 ((\beta_4))</td>
<td>0.0484</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Trend change after COVID-19 ((\beta_5))</td>
<td>0.0309</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td><strong>Proportion of patients receiving opioid prescriptions from multiple providers</strong></td>
<td></td>
<td></td>
<td>2.0825</td>
</tr>
<tr>
<td>Intercept ((\beta_0))</td>
<td>0.8141</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Baseline trend ((\beta_1))</td>
<td>−0.0076</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Level change after NIMS ((\beta_2))</td>
<td>0.6252</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>Trend change after NIMS ((\beta_3))</td>
<td>−0.0067</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>Level change after COVID-19 ((\beta_4))</td>
<td>0.3969</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>Trend change after COVID-19 ((\beta_5))</td>
<td>0.0323</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td><strong>Overlap rate of opioid prescriptions per patient</strong></td>
<td></td>
<td></td>
<td>2.0046</td>
</tr>
<tr>
<td>Intercept ((\beta_0))</td>
<td>1.4828</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Baseline trend ((\beta_1))</td>
<td>−0.0113</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Level change after NIMS ((\beta_2))</td>
<td>0.3349</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Trend change after NIMS ((\beta_3))</td>
<td>0.0101</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>Level change after COVID-19 ((\beta_4))</td>
<td>0.3709</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>Trend change after COVID-19 ((\beta_5))</td>
<td>−0.0442</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td><strong>Naloxone use rate among opioid users</strong></td>
<td></td>
<td></td>
<td>1.9854</td>
</tr>
<tr>
<td>Intercept ((\beta_0))</td>
<td>0.3685</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Baseline trend ((\beta_1))</td>
<td>0.0156</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Level change after NIMS ((\beta_2))</td>
<td>−0.2968</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>Trend change after NIMS ((\beta_3))</td>
<td>−0.0117</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>Level change after COVID-19 ((\beta_4))</td>
<td>0.0652</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>Trend change after COVID-19 ((\beta_5))</td>
<td>−0.0324</td>
<td>.18</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)AR: first-order autocorrelation maximum likelihood estimate.

\(^b\)D-W: Durbin-Watson test.
Impact of Narcotics Information Management System (NIMS) and COVID-19 on 4 outcome variables from January 2016 to December 2020: interrupted time-series analysis with observed and predicted regression lines for the (A) proportion of patients on high-dose opioid treatment; (B) proportion of patients receiving opioid prescriptions from multiple providers; (C) overlap rate of opioid prescriptions per patient; and (D) naloxone use rate among opioid users. The first interruption was held in May 2018 and the second was held in February 2020. The blue solid line represents the actual data, while the orange and green dot-dash lines represent the predicted data after NIMS and the COVID-19 outbreak, respectively.

The Proportion of Patients Receiving Opioid Prescriptions From Multiple Providers

The proportion of patients receiving opioid prescriptions from multiple providers was found to be the lowest in the pre-NIMS period (172/24,924, 0.69%), followed by the post-NIMS pre–COVID-19 period (160/15,642, 1.02%), and then the highest in the post–COVID-19 period (69/5016, 1.37%; Table 1). A sudden increase in the level was seen after the introduction of NIMS ($\beta_2 = .6252; P = .004$). A slight increase was observed after the COVID-19 outbreak, which was not significant ($\beta_4 = .3969; P = .09$). No significant change in the trend of patients receiving opioid prescriptions from multiple providers was observed after the implementation of NIMS or the COVID-19 outbreak. The segmented regression results of the trend and level changes are displayed in Table 2 and Figure 2B.

The Overlap Rate of Opioid Prescriptions Per Patient

The overlap rate of opioid prescriptions per patient showed a slight increase, with the lowest in the pre-NIMS period (486/24,924, 1.95%), followed by the post-NIMS pre–COVID-19 period (339/15,642, 2.17%), and then the highest in the post–COVID-19 period (102/5016, 2.30%; Table 1).

Both the NIMS and COVID-19 did not appear to significantly impact the level and trend of the overlap rate of opioid prescriptions per patient (Table 2). The trend and level changes are illustrated in Figure 2C.

The Naloxone Use Rate Among Opioid Users

The naloxone use rate among opioid users declined over time, from 0.57% (142/24,924) in the pre-NIMS period, to 0.51% (80/15,642) after NIMS, and then further down to 0.45% (23/5016) in the post–COVID-19 period (Table 1).

The baseline trend ($\beta_1$) before NIMS introduction was .0156 ($P = .01$), which was turned down by $-0.0117$ to 0.0039 after NIMS. COVID-19 changed little of the trend ($P = .18$). Abrupt decline was observed after the introduction of NIMS ($\beta_2 = -.2968; P = .04$). The regression results are shown in Table 2 and Figure 2D.

Discussion

Overview

Our study aimed to assess the impact of the Korean health care IT for narcotics management, known as the NIMS, on opioid misuse using nationally representative claims data. The study findings indicate that the NIMS did not demonstrate significant improvement during the 32 months since its introduction in the 4 specific indicators we selected. Rather, we observed a gradual increase in high-dose opioid prescribing and an abrupt increase in multiple providers’ prescriptions following the implementation of NIMS. While these changes were statistically significant, their clinical significance appears to be relatively minor, warranting further investigation.

The use rate of naloxone, which serves as a proxy for opioid misuse or abuse, had been increasing over time before the implementation of NIMS, and a temporary decrease in the use rate of naloxone was observed after NIMS implementation. This decrease suggests that the NIMS may have had a positive impact on opioid-related outcomes by reducing the occurrence of opioid misuse or abuse. However, it is inconclusive because...
a concurrent increase in the number of opioid users during the same period may have also contributed to the decrease in naloxone use rate.

We also conducted a second analysis, not treating COVID-19 as an interruption, and found no significant differences. While the primary focus of our study is to assess the impact of NIMS, this additional analysis confirms that even when not considering COVID-19 as an interruption, the NIMS still did not show a significant impact on the identified indicators of opioid overuse and misuse.

Given the limited existing research in this area, the reasons for NIMS not yielding anticipated results remain largely unexplored. In this regard, some recent studies have raised workforce concerns [36]. A recent study reported that the workforce to operate the NIMS is not adequately secured. Essential tasks such as analysis of causes of misuse and corrective actions for detected problems were carried out only to a limited extent due to the lack of workforce reinforcement [36]. Furthermore, recent in-depth interviews with 3 prescribers and 2 pharmacists confirmed that health care professionals primarily use the NIMS to fulfill legal reporting requirements [37]. These latest studies have raised the need for the introduction of a more comprehensive measure such as “opioid stewardship,” which encompasses guidelines for health care professionals, education for patients and the public, and enhanced communication and coordination among key stakeholders [38,39].

In comparison to similar systems overseas, the lack of real-time access to patient’s prescription histories within the NIMS may explain its suboptimal impact on drug usage enhancement [21]. The United States has a state-based electronic database called prescription drug monitoring programs (PDMPs), and some Canadian provinces operate the narcotics monitoring system (NMS), which mandates that pharmacies submit dispensing information for all monitored drugs [21]. NIMS, PDMPs, and NMS share the common goal of monitoring and managing the prescription and dispensing of narcotics to address misuse and abuse [21]. While they share this overarching objective, these systems differ in their operations and functionalities compared to the NIMS. All 3 systems collect, monitor, and analyze prescription data for controlled substances, including opioids. However, a key distinction is that PDMPs and NMS provide real-time access to patients’ prescription histories for prescribers and pharmacists, enabling them to promptly identify potential misuse, overuse, or diversion of prescription medications [40]. Numerous studies have investigated the effectiveness of PDMPs and NMS, with some demonstrating a significant reduction in the prescription of narcotic drugs and a reduction in opioids misuse [41-44]. These findings underscore the pivotal role of real-time monitoring in curbing prescription abuse and misuse.

The study population had a higher representation of women than men, who are known to experience more severe and long-lasting pain, leading to a higher prescription rate for opioids [45]. Additionally, it is noteworthy that our study population included a higher proportion of MedAid beneficiaries (ranging from 8.8% to 10.7%) compared to the general population, where the percentage of MedAid beneficiaries typically remains around 3% [46]. Several factors may contribute to the higher presentation of MedAid beneficiaries in our study. Research has shown that this population often experiences higher rates of chronic pain [47,48] and may have limited access to alternative pain management treatments [49,50]. Moreover, one important aspect to consider is the distinction in copayment requirements between NHI and MedAid. Under the NHI system, patients are required to make copayments for health care services, including medications such as opioids. On the other hand, MedAid beneficiaries generally pay lower or no copayments for medications, including opioids. This distinction can lead to different use patterns, a phenomenon often referred to as “moral hazard” [26,51,52]. While our study primarily focused on assessing the impact of NIMS on opioid use, the observed variation in MedAid beneficiaries’ opioid use patterns highlight an avenue for further research.

To the best of our knowledge, this research is the first study to directly evaluate the impact of the NIMS on opioid use using nationwide real-world data, providing invaluable insights into its effect on opioid prescription patterns in South Korea and suggesting areas for improvement. Nevertheless, our study has several limitations that should be noted. First, the inclusion of patients with musculoskeletal and connective tissue disorders may limit the generalizability of the findings to other patients with different medical conditions. Furthermore, the study’s setting in South Korea may limit the generalizability of the findings to other health care systems and countries with different opioid prescription practices and regulatory frameworks. Variations in health care policies, cultural factors, and clinical guidelines may influence opioid use patterns in distinct ways, warranting caution when extrapolating the study’s conclusions to other regions. Second, data analyzed in this study are based on reimbursement claims, meaning that any opioids prescribed outside of the national insurance market would not be recorded. This could result in an underrepresentation of actual opioid prescription quantities [53]. Third, the diagnostic codes assigned to patients may not always be completely accurate, which could impact the accuracy of the findings. Additionally, being a quasi-experimental study using ITS, there might be uncontrolled confounding factors that could influence the observed trends in opioid use.

It is worth noting that the NIMS has undergone improvements. For instance, the “Information Network to Prevent Doctor Shopping for Narcotics” feature, similar to PDMPs, included the entire narcotics data set since March 2021. This integration allows prescribers to assess a patient’s history of medical narcotic use, particularly when a doctor shopping for narcotics is suspected [20,21]. Feedback reports grounded in the NIMS big data are sent to doctors who demonstrate problematic prescribing behavior (eg, prescribing substantial amounts of medical narcotics). This series of changes is expected to gradually materialize in the form of opioid stewardship mentioned above in South Korea. Given these circumstances, there arises a clear need for future investigations to assess the efficacy of the enhanced NIMS.

Conclusion
This study suggests that, in its current form, the NIMS may not have brought significant improvements to the identified-negative impacts on opioid use. However, the potential for enhanced opioid stewardship through the NIMS highlights the need for continued efforts to address the misuse and abuse of opioids in South Korea.
indicators of opioid overuse and misuse. Additionally, the COVID-19 outbreak exhibited no significant influence on opioid use patterns. The absence of a real-time monitoring feature within the NIMS could be a key contributing factor. Further exploration and enhancements are needed to maximize the NIMS’ impact on curbing inappropriate opioid use.

Acknowledgments
We used the Health Insurance Review and Assessment Service–National Inpatient Sample from 2016 to 2020 (S20220126793) for this study; however, the findings are not associated with the Ministry of Health and Welfare or Health Insurance Review and Assessment Service.

Data Availability
The authors used the Health Insurance Review and Assessment Service–National Inpatient Sample data for this study and do not have permission to share these data. Raw data can be accessed with permission from the Health Insurance Review and Assessment.

Authors’ Contributions
IHL and NKJ conceived and designed the study. SYK and NKJ performed the analysis. SYK first drafted the manuscript. All authors participated in drafting the paper and approved the final version to be submitted for publication.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Disease and medication codes.

References


Kim SY, Cho NW, Yoo MS, Han SY, Oh JW. Narcotics information management system in South Korea: system development and innovation. BMC Health Serv Res 2023;23(1):73 [FREE Full text] [doi: 10.1186/s12913-023-09060-z] [Medline: 36694145]


Kim KY. Introduction to health insurance review and assessment services national patients sample. HIRA Policy Trends 2012;6(5):37-47.


Jena AB, Goldman D, Weaver L, Karaca-Mandic P. Opioid prescribing by multiple providers in Medicare: retrospective observational study of insurance claims. BMJ 2014;348:g1393 [FREE Full text] [doi: 10.1136/bmj.g1393] [Medline: 2455363]


Abbreviations

HIRA: Health Insurance Review and Assessment Service
HIRA-NPS: Health Insurance Review and Assessment Service–National Patient Sample
ICD-10: International Classification of Diseases, Tenth Revision
ITS: interrupted time series
MedAid: Medical Aid
MFDS: Ministry of Food and Drug Safety
MME: morphine milligram equivalents
NHI: National Health Insurance
NIMS: Narcotics Information Management System
NMS: narcotics monitoring system
PDMP: prescription drug monitoring program
PVI: Patriots and Veterans Insurance Plan

Edited by A Mavragani, T Sanchez; submitted 09.03.23; peer-reviewed by JR Medina, I Degen; comments to author 06.07.23; revised version received 09.09.23; accepted 07.01.24; published 21.02.24.

Please cite as:
Lee IH, Kim SY, Park S, Ryu JG, Je NK
Impact of the Narcotics Information Management System on Opioid Use Among Outpatients With Musculoskeletal and Connective Tissue Disorders: Quasi-Experimental Study Using Interrupted Time Series
JMIR Public Health Surveill 2024;10:e47130
URL: https://publichealth.jmir.org/2024/1/e47130
doi:10.2196/47130
PMID:38381481

©Iyn-Hyang Lee, So Young Kim, Susin Park, Jae Gon Ryu, Nam Kyung Je. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 21.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Use of Sentinel Surveillance Platforms for Monitoring SARS-CoV-2 Activity: Evidence From Analysis of Kenya Influenza Sentinel Surveillance Data

Daniel Owusu1*, BA, MPhil, DrPH; Linus K Ndegwa2*, MPHE, PhD; Jorim Ayugi3, BSc; Peter Kinuthia4, BSc, MSc; Rosalia Kalani5, BA, MS; Mary Okeyo6, HND; Nancy A Otieno3, MA, PhD; Gilbert Kikwai3, BSc; Bonventure Juma3, BS, MS, PhD; Peninah Munyu3a, BVMed, PhD; Francis Kuria7, MBChB, MMed; Emmanuel Okunga5, MBChB, MSc; Ann C Moen1, BA, BS, MPA; Gideon O Emukule2, PhD

1Influenza Division, US Centers for Disease Control and Prevention, Atlanta, GA, United States
2Global Influenza Branch, Influenza Division, US Centers for Disease Control and Prevention, Nairobi, Kenya
3Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya
4Henry Jackson Foundation, Nairobi, Kenya
5Disease Surveillance and Response Unit, Ministry of Health, Nairobi, Kenya
6National Influenza Centre Laboratory, National Public Health Laboratories, Ministry of Health, Nairobi, Kenya
7Directorate of Public Health, Ministry of Health, Nairobi, Kenya
*these authors contributed equally

Abstract

Background: Little is known about the cocirculation of influenza and SARS-CoV-2 viruses during the COVID-19 pandemic and the use of respiratory disease sentinel surveillance platforms for monitoring SARS-CoV-2 activity in sub-Saharan Africa.

Objective: We aimed to describe influenza and SARS-CoV-2 cocirculation in Kenya and how the SARS-CoV-2 data from influenza sentinel surveillance correlated with that of universal national surveillance.

Methods: From April 2020 to March 2022, we enrolled 7349 patients with severe acute respiratory illness or influenza-like illness at 8 sentinel influenza surveillance sites in Kenya and collected demographic, clinical, underlying medical condition, vaccination, and exposure information, as well as respiratory specimens, from them. Respiratory specimens were tested for influenza and SARS-CoV-2 by real-time reverse transcription polymerase chain reaction. The universal national-level SARS-CoV-2 data were also obtained from the Kenya Ministry of Health. The universal national-level SARS-CoV-2 data were collected from all health facilities nationally, border entry points, and contact tracing in Kenya. Epidemic curves and Pearson $r$ were used to describe the correlation between SARS-CoV-2 positivity in data from the 8 influenza sentinel sites in Kenya and that of the universal national SARS-CoV-2 surveillance data. A logistic regression model was used to assess the association between influenza and SARS-CoV-2 coinfection with severe clinical illness. We defined severe clinical illness as any of oxygen saturation <90%, in-hospital death, admission to intensive care unit or high dependence unit, mechanical ventilation, or a report of any danger sign (ie, inability to drink or eat, severe vomiting, grunting, stridor, or unconsciousness in children younger than 5 years) among patients with severe acute respiratory illness.

Results: Of the 7349 patients from the influenza sentinel surveillance sites, 76.3% (n=5606) were younger than 5 years. We detected any influenza (A or B) in 8.7% (629/7224), SARS-CoV-2 in 10.7% (768/7199), and coinfection in 0.9% (63/7165) of samples tested. Although the number of samples tested for SARS-CoV-2 from the sentinel surveillance was only 0.2% (60 per week vs 36,000 per week) of the number tested in the universal national surveillance, SARS-CoV-2 positivity in the sentinel
surveillance data significantly correlated with that of the universal national surveillance (Pearson $r=0.58$; $P<0.001$). The adjusted odds ratios (aOR) of clinical severe illness among participants with coinfection were similar to those of patients with influenza only (aOR 0.91, 95% CI 0.47-1.79) and SARS-CoV-2 only (aOR 0.92, 95% CI 0.47-1.82).

Conclusions: Influenza substantially cocirculated with SARS-CoV-2 in Kenya. We found a significant correlation of SARS-CoV-2 positivity in the data from 8 influenza sentinel surveillance sites with that of the universal national SARS-CoV-2 surveillance data. Our findings indicate that the influenza sentinel surveillance system can be used as a sustainable platform for monitoring respiratory pathogens of pandemic potential or public health importance.

(JMIR Public Health Surveill 2024;10:e50799) doi:10.2196/50799

KEYWORDS
SARS-CoV-2; COVID-19; influenza; sentinel surveillance; Kenya; epidemic; local outbreak; respiratory infection; surveillance; cocirculation; monitoring; respiratory pathogen

Introduction

Influenza can cause local outbreaks and seasonal epidemics and has pandemic potential, which makes it an important respiratory infection to monitor for prompt public health response [1]. Influenza occurs in multiple epidemics all year round in the tropics and mostly in winter or cold seasons in temperate regions [1,2]. Although influenza virus infection can be asymptomatic in about 30%-50% of infections or cause mild illness [3-5], in some people, influenza virus infection can cause severe illness leading to hospitalization, intensive care admission, mechanical ventilation, or death [1,6]. It is estimated that influenza is associated with 290,000 to 650,000 annual deaths globally [7,8]. Sub-Saharan Africa is one of the highest-burden regions [7,8], but influenza remained poorly monitored until recently when countries began building influenza surveillance systems to monitor circulating strains [2] and contribute to global data to inform strain selections for annual influenza vaccines [9,10].

In December 2019, an outbreak of severe respiratory illness caused by a novel coronavirus (SARS-CoV-2) was reported in Wuhan, China [11]. The World Health Organization declared SARS-CoV-2 infection that causes COVID-19 as a pandemic on March 11, 2020. Given that this outbreak occurred during the Northern Hemisphere influenza season, the public health community was concerned about the potential cocirculation of influenza and SARS-CoV-2 that could overburden health care systems. During the influenza months in winter, surges in medical visits and increase in health care use have been well documented [12]. Data from the Northern Hemisphere or countries with strong public health surveillance systems showed that influenza activity unexpectedly remained minimal during the COVID-19 pandemic [13-19]. Similarly, countries in the sub-Saharan African region, which has some of the highest burden of influenza, reported very minimal influenza activity [20]. However, given that most of these countries are still developing their influenza surveillance systems, it is unclear whether the minimal activity reported was due to a break in surveillance because of the response measures to the pandemic. In addition, little is known about how data from sentinel influenza surveillance platforms could be used to gauge SARS-CoV-2 circulation in sub-Saharan Africa.

Kenya, an equatorial country in sub-Saharan Africa, has implemented sentinel influenza surveillance since 2007 and developed a strong platform that continued to function during the COVID-19 pandemic. This provides an opportunity to understand influenza circulation in a tropical setting during the COVID-19 pandemic. Here, we describe influenza circulation and coinfection with SARS-CoV-2 in Kenya during the COVID-19 pandemic and compare SARS-CoV-2 virus detection from the influenza sentinel surveillance system with that of the universal national SARS-CoV-2 surveillance that included data from all facilities and border points in Kenya. The findings of this evaluation could inform the implementation or strengthening of national sentinel influenza systems and leveraging these platforms to monitor respiratory pathogens of pandemic potential.

Methods

Study Sites

The Kenya Ministry of Health (MoH), in collaboration with the US Centers for Disease Control and Prevention (CDC), established sentinel surveillance for influenza in selected health care facilities in 2007 [21]. Currently, there are 8 sentinel surveillance sites (Figure 1) located within 6 county-level referral hospitals (Coast General Teaching and Referral Hospital, Kakamega, Nakuru, Nyeri, Marsabit, and Siaya [22]), 1 tertiary referral hospital (Kenyatta National Hospital [23]), and the International Rescue Committee hospital in the Kakuma Refugee Camp. All sites enroll patients of all ages except the site in the Kenyatta National Hospital, which enrolls only pediatric cases.

https://publichealth.jmir.org/2024/1/e50799

XSL-FO
RenderX
Surveillance Procedures

Trained surveillance officers (either a nurse or a clinical officer) identified and enrolled patients who met the severe acute respiratory illness (SARI) or influenza-like illness (ILI) case definitions [24]. We defined a SARI case as an acute onset of respiratory illness (within the last 10 days) with fever (reported or measured fever of $\geq 38^\circ C$) and cough, requiring hospitalization. Eligible patients with SARI were identified and enrolled 5 days a week (Monday to Friday) within 48 hours of hospital admission for all sites except for Kakuma Refugee Camp, where enrollment was done Monday to Saturday. In addition, surveillance staff reviewed admission logbooks every Monday to identify and enroll any eligible patients with SARI admitted during the weekends. We defined ILI as an acute onset of respiratory illness (within the last 10 days) with measured fever of $\geq 38^\circ C$ and cough. The first 3 patients with ILI were enrolled per week at the outpatient clinics. Surveillance for SARS-CoV-2 was initiated and integrated into the routine surveillance for influenza among patients presenting with SARI or ILI (as defined earlier) in the sentinel sites in April 2020.

This analysis used data collected from the sentinel sites (Figure 1) and the national SARS-CoV-2 surveillance data reported to the MoH from April 2020 through March 2022.

Specimen Collection and Testing

Trained surveillance officers collected nasopharyngeal or nasal and oropharyngeal swabs from the enrolled patients. Nasopharyngeal or nasal and oropharyngeal swabs were placed in a single cryovial containing 3-mL sterile viral transport media and labeled with the patient’s unique identifier, a barcode number, and the date of specimen collection. Samples were stored at 2-8 $^\circ C$ at the sentinel surveillance sites for up to 36 hours and then shipped to the National Influenza Center laboratory in Nairobi, where each specimen was made into 3 aliquots and tested.

We tested an aliquot of each specimen for SARS-CoV-2 and influenza-viral ribonucleic acid at the CDC-supported Kenya Medical Research Institute laboratory and National Influenza Center laboratory, respectively, by real-time reverse transcription polymerase chain reaction (rtRT-PCR). Each assay had a target for Homo sapiens (human) RNase P gene for
monitoring specimen quality. Influenza testing was done as previously described [21,22]. The primer and probe sequences for the influenza A, influenza B, and ribonuclease P gene targets were as per the approved CDC’s Human Influenza Virus Real-Time RT-PCR Detection and Characterization Panel [21,22]. All specimens positive for influenza type A were subtyped for H3, H5, and A(H1N1) pdm09 using rtRT-PCR. The SARS-CoV-2 laboratory testing protocol has been described elsewhere [25]. A cycle threshold (Ct) of ≥40.0 was considered positive and ≥40.0 or no amplification was considered negative.

Data Collection and Variable Definitions

The surveillance staff enrolled patients using an electronic structured questionnaire with questions on demographics, clinical presentation, underlying medical conditions, medical history, vaccination history, and exposures. The questionnaire was administered to participants or their parents or legal guardians using a password-protected netbook. In addition, surveillance staff conducted a medical chart review to complete a follow-up exit questionnaire for patients with SARI. This review was done to abstract data on clinical outcomes of hospitalization. Laboratory data were recorded into Freezerworks software (Dataworks Development, Inc) and merged with epidemiological information. All data were stored in password-protected Microsoft Access or SQL databases on a central server at the Kenya Medical Research Institute, Center for Global Health Research campus in Nairobi.

We defined influenza virus or SARS-CoV-2 detection as a positive rtRT-PCR test of influenza (A or B) or SARS-CoV-2, respectively. A participant was assumed to have been coinfected if the respiratory sample was rtRT-PCR positive for both influenza (A or B) and SARS-CoV-2. Severe clinical illness was defined as any of oxygen saturation <90%, in-hospital death, admission to intensive care unit, mechanical ventilation, or a report of any danger sign (ie, inability to drink or eat, severe vomiting, grunting, stridor, or unconsciousness among children younger than 5 years).

Kenya National SARS-CoV-2 Surveillance

In Kenya, the universal national SARS-CoV-2 surveillance was led by MoH focusing initially on points of entry and contact of people exposed to individuals who tested positive for SARS-CoV-2. Once there was evidence of community spread of SARS-CoV-2 infections in Kenya, all persons suspected to have COVID-19 or presenting with ILI in health facilities were tested for COVID-19 [26]. In addition, cargo vessel crew at all entry points were screened for SARS-CoV-2. Testing was initially limited to a few laboratories but later expanded to 96 public and private laboratories conducting rtRT-PCR tests and 472 conducting rapid diagnostic tests. All 8 influenza sentinel sites contributed data to the universal national SARS-CoV-2 surveillance throughout the study period.

Statistical Analysis

Frequencies and percentages were calculated to describe the characteristics of participants overall and by patient type (in- or outpatients). Microsoft Excel (Microsoft Corp) charts were used to describe influenza or SARS-CoV-2 trends based on a calculated monthly percentage of influenza and SARS-CoV-2 positivity. Pearson correlation coefficient (r) was used to compare trends in the 3-week moving average of SARS-CoV-2 positivity from the influenza sentinel surveillance data with that of the data from the universal national SARS-CoV-2 surveillance.

Logistic regression was conducted to assess the association between coinfection and severe clinical illness among SARI cases. Univariate and multivariable logistic regression analyses were also conducted to examine factors associated with the detection of influenza (A or B) only, SARS-CoV-2 only, and influenza and SARS-CoV-2 coinfection among enrolled patients, stratified by ILI versus SARI cases. In all multivariable models, we included age, surveillance site, and in-patient or outpatient status a priori and any other variable of interest (underlying conditions, sex, smoker in household, hospitalization in past year, symptoms) that was significant at P<0.05 in the univariate analysis. All analyses were performed using Stata Statistical Software (version 13.0; StataCorp).

Ethical Considerations

The Kenya MoH determined the sentinel influenza surveillance to be a routine public health activity, and a nonresearch project determination was also received from CDC (project 0900f3eb81e74404). Thus, ethical review was not required to enroll and collect data from patients at the sentinel sites. We obtained informed verbal consent from participants or parents or legal guardians of children before the administration of questionnaires and collection of specimens. In addition to verbal consent of parents or guardians, children aged 7 years and older provided assent before participation.

Results

Demographic and Clinical Characteristics

Patient characteristics are summarized in Table 1. From April 2020 through March 2022, we enrolled 7349 patients from the sentinel influenza surveillance sites in Kenya. Most (n=5606, 76.3%) of the patients were children aged younger than 5 years. The median age of all the enrolled patients was 1.5 (IQR 0.7-4.6) years. Of all the enrolled patients, 3331 (45.3%) were female participants, and 899 (12.2%) had at least one underlying medical condition (eg, HIV positive: n=171, 2.3%; heart disease: n=281, 3.8%; chronic neurological or neuromuscular disease: n=185, 2.5%; asthma: n=223, 3%; diabetes: n=81, 1.1%; and other: n=93, 1.3%). Overall, 5932 (80.7%) were patients with SARI, while the rest were patients with ILI who were seen as outpatients. The Kakuma site provided the highest percentage of enrolled patients (n=2316, 31.5%), followed by Siaya County Referral Hospital with 18.1% (n=1332). Kakamega County Referral Hospital had the least number of enrolled patients (n=174, 2.4%). Overall, 46% (n=2727) of inpatients (SARI cases) presented with severe clinical illness. Among patients presenting with severe clinical illness, 20.9% (n=1238) were admitted to the intensive care unit, 8.9% (n=527) had oxygen saturation <90%, 2.3% (n=134) were put on mechanical ventilation during admission, 41% (2084/5080) were children younger than 5 years who had at least one danger sign reported, and 5.4% (n=319) died during admission (Table 1).

https://publichealth.jmir.org/2024/1/e50799
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Outpatients (n=1417)</th>
<th>Inpatients (n=5932)</th>
<th>All (N=7349)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), median (IQR)</strong></td>
<td>14.0 (2.4-32.9)</td>
<td>1.2 (0.6-2.7)</td>
<td>1.5 (0.7-4.6)</td>
</tr>
<tr>
<td><strong>Age groups, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-11 months</td>
<td>159 (11.2)</td>
<td>2627 (44.3)</td>
<td>2786 (37.9)</td>
</tr>
<tr>
<td>12-23 months</td>
<td>135 (9.5)</td>
<td>1312 (22.1)</td>
<td>1447 (19.7)</td>
</tr>
<tr>
<td>2-4 years</td>
<td>232 (16.4)</td>
<td>1141 (19.2)</td>
<td>1373 (18.7)</td>
</tr>
<tr>
<td>5-14 years</td>
<td>198 (14)</td>
<td>348 (5.9)</td>
<td>546 (7.4)</td>
</tr>
<tr>
<td>15-64 years</td>
<td>658 (46.4)</td>
<td>365 (6.2)</td>
<td>1023 (13.9)</td>
</tr>
<tr>
<td>≥65 years</td>
<td>35 (2.5)</td>
<td>139 (2.3)</td>
<td>174 (2.4)</td>
</tr>
<tr>
<td>&lt;5 years</td>
<td>526 (37.1)</td>
<td>5080 (85.6)</td>
<td>5606 (76.3)</td>
</tr>
<tr>
<td>≥5 years</td>
<td>891 (62.9)</td>
<td>852 (14.4)</td>
<td>1743 (23.7)</td>
</tr>
<tr>
<td>&lt;13 years</td>
<td>17 (0.7)</td>
<td>677 (47.8)</td>
<td>6095 (82.9)</td>
</tr>
<tr>
<td>≥13 years</td>
<td>514 (8.7)</td>
<td>740 (25.2)</td>
<td>1254 (17.1)</td>
</tr>
<tr>
<td><strong>Female, n (%)</strong></td>
<td>711 (50.2)</td>
<td>2620 (44.2)</td>
<td>3331 (45.3)</td>
</tr>
<tr>
<td><strong>Site name, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenyatta National Hospital</td>
<td>5 (0.4)</td>
<td>1053 (31.1)</td>
<td>1058 (14.4)</td>
</tr>
<tr>
<td>Coast General Teaching and Referral Hospital</td>
<td>11 (0.8)</td>
<td>446 (7.5)</td>
<td>457 (6.2)</td>
</tr>
<tr>
<td>Nyeri County Referral Hospital</td>
<td>291 (20.5)</td>
<td>506 (8.5)</td>
<td>797 (10.9)</td>
</tr>
<tr>
<td>Nakuru County Referral Hospital</td>
<td>26 (1.8)</td>
<td>916 (15.4)</td>
<td>942 (12.8)</td>
</tr>
<tr>
<td>Kakamega County Referral Hospital</td>
<td>14 (1)</td>
<td>160 (2.7)</td>
<td>174 (2.4)</td>
</tr>
<tr>
<td>Siaya County Referral Hospital</td>
<td>512 (36.1)</td>
<td>820 (13.8)</td>
<td>1332 (18.1)</td>
</tr>
<tr>
<td>Marsabit Community Referral Hospital</td>
<td>85 (6)</td>
<td>188 (3.2)</td>
<td>273 (3.7)</td>
</tr>
<tr>
<td>Kakuma Refugee Camp</td>
<td>473 (33.4)</td>
<td>1843 (31.1)</td>
<td>2316 (31.5)</td>
</tr>
<tr>
<td><strong>Underlying medical condition, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV infection</td>
<td>47 (3.3)</td>
<td>124 (2.1)</td>
<td>171 (2.3)</td>
</tr>
<tr>
<td>HIV unknown</td>
<td>467 (33)</td>
<td>837 (14.1)</td>
<td>1304 (17.7)</td>
</tr>
<tr>
<td>Heart disease</td>
<td>61 (4.3)</td>
<td>220 (3.7)</td>
<td>281 (3.8)</td>
</tr>
<tr>
<td>Chronic neurological or neuromuscular disease</td>
<td>14 (1)</td>
<td>171 (2.9)</td>
<td>185 (2.5)</td>
</tr>
<tr>
<td>Asthma</td>
<td>64 (4.5)</td>
<td>159 (2.7)</td>
<td>223 (3)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>18 (1.3)</td>
<td>63 (1.1)</td>
<td>81 (1.1)</td>
</tr>
<tr>
<td>Othera</td>
<td>18 (1.3)</td>
<td>75 (1.3)</td>
<td>93 (1.3)</td>
</tr>
<tr>
<td>Hospitalized in the past 12 months, n (%)</td>
<td>73 (5.2)</td>
<td>810 (13.7)</td>
<td>883 (12)</td>
</tr>
<tr>
<td>Current smoker in the household, n (%)</td>
<td>39 (2.8)</td>
<td>258 (4.4)</td>
<td>297 (4)</td>
</tr>
<tr>
<td><strong>SARS-CoV-2, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tested</td>
<td>1400 (98.8)</td>
<td>5799 (97.8)</td>
<td>7199 (98)</td>
</tr>
<tr>
<td>Positive</td>
<td>235 (16.8)</td>
<td>533 (9.2)</td>
<td>768 (10.7)</td>
</tr>
<tr>
<td><strong>Influenza, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tested</td>
<td>1402 (98.9)</td>
<td>5822 (98.2)</td>
<td>7224 (98.3)</td>
</tr>
<tr>
<td>Influenza positive</td>
<td>103 (7.4)</td>
<td>526 (9)</td>
<td>629 (8.7)</td>
</tr>
<tr>
<td>Influenza A positive</td>
<td>47 (45.6)</td>
<td>259 (49.2)</td>
<td>306 (48.7)</td>
</tr>
<tr>
<td>Influenza A (H3N2) positive</td>
<td>39 (83)</td>
<td>238 (92)</td>
<td>277 (90.5)</td>
</tr>
<tr>
<td>Influenza A (H1N1) pdm09 positive</td>
<td>0 (0)</td>
<td>3 (1.2)</td>
<td>3 (1)</td>
</tr>
</tbody>
</table>
Influenza and SARS-CoV-2 Circulation

We tested 7199 (98%) specimens for SARS-CoV-2, 7224 (98.3%) specimens for influenza, and 7165 (97.5%) of the specimens for both influenza and SARS-CoV-2. We detected influenza (A or B) in 8.7% (n=629); 7.4% (n=103) tested positive among outpatients and 9% (n=526) among inpatients. Of the patients with influenza, 48.7% (n=306) had influenza A, while 53.6% (n=337) had influenza B. Among the influenza A cases, 90.5% (n=277) were A (H3N2), 1% (n=3) A (H1N1) pdm09, and 8.5% (n=26) not subtyped. Among patients who were tested for SARS-CoV-2, 10.7% (n=768) tested positive; 16.8% (235/1400) tested positive among outpatients and 9.2% (533/5799) among inpatients (Table 1).

Detection of both viruses continued throughout the evaluation period, with influenza positivity remaining higher than SARS-CoV-2 positivity from mid-May 2021 through late December 2021. The highest influenza positivity was 21.4% (71/333) in August 2021 when SARS-CoV-2 positivity was 17.2% (57/333). The highest peak of SARS-CoV-2 positivity was in March 2021 at 25.1% (109/434) when influenza positivity was 9.4% (41/434; Figure 2).
Figure 2. Monthly trends of influenza activity (including circulating subtypes), SARS-CoV-2 activity, and coinfections of influenza and SARS-CoV-2 using data from the 8 influenza sentinel surveillance sites in Kenya, April 2020 to March 2022. From April 2020 through March 2022, influenza cocirculated with SARS-CoV-2, with influenza positivity rates sometimes surpassing that of SARS-CoV-2. Data source: Kenya influenza sentinel surveillance system.

SARS-CoV-2 Positivity in the Influenza Sentinel Surveillance Versus the Universal National SARS-CoV-2 Surveillance Data

SARS-CoV-2 positivity in the data collected from the sentinel sites, with a median of 60 (IQR 38-71) weekly tests during the study period, significantly correlated (Pearson r=0.58; P<.001) with SARS-CoV-2 positivity in the Kenya universal national SARS-CoV-2 surveillance data, which had a median of 36,000 (IQR 28,000-43,000) weekly tests (Figure 3).

Figure 3. Three-week MA of SARS-CoV-2 positivity using data from the 8 influenza sentinel surveillance sites in Kenya versus universal national SARS-CoV-2 surveillance data from all tests conducted in Kenya, April 2020 to March 2022. This figure shows number of specimens tested from the sentinel sites and 3-week MA of SARS-CoV-3 positivity from the sentinel sites and the Kenya national SARS-CoV-2 surveillance data. SARS-CoV-2 detection and peak from the sentinel surveillance, with a median of 60 (IQR 38-71) weekly tests, significantly correlated (Pearson r=0.58; P<.001) with the detection and peak activity from the Kenya universal national SARS-CoV-2 surveillance data, which had a median of 36,000 (IQR 28,000-43,000) weekly tests. Epi: epidemiology; MA: moving average; SARI: severe acute respiratory infection.

Coinfection and Severe Clinical Illness

We detected influenza and SARS-CoV-2 coinfection in 0.9% (63/7165) of those who were tested for both infections. Thus, 8.2% (63/765) of those who tested positive for SARS-CoV-2 also tested positive for influenza. The odds ratios of clinical severe illness among participants with coinfection were similar to those of patients with influenza only (adjusted odds ratio [aOR] 1.06, 95% CI 0.72-1.39) and SARS-CoV-2 only (aOR 1.00, 95% CI 0.54-2.10) infection (Table 2).
Table 2. Association between influenza and SARS-CoV-2 coinfection and severe clinical illness among patients hospitalized with severe acute respiratory illness (SARI) from 8 influenza sentinel surveillance sites in Kenya (n=1004), April 2020 to March 2022.\(^a\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Had severe illness,(^b) n (%)</th>
<th>uOR(^c) (95% CI)</th>
<th>aOR(^d) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Detection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARS-CoV-2 and influenza A or B coinfection</td>
<td>20 (4.9)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Influenza A or B</td>
<td>190 (46.2)</td>
<td>1.23 (0.69-2.21)</td>
<td>0.91 (0.47-1.79)</td>
</tr>
<tr>
<td>SARS-CoV-2</td>
<td>201 (48.9)</td>
<td>1.15 (0.64-2.05)</td>
<td>0.92 (0.47-1.82)</td>
</tr>
<tr>
<td><strong>Age group</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-11 months</td>
<td>188 (45.7)</td>
<td>1.35 (0.91-1.98)</td>
<td>1.07 (0.61-1.85)</td>
</tr>
<tr>
<td>12-23 months</td>
<td>84 (20.4)</td>
<td>0.82 (0.53-1.25)</td>
<td>0.67 (0.37-1.20)</td>
</tr>
<tr>
<td>2-4 years</td>
<td>66 (16.1)</td>
<td>0.71 (0.46-1.11)</td>
<td>0.57 (0.31-1.04)</td>
</tr>
<tr>
<td>5-12 years</td>
<td>13 (3.2)</td>
<td>0.56 (0.27-1.15)</td>
<td>0.25 (0.10-0.62)</td>
</tr>
<tr>
<td>≥13 years</td>
<td>60 (14.6)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td><strong>Any underlying medical condition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>332 (80.8)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>79 (19.2)</td>
<td>1.75 (1.23-2.48)</td>
<td>N/A(^e)</td>
</tr>
<tr>
<td><strong>Heart disease</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>42 (10.2)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>369 (89.8)</td>
<td>2.48 (1.50-4.12)</td>
<td>1.44 (0.76-2.73)</td>
</tr>
<tr>
<td><strong>Diabetes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>9 (2.2)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>402 (97.8)</td>
<td>1.08 (0.45-2.60)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Gastroenteritis or diarrhea</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>112 (27.3)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>299 (72.8)</td>
<td>1.74 (1.29-2.36)</td>
<td>1.22 (0.84-1.77)</td>
</tr>
<tr>
<td><strong>Duration of SARI symptoms at presentation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3 days</td>
<td>379 (92.2)</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>≥4 days</td>
<td>32 (7.8)</td>
<td>0.92 (0.58-1.46)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\)Odds ratios were adjusted for age, site of data collection, and any variable that was significant at \(P<.05\) in the univariate analysis.

\(^b\)Severe clinical illness was defined as any of oxygen saturation <90%, in-hospital death, admission to intensive care unit or high dependence unit, mechanical ventilation, or a report of any danger sign among children younger than 5 years (inability to drink or eat, severe vomiting, grunting, stridor, or unconsciousness).

\(^c\)uOR: unadjusted odds ratio.

\(^d\)aOR: adjusted odds ratio.

\(^e\)N/A: Not applicable.

**Factors Associated With Influenza and SARS-CoV-2 Infection Detection Among SARI and ILI Cases**

None of the variables included in the multivariable analysis were associated with influenza-only detection or influenza and SARS-CoV-2 coinfection among SARI cases. However, detection of SARS-CoV-2 only was significantly higher in patients aged 13 years and older compared to those aged 0-11 months (aOR 3.85, 95% CI 1.86-8.02; Multimedia Appendix 1). Adjusting for other covariates, influenza-only detection was less common in patients with ILI diagnosed with malaria (aOR 0.09, 95% CI 0.03-0.27). Among ILI cases, SARS-CoV-2–only detection was associated with age group, malaria diagnosis, and presenting with chills (Multimedia Appendix 2).

**Discussion**

**Summary of Main Findings**

In this study, we found that influenza continued to circulate in Kenya throughout the COVID-19 pandemic. Although the number of weekly tests in the influenza surveillance system was only 0.2% (60 per week vs 36,000 per week) of the number of tests conducted in the universal national surveillance for SARS-CoV-2, we found that peaks in SARS-CoV-2 positivity from our sentinel surveillance sites coincided with peaks in...
SARS-CoV-2 positivity from the universal national SARS-CoV-2 surveillance data. Our analysis further revealed that about 1 in 12 patients who tested positive for SARS-CoV-2 also tested positive for influenza; however, we did not find any significant evidence of increased severe clinical illness among these patients compared to those who tested positive for influenza only or SARS-CoV-2 only.

**Influenza and SARS-CoV-2 Cocirculation**

Multiple reports from elsewhere [13,14,17,19,27] indicated that influenza activity was minimal and far below expectation at the height of the SARS-CoV-2 pandemic from 2020 through 2021. Our results suggest that unlike these countries, Kenya continued to record substantially high levels of influenza activity during the pandemic. Some experts have attributed the minimal influenza activity observed in many countries during the SARS-CoV-2 pandemic to the impacts of widespread mitigation measures such as social distancing, facial covering, and stay-at-home measures [15,16,18,28]. However, Kenya implemented mitigation measures with similar stringency to those of countries that recorded minimal influenza activity such as South Africa and the United States (Multimedia Appendix 3) [29]. Although our data are limited to explain why Kenya experienced high influenza activity despite implementing some of the stringent mitigation measures, our results suggest the importance of local surveillance of influenza and other pathogens of public health importance and pandemic potential. Because of geographic and sociocultural differences, data from one country may not be applicable to another country.

**Influenza Sentinel Surveillance System Versus Universal National SARS-CoV-2 Surveillance**

We observed that the detection rates and peak activity of SARS-CoV-2 from the sentinel surveillance were well aligned with that of the universal national surveillance, suggesting that the data from the sentinel system was able to detect increased SARS-CoV-2 activity during the study period. This is consistent with findings reported from Indonesia, in which SARS-CoV-2 trends in the influenza surveillance data aligned with COVID-19 national surveillance data [30]. When the pandemic started, the Kenyan government, like many other countries, implemented several ad hoc systems to respond to the pandemic, including opening border surveillance units and national influenza surveillance coordination centers and increasing laboratory capacity to collect SARS-CoV-2 data and implement interventions to control the pandemic. Although it might be good to identify all cases to respond appropriately, a universal national surveillance system is labor- and resource-intensive and not sustainable in the long term in Kenya and other settings. Therefore, there is a need for a sustainable approach for ongoing monitoring of SARS-CoV-2 and other pathogens of pandemic potential in general to detect early signals and implement effective interventions. Our results support the World Health Organization’s [31] recommendation for integrated surveillance of SARS-CoV-2 and influenza and suggest that influenza sentinel surveillance systems can serve this purpose for routine monitoring and provide signals of increased activity of SARS-CoV-2 to aid the development of appropriate public health interventions and decisions. These findings provide further evidence about the need to implement and sustain a strong influenza surveillance system in all countries, especially, those in sub-Saharan Africa.

**Influenza and SARS-CoV-2 Coinfection and Severe Clinical Illness**

Besides concerns about health systems being overwhelmed by influenza and SARS-CoV-2 cocirculation, one major concern about this cocirculation was the potential for coinfection with both viruses and consequent severe outcomes in patients. Our understanding of the extent and potential impact of influenza and SARS-CoV-2 coinfection on clinical illness is still evolving. A few studies have reported varying results regarding the prevalence and severity of coinfection [32-36]. Prevalence of influenza coinfection in people with SARS-CoV-2 as low as 0.6% has been reported in the United States [35] and as high as 12.6% has been reported in Guangzhou, China [33]. In this study, we found 8.2% (63/768) influenza coinfection in participants who tested positive for SARS-CoV-2. Unlike previous findings from animal models [32] and UK adults [36], but consistent with findings from India [34], we did not find a significant increase in the frequency of severe clinical presentation among these individuals compared to those infected with only influenza or only SARS-CoV-2. Our results may have been impacted by the age composition of the study sample. The majority were children, who are already at increased risk of severe influenza outcomes [37] and often have less severe COVID-19 [38]. Our findings together with the existing literature on influenza and SARS-CoV-2 coinfection suggest a need to continue to monitor cases of coinfection and their clinical presentations and outcomes to inform preventive measures and clinical management.

**Limitations**

Although we used data from a good influenza sentinel surveillance system that collects data from all ages and our analysis controlled for multiple potential confounders, there are some limitations of the study. The criteria used to define severe clinical illness were in-hospital presentations and outcomes and did not include outcomes outside the sentinel site hospitals such as death prior to admission or after discharge. Therefore, more studies are needed to examine disease severity among people with influenza and SARS-CoV-2 coinfection. The age of 65 years and older is a risk factor for severe outcomes of SARS-CoV-2 infection; however, this population is underrepresented in this analysis (139/5932, 2.3%), so we cannot draw inferences on coinfections for this population. Most of the participants were children younger than 13 years of age; therefore, we could not assess differences among older patients. Data used in this study may not be representative of the Kenyan population because the sites were within only selected counties that do not necessarily represent the rest of the country.

**Conclusions**

Unlike many other countries, influenza significantly cocirculated with SARS-CoV-2 during the early weeks of 2020 through the first quarter of 2022 in Kenya, and SARS-CoV-2 positivity data from the influenza sentinel surveillance sites were comparable to that of the national surveillance system. Our findings reinforce
the need to implement or strengthen national influenza sentinel surveillance systems to monitor respiratory diseases of public health importance or those with pandemic potential. Influenza or respiratory disease sentinel surveillance sites could be used as a sustainable platform for national monitoring of SARS-CoV-2 activity.

Acknowledgments
The authors thank the patients who enrolled in the surveillance and who provided data for their generosity. The authors thank the staff of the following hospitals for their invaluable contributions to this effort: Kenyatta National Hospital; Coast General Teaching and Referral Hospital; Kakamega, Nakuru, Nyeri, Marsabit, and Siaya County Referral Hospitals; and the International Rescue Committee hospital in the Kakuma Refugee Camp; and the hospital managers for their support of this research. The authors would also like to thank the surveillance officers for their great work in identifying cases and collecting data and samples: Molly Mugabe, Esther Chomba, Jonathan Onyango, Milka Bunei, Juma Maleve, Musa Nawe, Elizabeth Simiyu, and Peter Muturi. The authors appreciate the data staff for the enormous work of entering and cleaning the data at the National Influenza Center and Centers for Disease Control and Prevention–supported Kenya Medical Research Institute laboratory. This study was conducted as part of the authors’ usual employment. No author received outside support or funding to conduct this study. The Kenya influenza surveillance system was supported by funds from the US Centers for Disease Control and Prevention.

Data Availability
The data sets generated and analyzed during this study are not publicly available because they were collected for public health surveillance by the Kenya Ministry of Health but are available from RK (focal person for influenza surveillance, division of disease surveillance and response, Kenya Ministry of Health) on reasonable request.

Disclaimer
The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Factors associated with influenza and SARS-CoV-2 infection detection among patients hospitalized with severe acute respiratory illness from 8 influenza sentinel surveillance sites in Kenya (n=1004), April 2020 to March 2022.

Multimedia Appendix 2
Factors associated with influenza and SARS-CoV-2 infection detection among outpatients with influenza-like illness from 8 influenza sentinel surveillance sites in Kenya (n=1004), April 2020 to March 2022.

Multimedia Appendix 3
Kenya implemented SARS-CoV-2 mitigation measures with similar stringency to those of countries that recorded minimal influenza activity such as South Africa and the United States.

References


https://publichealth.jmir.org/2024/1/e50799 JMIR Public Health Surveill 2024 | vol. 10 | e50799 | p.2226 (page number not for citation purposes)


**Abbreviations**
- **aOR**: adjusted odds ratio
- **CDC**: Centers for Disease Control and Prevention
- **ILI**: influenza-like illness
- **MoH**: Ministry of Health
- **rtRT-PCR**: real-time reverse transcription polymerase chain reaction
- **SARI**: severe acute respiratory illness
Development of New Stringency Indices for Nonpharmacological Social Distancing Policies Implemented in Korea During the COVID-19 Pandemic: Random Forest Approach

Catherine Apio\textsuperscript{1,}, BSc, MSc; Kyulhee Han\textsuperscript{1,}, BSc; Doeun Lee\textsuperscript{1}, BSc, MSc; Bogyeom Lee\textsuperscript{2}, BSc; Taesung Park\textsuperscript{3}, Prof Dr, PhD

\textsuperscript{1}Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea
\textsuperscript{2}Ross School of Business, University of Michigan-Ann Arbor, Ann Arbor, MI, United States
\textsuperscript{3}Department of Statistics, Seoul National University, Seoul, Republic of Korea

*these authors contributed equally

Corresponding Author:
Taesung Park, Prof Dr, PhD
Department of Statistics
Seoul National University
Building 25, Room 310
Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 010 5557 8924
Fax: 82 2 883 6144
Email: tspark@stats.snu.ac.kr

Abstract

Background: In the absence of an effective treatment method or vaccine, the outbreak of the COVID-19 pandemic elicited a wide range of unprecedented restriction policies aimed at mitigating and suppressing the spread of the SARS-CoV-2 virus. These policies and their Stringency Index (SI) of more than 160 countries were systematically recorded in the Oxford COVID-19 Government Response Tracker (OxCGRT) data set. The SI is a summary measure of the overall strictness of these policies. However, the OxCGRT SI may not fully reflect the stringency levels of the restriction policies implemented in Korea. Korea implemented 33 COVID-19 restriction policies targeting 4 areas: public facilities, public events, social gatherings, and religious gatherings.

Objective: This study aims to develop new Korea Stringency Indices (KSI) that reflect the stringency levels of Korea’s restriction policies better and to determine which government-implemented policies were most effective in managing the COVID-19 pandemic in Korea.

Methods: The random forest method was used to calculate the new KSI using feature importance values and determine their effectiveness in managing daily COVID-19 confirmed cases. Five analysis periods were considered, including November 01, 2020, to January 20, 2021 (Period 1), January 20, 2021, to June 27, 2021 (Period 2), November 01, 2020, to June 27, 2021 (Period 3), June 27, 2021, to November 01, 2021 (Period 4), and November 01, 2021, to April 24, 2022 (Period 5).

Results: Among the KSI, public facilities in period 4, public events in period 2, religious gatherings in periods 1 and 3, and social gatherings in period 5 had the highest importance. Among the public facilities, policies associated with operation hour restrictions in cinemas, restaurants, PC rooms, indoor sports facilities, karaoke, coffee shops, night entertainment facilities, and baths or saunas had the highest importance across all analysis periods. Strong positive correlations were observed between daily confirmed cases and public facilities, religious gatherings, and public events in period 1 of the pandemic. From then, weaker and negative correlations were observed in the remaining analysis periods. The comparison with the OxCGRT SI showed that the SI had a relatively lower feature importance and correlation with daily confirmed cases than the proposed KSI, making KSI more effective than SI.

Conclusions: Restriction policies targeting public facilities were the most effective among the policies analyzed. In addition, different periods call for the enforcement of different policies given their effectiveness varies during the pandemic.
Introduction

The spread of COVID-19 [1-3], which became a global pandemic [4], elicited a wide range of unprecedented restriction policies from different governments aimed to mitigate and suppress the spread of the SARS-CoV-2 virus [5-7]. Without a vaccine or an effective treatment, these social distancing policies included school closures, travel restrictions, bans on social and private gatherings, stay-at-home orders, workplace closures, closure of public transportation, and so forth [6]. The transmission of the SARS-CoV-2 virus spreads through contacts made between susceptible and infectious individuals depending on the spatial distance between contacts. Therefore, the suppression of contacts is the goal of these policies, designed to slow the growth rate of infections, as inferred from past epidemics [8,9].

The impact of implemented government policies in lowering COVID-19 cases has been demonstrated in many studies as the pandemic progressed [6,10-15]. One country that implemented social distancing policies was South Korea. When the first patient with COVID-19 was confirmed in Korea on January 20, 2020 [16-18], voluntary social distancing started to be practiced by citizens in their daily lives until it was declared a global pandemic. With the declaration of COVID-19 as a worldwide pandemic on March 11, 2020 [4], the Korean government instituted its first social distancing for 2 weeks starting March 22, 2020, when the number of cases rose to approximately 100-150 daily cases [19]. The government combined testing, contact tracing, early isolation, and the free treatment of positive cases together with digital technologies without considering “lockdown” measures [20-22]. Many countries and governments worldwide applauded South Korea’s response as the most influential and was considered as one of the most effective models against COVD-19 in the early months that followed the pandemic [20]. The COVID-19 curve was flattened to an average of 6.4 daily cases in the first week of May 2020 [20-22]. However, as the pandemic progressed, this approach could not hold up due to worsening COVID-19 situations in the country, mainly in the metropolitan areas tied to small church gatherings, restaurants, social gatherings, nightclubs, and schools [17,18,23,24]. From then on, the government imposed restrictions in these areas and would strengthen or lower the restriction levels depending on the average daily confirmed cases. The policies and their levels would be announced via official government websites and press releases [25].

The policies implemented by different countries worldwide were systematically collected and recorded by the Blavatnik School of Government, and the University of Oxford called the Oxford COVID-19 Government Response Tracker (OxCGRT) data set [26]. This data recorded more than 17 policies and indices from more than 160 countries. This also included the Stringency Index (SI), a summary of the measure of the overall strictness of these policies. Previously, we conducted studies showing the relationship between these policies and indices in lowering COVID-19 cases in more than 90 countries and South Korea [13,27]. However, we observed that the SI provided by the OxCGRT data may not fully capture Korea’s containment and closure policies. To describe government responses simply, OxCGRT would summarize the policies implemented in each country into the same number of levels and then use an additive method to calculate the unweighted SI since the approach is transparent and easiest to interpret. However, this approach cannot capture small policy changes across all countries and adequately reflect the influence of different policies on an individual basis. Some countries applied more stringent policy changes that cannot be captured by the calculated OxCGRT SI. For example, South Korea implemented restriction policies targeting social gatherings and would relax or tighten the policies by increasing or decreasing the maximum number of people allowed at these social gatherings like for 8 to 10 or vice versa. These changes were too subtle to be reflected in the OxCGRT’s SI. Second, OxCGRT did not consider the impact of each policy on the spread of COVID-19 when calculating SI. This assumes that each policy will have an equal influence and contribution to the final policy score. OxCGRT encouraged users to carefully consider which combination and weighting of policies would best capture the dimensions they are seeking to measure [26]. Therefore, Korea Stringency Indices (KSI) better capture the subtle changes present in Korea’s COVID-19 response policies. In addition, we can focus on the target areas of the policies and tell whether the implemented policies and their level of strictness are having an impact on the people who frequent those target areas.

Although Korea was one of the first countries to be affected by COVID-19 [28], many countries, like China, Uganda, Europe, and the United States, implemented a wide range of stringent restriction policies in response to the rapid increase in the daily number of COVID-19 cases. But, after experiencing a sharp growth in COVID-19 cases early in the pandemic, Korea rapidly controlled transmission while implementing less stringent social distancing measures than other countries. The government only implemented free nationwide public testing, isolation of infected individuals, vaccination, and information of the public through text messages together with government press conferences [21,29]. The government maintained this course of less strict policies even when the daily cases rose to tens of thousands without locking down regions and causing severe economic damage [28]. Analyzing the effectiveness of these policies in Korea can answer questions about which policies were most effective at managing the pandemic situation in Korea and provide a framework for responses that can be adopted against future pandemics without resorting to stringent measures that affect people mentally and the economy. Maybe we can also find loopholes in Korea’s response strategy and improve its efficiency.
In our analysis, we aim to answer the question of which implemented government policies were most effective in managing the COVID-19 situation in Korea and calculate KSI using the COVID-19 restriction policies implemented in Korea. The Korean government implemented restriction policies targeting 4 target areas: public facilities, public events, religious gatherings, and social gatherings. The policies and their level of strictness were manually obtained from this website [25]. Our final goal is to provide KSI, which is the summarized score to express the strictness of the policies considering their impact on the spread of COVID-19. We first assumed that each policy may have different impacts on the spread of COVID-19. Given this reason, the impact of these policies on the spread of COVID-19 was estimated using a random forest (RF) model. In addition, RF is not restricted by the multiple levels that are present in a given policy. All the information given by the levels in each given policy is used to estimate the impact (feature importance) of the policy on the spread of COVID-19. These feature importance values are taken as the weights of the policies and used for calculating the summarized score of each policy category. The basis for this assumption is that even if the amount of change in 2 policies is the same, the more important policy should be weighted more in the summarized score. Finally, we assumed that the impact of each policy on COVID-19 could be different by each period. Therefore, the analysis was carried out in 5 segments or periods corresponding to waves of the pandemic or turning points during the pandemic. Such turning points classify the behavior of a country’s trajectory throughout the pandemic as being in (or over) their subsequent waves.

**Methods**

**Data**

The COVID-19 daily series of confirmed cases in Korea was downloaded from the “Our World in Data” (OWID) website [30]. Countries implemented different restriction policies with the progress of the pandemic, and these policies were recorded by the OxCGRT [26,31]. The OxCGRT data systematically recorded the different levels of the policies on an ordinal or numeric scale. The SI was obtained from the OxCGRT data and records the average level of strictness of the policies that primarily contain and restrict people’s behavior and movements (a summary of the measure of the overall strictness of these policies).

**Korea’s Restriction Policies**

The COVID-19 restriction policies implemented in Korea and their strictness were collected manually from the government’s official website of the Korea Disease Control and Prevention Agency (KDCA) in response to COVID-19 [25]. The KDCA gave press releases about the COVID-19 situation in the country. The social distancing policies and their strictness levels for a given period were announced during these press releases and also on the official KDCA website. All press release posts by the KDCA for our analysis period (November 01, 2020, to April 24, 2022) were checked (1532 posts), and the social distancing policy level for the announced given period was recorded in a spreadsheet with their corresponding dates. If there were changes to the social distancing policy, they were summarized in the spreadsheet for the given date. However, if there was no updated policy notice for a certain period, we assumed the same policy level had been maintained. These policies focused on 4 main activities with different targets under each, including public facilities, religious gatherings, public events, and social gatherings, as summarized in Table 1. The level of strictness or stringency of the policies varied with the number of confirmed cases and is outlined in the “score” column of Table 2. A higher score value means a more vital constraint.
<table>
<thead>
<tr>
<th>Category and target</th>
<th>Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Public facilities</strong></td>
<td></td>
</tr>
<tr>
<td>Night entertainment facilities</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 1</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 2</td>
</tr>
<tr>
<td>Restaurant</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 3</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 4</td>
</tr>
<tr>
<td>Coffee shop</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 5</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 6</td>
</tr>
<tr>
<td>Karaoke</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 7</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 8</td>
</tr>
<tr>
<td>Baths or saunas</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 9</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 10</td>
</tr>
<tr>
<td>Indoor sports facilities</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 11</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 12</td>
</tr>
<tr>
<td>Sales promotion center</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 13</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 14</td>
</tr>
<tr>
<td>Cinema</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 15</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 16</td>
</tr>
<tr>
<td>Supermarket</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 17</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 18</td>
</tr>
<tr>
<td>PC rooms</td>
<td></td>
</tr>
<tr>
<td>Operating hours restrictions</td>
<td>Policy 19</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>Policy 20</td>
</tr>
<tr>
<td><strong>Public events</strong></td>
<td></td>
</tr>
<tr>
<td>Personnel</td>
<td></td>
</tr>
<tr>
<td>Number of gathering restrictions (all people)</td>
<td>Policy 21</td>
</tr>
<tr>
<td>Number of gathering restrictions (vaccinated people)</td>
<td>Policy 22</td>
</tr>
<tr>
<td>Religious gatherings</td>
<td></td>
</tr>
<tr>
<td>Personnel</td>
<td></td>
</tr>
<tr>
<td>Number of gathering restrictions (all people)</td>
<td>Policy 23</td>
</tr>
<tr>
<td>Number of gathering restrictions (vaccinated people)</td>
<td>Policy 24</td>
</tr>
<tr>
<td>Social gathering</td>
<td></td>
</tr>
<tr>
<td>Personnel</td>
<td></td>
</tr>
<tr>
<td>Number of gathering restrictions (all people)</td>
<td>Policy 25</td>
</tr>
<tr>
<td>Category and target</td>
<td>Policy</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Number of gathering restrictions (unvaccinated people)</td>
<td>Policy 26</td>
</tr>
<tr>
<td>Number of gathering restrictions (penalty after 6 PM)</td>
<td>Policy 27</td>
</tr>
<tr>
<td><strong>Exceptions</strong></td>
<td></td>
</tr>
<tr>
<td>Number of gathering restrictions (family living together, families needing care, and those on their deathbed)</td>
<td>Policy 28</td>
</tr>
<tr>
<td>Number of gathering restrictions (sports)</td>
<td>Policy 29</td>
</tr>
<tr>
<td>Number of gathering restrictions (immediate family)</td>
<td>Policy 30</td>
</tr>
<tr>
<td><strong>Restaurants and coffee shops</strong></td>
<td></td>
</tr>
<tr>
<td>Number of gathering restrictions (vaccinated people)</td>
<td>Policy 31</td>
</tr>
<tr>
<td>Number of gathering restrictions (unvaccinated people)</td>
<td>Policy 32</td>
</tr>
<tr>
<td>Number of gathering restrictions (all people)</td>
<td>Policy 33</td>
</tr>
</tbody>
</table>
Table 2. A summary of Korea’s level of strictness for each restriction policy.

<table>
<thead>
<tr>
<th>Category, policy, and explanation</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Public facilities</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Operating hours restrictions</strong></td>
<td></td>
</tr>
<tr>
<td>Prohibition of operating</td>
<td>10</td>
</tr>
<tr>
<td>Allowing operation until 9:00 PM</td>
<td>7</td>
</tr>
<tr>
<td>Allowing operation until 10:00 PM</td>
<td>5</td>
</tr>
<tr>
<td>Allowing operation until 11:00 PM</td>
<td>4</td>
</tr>
<tr>
<td>Allowing operation until 12:00 PM</td>
<td>3</td>
</tr>
<tr>
<td>No restrictions</td>
<td>1</td>
</tr>
<tr>
<td><strong>Prohibition of indoor eating</strong></td>
<td></td>
</tr>
<tr>
<td>Prohibition of operating</td>
<td>10</td>
</tr>
<tr>
<td>Prohibition of indoor eating</td>
<td>5</td>
</tr>
<tr>
<td>No restrictions</td>
<td>1</td>
</tr>
<tr>
<td><strong>Public events</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Number of gathering restrictions—for all people</strong></td>
<td></td>
</tr>
<tr>
<td>Prohibition of gathering</td>
<td>10</td>
</tr>
<tr>
<td>Limited to 49</td>
<td>6</td>
</tr>
<tr>
<td>Limited to 99</td>
<td>5</td>
</tr>
<tr>
<td>Limited to 250</td>
<td>4</td>
</tr>
<tr>
<td>Limited to 299</td>
<td>3</td>
</tr>
<tr>
<td><strong>Number of gathering restrictions—for vaccinated group</strong></td>
<td></td>
</tr>
<tr>
<td>Prohibition of gathering</td>
<td>10</td>
</tr>
<tr>
<td>Limited to 49</td>
<td>6</td>
</tr>
<tr>
<td>Limited to 99</td>
<td>5</td>
</tr>
<tr>
<td>Limited to 250</td>
<td>4</td>
</tr>
<tr>
<td>Limited to 299</td>
<td>3</td>
</tr>
<tr>
<td>Limited to 499</td>
<td>2</td>
</tr>
<tr>
<td>No restrictions</td>
<td>1</td>
</tr>
<tr>
<td><strong>Religious gatherings</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Number of gathering restrictions—for all people</strong></td>
<td></td>
</tr>
<tr>
<td>Prohibition of gathering</td>
<td>10</td>
</tr>
<tr>
<td>Limited to 10% of maximum occupancy</td>
<td>6</td>
</tr>
<tr>
<td>Limited to 20% of maximum occupancy</td>
<td>5</td>
</tr>
<tr>
<td>Limited to 30% of maximum occupancy</td>
<td>4</td>
</tr>
<tr>
<td>Limited to 50% of maximum occupancy</td>
<td>3</td>
</tr>
<tr>
<td>Limited to 70% of maximum occupancy</td>
<td>2</td>
</tr>
<tr>
<td>No restrictions</td>
<td>1</td>
</tr>
<tr>
<td><strong>Number of gathering restrictions—for vaccinated group</strong></td>
<td></td>
</tr>
<tr>
<td>Prohibition of gathering</td>
<td>10</td>
</tr>
<tr>
<td>Limited to 10% of maximum occupancy</td>
<td>6</td>
</tr>
<tr>
<td>Limited to 20% of maximum occupancy</td>
<td>5</td>
</tr>
<tr>
<td>Limited to 30% of maximum occupancy</td>
<td>4</td>
</tr>
<tr>
<td>Limited to 70% of maximum occupancy</td>
<td>2</td>
</tr>
</tbody>
</table>
### Category, policy, and explanation

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No restrictions</td>
</tr>
</tbody>
</table>

#### Social gatherings

**Number of gathering restrictions—for all people**

- Limited to 4: 6
- Limited to 6: 5
- Limited to 8: 4
- Limited to 10: 3
- No restrictions: 1

**Number of gathering restrictions—for nonvaccinated group**

- Limited to 2: 7
- Limited to 4: 6
- Limited to 6: 5
- Limited to 7: 4
- Limited to 10: 3
- No restrictions: 1

**Number of gathering restrictions—a penalty after 6 PM**

- Penalty on the number of gatherings after 6:00 PM: 2
- No penalty: 1

**Exceptions: family living together, a family who needs care, a deathbed**

- Same restrictions on these people (No exception): 2
- No restrictions on these people: 1

**Exceptions: sports**

- Same restrictions on these people (No exception): 2
- No restrictions on these people: 1

**Exceptions: immediate family**

- Same restrictions on these people (No exception): 2
- No restrictions on these people: 1

#### Restaurant and coffee shop: vaccinated people

- Limited to 4: 6
- Limited to 6: 5
- Limited to 8: 4
- Limited to 10: 3
- No restrictions: 1

#### Restaurant and coffee shop: nonvaccinated people

- Limited to 1: 8
- Limited to 2: 7
- Limited to 4: 6
- Limited to 6: 5
- Limited to 7: 4
- Limited to 10: 3
- No restrictions: 1

#### Restaurant and coffee shop: sharing a table with nonvaccinated people

- No table sharing between nonvaccinated and vaccinated persons: 2
Analysis Periods
The analysis was carried out in 5 periods. Period 1 corresponds to the third wave of the pandemic ranging from November 01, 2020, to January 20, 2021. Period 2 comes after the 3rd wave of the pandemic and ranges from January 20, 2021 to June 27, 2021. Period 3 combines the first and second analysis periods (November 01, 2020, to June 27, 2021). Period 4 consists of the summer vacation period and right after the summer vacation period, from June 27, 2021, to November 01, 2021. Period 5 is from November 01, 2021, to April 24, 2022, the step-by-step recovery period. This period is characterized by the major cities in the country taking steps to go back to normal operations like before the pandemic and the end of the social distancing policies except masking. The analysis period is summarized in Figure 1 below. The end of the analysis period 4 is the beginning of period 5 but with a different scale for the y-axis.

Figure 1. The 5 analysis periods used in the study.

Statistical Analysis
In the first step, an RF model was used to estimate the feature importance of policies corresponding to each target level policy, respectively against COVID-19 daily confirmed cases. Since there are 15 target policies, 15 RF models for each target policy were fitted each independent of the other target policies. Then, the estimated feature importance of the 33 policies was multiplied by their respective policy “score” values and summed to form the 15 policies according to the target category level as shown in Table 1. Public facilities had 10 target policies, public events and religious gatherings had 1 target policy each (personnel), and social gatherings had 3 target policies (restaurants and cafes, exceptions, and personnel). In the second step, another RF model was fitted with the 15 target-level policies and their feature importance was estimated. Then, the KSI for each category is estimated by multiplying the estimated feature importance of each target policy with the calculated target policies estimated above and summed corresponding to each category. The RF model is an ensemble learner based on randomized decision trees and provides different feature importance measures, one from statistical permutation tests and the other from training an RF classifier. Both measures have been found to correlate reasonably well and provide excellent means of measuring feature relevance [32]. Therefore, although simple, the calculation of KSIs using this method is justifiable. Since there are 4 categories, we will get 4 KSI values, 1 for each category. In the last step, an RF model is used to rank the impact of the restriction policies using KSIs on COVID-19 daily cases using feature importance. The SI from the OxCGRT data was included in the analysis.

During the analysis, a training validation approach was used to find optimal hyperparameters, with the last 14 days used as a validation set in each analysis period. A detailed explanation of hyperparameter settings is explained in Table 3 below.
Table 3. Optimal hyperparameter settings used in the analysis.

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Range</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td># of trees</td>
<td>[4, 8, 16, 32, 64, 100]</td>
<td>The number of trees in the forest.</td>
</tr>
<tr>
<td>Max features</td>
<td>['auto', 'sqrt']</td>
<td>The number of features to consider at every split.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Auto: max features = # of features.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sqrt: max features = sqrt (# of features)</td>
</tr>
<tr>
<td>Max depth</td>
<td>[None, 10, ..., 110]</td>
<td>The maximum number of levels in the tree.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None: no limit to the depth of the tree.</td>
</tr>
<tr>
<td>Min sample split</td>
<td>[2, 5, 10]</td>
<td>The minimum number of samples required to split a node.</td>
</tr>
<tr>
<td>Min sample leaf</td>
<td>[1, 2, 4]</td>
<td>The minimum number of samples required at each leaf node.</td>
</tr>
</tbody>
</table>

Ethical Considerations

No individual data were used in our study. Only cumulative count data of anonymous individuals were used. Since publicly available data were used for this study, ethical approval from a research board was unnecessary.

Results

Feature Importance of the Policies

The feature importance of the 33 policies and the 15 target policies for the 5 analysis periods is summarized in Figure 2. The feature importance of the 33 policies varies much across the analysis periods and we observe a pattern of low, intermediate, and high feature importance among the policies and across the 5 analysis periods. Most restriction policies targeting public events and religious gatherings had their feature importance around 0.5 across the analysis periods except for period 5. Restriction policies targeting public facilities also had their importance mainly around 0.50, except in period 5 which had importance values of either 0 or 1.00. The most prominent impact was observed with the restriction policies that targeted the use of public facilities. A few policies stood out, especially in the step-by-step recovery period (period 5), including operating hour restrictions on cinemas, restaurants, PC rooms, indoor sports facilities, karaoke, coffee shops, night entertainment facilities, and baths or saunas. After the summer vacation (period 4), operating hour restrictions were imposed on supermarkets, cinemas, sales promotion centers, restaurants and coffee shops. Period 3 which combines periods 1 and 2 had feature importance for most policies at around 0.50. We conclude that most policies associated with operating hour restrictions followed by the number of gathering restrictions had higher feature importance values than the other policies for a given analysis period, respectively.

There was not much variation in the feature importance of the 15 target policies (Figure 2B) obtained from summing up the 33 policies across the 5 analysis periods. All feature importance was below 0.40, except for the restriction on the number of gatherings in public places in period 2 (0.80). For some policies, their importance was either equal to or near 0 at a given analysis period.
Figure 2. The feature importance values were obtained from RF models across the 5 analysis periods. (A) Results from the 15 RF models for the 33 policies for each period. (B) Results from 1 RF model for the 15 target policies for each period. RF: random forest.

Korea Stringency Indices

Four KSIs for public events, religious gatherings, social gatherings, and public facilities were calculated using the feature importance values and score values of policies under each target policy and category, respectively. In addition to SI obtained from the OxCGRT data, their variation with daily confirmed cases for the 5 analysis periods is shown in Figure 3. SI had higher indices across all analysis periods. Public facilities had relatively higher stringency indices than the other KSIs from periods 1 to 5, followed by religious gatherings, social gatherings, and public events, respectively. Furthermore, we observed that the stringency indices of KSIs tended to increase with the number of daily confirmed cases and decrease with decreasing daily cases.

The KSIs had their feature importance below 0.20 in most analysis periods as shown in Figure 4. Exceptions were observed with public facilities in period 4, public events in period 2, social gatherings in period 5, and religious gatherings for period 1 and period 3.
Figure 3. Variation of the KSIs and the OxCGRT SI with daily confirmed cases for (A) analysis periods 1, 2, and 4; (B) analysis period 5; and (C) analysis period 3+4. KSI: Korea Stringency Index; SI: Stringency Index.
Correlation Analysis

The correlation analysis was carried out between KSIs, SI, and daily confirmed cases, as shown in Figure 5. In period 1 of the pandemic, all correlations were positive and statistically significant, with restrictions in the use of public facilities ($r=0.86$, $P<.001$), religious gatherings ($r=0.86$, $P<.001$), public events ($r=0.76$, $P<.001$), and SI ($r=0.71$, $P<.001$) being the strongest, respectively. Strong correlations were also observed between the KSIs, especially between public facilities and religious gatherings ($r=0.99$, $P<.001$), public facilities and public events ($r=0.88$, $P<.001$), and religious gatherings and public events ($r=0.92$, $P<.001$). However, a weak relationship was observed with restrictions in social gatherings ($r=0.38$, $P<.001$). In period 2 of the pandemic, a weaker and negative relationship between KSIs, SI, and daily confirmed cases was observed. However, there were strong correlations between public facilities and religious gatherings ($r=0.81$, $P<.001$) and religious and social gatherings ($r=0.82$, $P<.001$). Weak and sometimes negative correlations were observed in periods 4 and 5 between the KSIs and daily confirmed cases. However, strong correlations were observed between social and religious gatherings ($r=0.82$, $P<.001$) and public facilities and public events ($r=0.81$, $P<.001$) in period 5. The detailed results are summarized in Table S1 in Multimedia Appendix 1.
Figure 5. Correlation analysis plot for (A) analysis period 1; (B) analysis period 2; (C) analysis period 4; and (D) analysis period 5. The asterisk (*) indicates a statistically significant correlation. NC: new cases; PF: public facilities; PH: public events; RG: religious gatherings; SG: social gatherings; SI: Stringency Index.

Discussion

From our analysis, we observed that policies associated with operation hour restrictions and the number of gathering restrictions had a higher impact than other policies in the same analysis periods. The greatest impact was observed with the restrictions on the use of public facilities, especially in operating hour restrictions of cinemas, restaurants, coffee shops, indoor sports facilities, PC rooms, karaoke, and baths or saunas. The next impact was with policies restricting the number of people gatherings in public, religious, and social gatherings. For KSIs, their feature importance was below 0.20 in most analysis periods except public events in period 2, social gatherings in period 5, religious gatherings in periods 1 and 3, and public events in period 4. Although the stringency of SI was higher than KSIs, its feature importance with COVID-19 daily cases was lower than 0.40 across all analysis periods. Furthermore, we observed that different policies are effective at different analysis periods. Other periods of the pandemic require the enforcement of varying restriction policies. For example, the summer period is accompanied by a lot of indoor and outdoor activities and travel which increases the contact among people in the population, thus the rate of daily confirmed cases. To lower the wave of daily cases, the number of gathering restrictions and the use of public facilities, especially in PC rooms, restaurants and coffee shops, and cinemas would be more effective. These policies would reduce the mixing of people, allowing confirmed cases to be isolated and recovered. After flattening the curve, we can ease the strictness of using public facilities and social gatherings while maintaining a high stringency level on religious and public gatherings. Therefore, different policies have to be strengthened depending on the situation of the pandemic and significant activities happening in the country.

Correlation analysis showed a strong positive relationship between KSIs and daily confirmed cases in period 1 of the pandemic. A rise in the average daily confirmed cases corresponded with an increase in the stringency levels of the policies. The increase in the KSIs stringency indices leads to lower numbers of COVID-19 daily confirmed cases. In period 2, the relationship between the KSIs and daily confirmed cases was negative and weaker, maybe due to the easing of the strictness of the policies because of the flattening of the cases’ curve. From then on, the correlations stayed weak, and the relationship was either negative or positive, depending on the KSI. This can be attributed to easing the restriction policies’ strictness because most of the population was already vaccinated and opening the country back to normal operations. The correlation of SI was generally lower than those of the KSIs, showing that using KSIs to evaluate the effectiveness of restriction policies on COVID-19 cases is better than the OxCGRT SI. The KSIs provide more information about the effectiveness of the different policies at different target categories as the pandemic progresses.

Some argue that the COVID-19 pandemic caused by the contact-transmissible SARS-CoV-2 virus is our most significant health crisis [33]. However, its global impact on health, business, education, travel, international security, and other aspects of life cannot be disputed. Its outbreak exposed the weakness of
health care systems, countries, and world preparedness in handling a sudden global crisis. Due to the lack of an effective treatment or vaccine, the pandemic elicited unprecedented restriction (or social distancing) policies to mitigate and suppress the spread of the virus. Around 110 countries have implemented at least 1 restriction policy against COVID-19 [26]. Standard policies included school closures, travel restrictions, bans on public gatherings, stay-at-home orders, closure of public transportation, emergency investments in the health care system, contact tracing, and investments in COVID-19 vaccines. Due to its high recovery rate of above 95% [34], the implementation of the restriction policies was also met with resistance from the general population as the social distancing period lengthened mainly toward vaccination and wearing masks [35-43]. Opposition came primarily from business holders, workers, and parents of school-going children because of the impact of these restrictions on education and businesses. However, evidence from several scientific studies and disease modeling continued to support the effectiveness of these policies on COVID-19 cases, thereby motivating governments to continue their implementation [44], especially with the emergence of more severe and transmissible SARS-CoV-2 variants like the delta and omicron variants, respectively [45-49].

The Korean government without resorting to any draconian “lockdown” policies like other countries of Uganda, China, the United States, etc [21,28,29], resorted to population testing combined with contact tracing, early isolation, free treatment of positive cases and digital technologies to fight the pandemic. Thirty-three restriction policies targeting public facilities, public events, social gatherings, and religious gatherings were implemented in Korea during the pandemic. These policies mainly targeted operating hour restrictions of public facilities or places and restrictions on the number of people gathering. Therefore, analysis of these policies to determine which government-implemented policies were most effective in managing the COVID-19 situation in Korea can provide a framework for responses that can be adopted against future pandemics. From our analysis, we observed the greatest impacts came from restrictions in public facilities, especially in operating hour restrictions of cinemas, restaurants, coffee shops, indoor sports facilities, PC rooms, karaoke, and baths or saunas, followed by restrictions to the number of people in public, religious and social gatherings. From the literature, an analysis of restriction policies in more than 90 countries showed the positive impact of containment policies like cancellation of public events, school and workplace closures, stay-at-home requirements, and restriction on gatherings, on the spread of COVID-19 [17]. Travel restrictions were shown to be effective in reducing the risk of imported SARS-CoV-2 cases [50].

Masking was found to be the most cost-effective nonpharmacological intervention implemented, delivering 4 times more impact than school closures and approximately double that of other mobility restrictions [51]. Gathering restrictions were the second most effective while international travel controls and public information campaigns had negligible effects [51]. From the above literature review, we consistently observe policies associated with gathering restrictions or restrictions on places where many people meet and mix like public places, workplaces, and schools, to always stand out among other policies, which is consistent with our findings of restrictions on public facilities and number of gathering restrictions.

The major limitation of our analysis was the collection of the restriction policy levels, which was done manually. This was time-consuming. Furthermore, since the policy levels were summarized from the KDCA website, some unnoticed changes may have been missed and thus not incorporated. In addition, the strictness levels and their compliance varied greatly between metropolitan areas like Seoul and Gyeonggi-do, which had a higher average number of daily cases, and the nonmetropolitan areas. However, we only focused on the national restriction levels when summarizing the restriction policy levels and calculated only nationwide-specific indices. Local or regional information was neglected and not incorporated when calculating indices or summarizing policy levels. Local or regional information must be considered as it is more insightful, especially considering countries such as the United States, which have large territories and large differences in regional policies. Furthermore, although our analysis focused only on Korea due to a lack of detailed information about the restriction policies of other countries, our method of calculating the stringency indices is simple and can easily be applied to other countries. However, the reliability of the developed indices will vary greatly depending on the types of policies implemented and the frequency of policy information provided by each country. In the future, the crawl-based semiautomated systems can be adapted to collect policy data, after which the validity of the data can be confirmed manually. To increase the generalizability of the result, we can consider developing the region-specific stringency indices using regional data. Furthermore, considering the population demographic information like age may provide more in-depth results and compliance levels across age groups. In addition, more robust methods of calculating indices need to be explored.

In conclusion, nonpharmacological restriction policies that aim for physical distancing have a strong potential to suppress the spread of COVID-19 and lead to a smaller number of overall cases. From our analysis, we observed that restrictions on the use of public facilities focusing on restaurants, coffee shops, cinemas, PC rooms, saunas or baths, indoor sports facilities, and restrictions on the number of gatherings, have played a significant role in slowing down the spread of the disease, thus buying time for the health care systems and governments. Given the high impact of restriction policies on public facilities, a nonpharmacological restriction policy framework can be designed with restrictions on public facilities being the main focus for diseases that spread through contact between people in the population. However, different periods call for enforcing different policies as their effectiveness can vary during the pandemic. Since nonpharmacological restriction policies alone cannot effectively combat a spreading pandemic, other factors like an effective health care plan or treatment, population demographics, and population mobility must be considered for either long-term or short-term impact. The compliance from the public must also be considered for both short and long-term effectiveness.
Acknowledgments
This research was supported by research grants from the Ministry of Science and Information & Communication Technology, South Korea (Number 2021M3E5E3081425).

Data Availability
The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions
Conceptualization was done by TP. Methodology was developed by TP and KH. Formal analysis was performed by DL. Resources were managed by DL and BL. Visualization was undertaken by KH. Writing of the original draft was carried out by CA. Review and editing were contributions from CA, TP, and KH. Supervision was provided by TP. Project administration was overseen by TP. Funding acquisition was secured by TP.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Detailed results of the correlation analysis.

References


29. South Korea’s combating COVID-19 under the rule of law. Law E, Far NUS. 2021. URL: https://intr2dok.viforea.co.kr/en/ [accessed 2023-02-16]


33. Apio et al. JMIR Public Health and Surveillance 2024 | vol. 10 | e47099 | p.2244 https://publichealth.jmir.org/2024/1/e47099


Abbreviations

- KDCA: Korea Disease Control and Prevention Agency
- KSI: Korea Stringency Index
- OWID: Our World in Data
- OxCGRT: Oxford COVID-19 Government Response Tracker
- RF: random forest
- SI: Stringency Index

Edited by A Mavragani, T Sanchez; submitted 08.03.23; peer-reviewed by Y Shao, DA Mertes; comments to author 01.06.23; revised version received 28.07.23; accepted 21.11.23; published 08.01.24.

Please cite as:

Apio C, Han K, Lee D, Lee B, Park T
Development of New Stringency Indices for Nonpharmacological Social Distancing Policies Implemented in Korea During the COVID-19 Pandemic: Random Forest Approach
JMIR Public Health Surveill 2024 | vol. 10 | e47099 | p.2245
https://publichealth.jmir.org/2024/1/e47099

PMID: 38190233
Impact of the COVID-19 Pandemic on Health Care Utilization in the Vaccine Safety Datalink: Retrospective Cohort Study

Lei Qian, PhD; Lina S Sy, MPH; Vennis Hong, MPH; Sungching C Glenn, MS; Denison S Ryan, MPH; Jennifer C Nelson, PhD; Simon J Hambidge, MD, PhD; Bradley Crane, MD; Ousseny Zerbo, PhD; Malini B DeSilva, MD, MPH; Jason M Glanz, PhD; James G Donahue, DVM, PhD; Elizabeth Liles, MD, MCR; Jonathan Duffy, MD, MPH; Stanley Xu, PhD

1Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States
2Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States
3Denver Health Ambulatory Care Services, Denver, CO, United States
4Center for Health Research, Kaiser Permanente Northwest, Portland, OR, United States
5Kaiser Permanente Vaccine Study Center, Kaiser Permanente Northern California, Oakland, CA, United States
6HealthPartners Institute, Bloomington, MN, United States
7Institute for Health Research, Kaiser Permanente Colorado, Aurora, CO, United States
8Marshfield Clinic Research Institute, Marshfield, WI, United States
9Centers for Disease Control and Prevention, Atlanta, GA, United States
10Department of Health Systems Science, Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, United States

Abstract

Background: Understanding the long-term impact of the COVID-19 pandemic on health care utilization is important to health care organizations and policy makers for strategic planning, as well as to researchers when designing studies that use observational electronic health record data during the pandemic period.

Objective: This study aimed to evaluate the changes in health care utilization across all care settings among a large, diverse, and insured population in the United States during the COVID-19 pandemic.

Methods: We conducted a retrospective cohort study within 8 health care organizations participating in the Vaccine Safety Datalink Project using electronic health record data from members of all ages from January 1, 2017, to December 31, 2021. The visit rates per person-year were calculated monthly during the study period for 4 health care settings combined as well as by inpatient, emergency department (ED), outpatient, and telehealth settings, both among all members and members without COVID-19. Difference-in-difference analysis and interrupted time series analysis were performed to assess the changes in visit rates from the prepandemic period (January 2017 to February 2020) to the early pandemic period (April-December 2020) and the later pandemic period (July-December 2021), respectively. An exploratory analysis was also conducted to assess trends through June 2023 at one of the largest sites, Kaiser Permanente Southern California.

Results: The study included more than 11 million members from 2017 to 2021. Compared with the prepandemic period, we found reductions in visit rates during the early pandemic period for all in-person care settings. During the later pandemic period, overall use reached 8.36 visits per person-year, exceeding the prepandemic level of 7.49 visits per person-year in 2019 (adjusted percent change 5.1%, 95% CI 0.6%-9.9%); inpatient and ED visits returned to prepandemic levels among all members, although they remained low at 0.095 and 0.241 visits per person-year, indicating a 7.5% and 8% decrease compared to pre-pandemic levels among members without COVID-19, respectively. Telehealth visits, which were approximately 42% of the volume of outpatient visits during the later pandemic period, were increased by 97.5% (95% CI 86.0%-109.7%) from 0.865 visits per person-year in
2019 to 2.35 visits per person-year in the later pandemic period. The trends in Kaiser Permanente Southern California were similar to those of the entire study population. Visit rates from January 2022 to June 2023 were stable and appeared to be a continuation of the use levels observed at the end of 2021.

**Conclusions:** Telehealth services became a mainstay of the health care system during the late COVID-19 pandemic period. Inpatient and ED visits returned to prepandemic levels, although they remained low among members without evidence of COVID-19. Our findings provide valuable information for strategic resource allocation for postpandemic patient care and for designing observational studies involving the pandemic period.

**Methods**

**Study Setting**

We conducted a retrospective cohort study within 8 health care organizations (Kaiser Permanente: Southern California, Northern California, Northwest, Washington, and Colorado; HealthPartners; Denver Health; and Marshfield Clinic) participating in the Vaccine Safety Datalink (VSD) Project [15]. The VSD is a collaborative project between the Centers for Disease Control and Prevention’s Immunization Safety Office and integrated health care organizations, with over 9 million individuals annually enrollment representing approximately 3% of the US population. Each participating site creates standardized data sets with information on members’ demographic characteristics, membership enrollment, vaccination history, and medical encounters in the inpatient, ED, outpatient, and telehealth settings.

**Ethical Considerations**

The institutional review board at each study site approved the study with a waiver of informed consent, given that the data-only retrospective research activities were minimal risk. To protect the privacy and confidentiality of human participants, all staff working on the research study were trained in procedures to protect the privacy of medical record information. Study participants were not compensated given the observational nature of the study.

**Study Population**

The study population consisted of individuals of all ages who were active health plan members at any time from January 1, 2017, to December 31, 2021 from 8 participating VSD sites. Members who did not test positive for COVID-19 or have a medically attended COVID-19 diagnosis at any time during the study period were classified as “members without COVID-19.” The exploratory analysis to assess the more recent trends included KPSC members from January 1, 2022, to June 30, 2023.
Health Care Utilization

Four visit types were evaluated based on care setting: inpatient visit, ED visit, outpatient visit, and telehealth visit. Telehealth visits were conducted synchronously using telephone or live video-audio interaction. The visit rates per person-year from these 4 care settings were calculated monthly during the study period from 2017 to 2021. The numerator was the visit count of each type, and the denominator was person-years of membership during a given month. We also calculated the overall visit rates (all 4 settings combined), the rates of combined outpatient and telehealth visits (as total ambulatory care), and the rates of combined in-person (inpatient, ED, and outpatient) visits.

Statistical Analysis

Overview

Demographic characteristics (age, sex, and race or ethnicity) of the study population were described by year. Monthly visit rates overall and by visit type were plotted over time before and during the COVID-19 pandemic to examine trends.

We selected April-December 2020 and July-December 2021 as the early and later pandemic periods, respectively, to assess the short- and longer-term impact of the pandemic on health care utilization. The July-December 2021 period was selected as the later pandemic period because COVID-19–related restrictions were relaxed during this period.

Short-Term Impact Assessment

To assess the short-term impact of the pandemic on health care utilization during the early pandemic period, due to data volatility during this period, we performed difference-in-difference (DiD) analyses by comparing visit rates in each of the 3 periods of the early pandemic (April-June, July-September, and October-December) in 2020 to the January-February 2020 before the pandemic period; visit rates in the 4 corresponding periods in 2017-2019 were also calculated and used to control for seasonality (Figure S1A in Multimedia Appendix 1) [16]. We excluded March 2020 from the analysis because the COVID-19 pandemic was declared a national emergency on March 13, 2020. We estimated the ratio of rate ratios (RRRs) and 95% CIs by fitting Poisson regression models with the visit count as the dependent variable and the natural logarithm of person-years of membership as an offset, adjusting for overdispersion of the count data. The percent change of visit rates during the early pandemic period from the prepandemic period was calculated as \((\text{RRR} - 1) \times 100\). Separate analyses were conducted for all visit types combined and by visit type.

Long-Term Impact Assessment

To assess the longer-term impact of the pandemic on health care utilization during the later pandemic period, interrupted time series (ITS) analyses were conducted with adjustment for prepandemic secular trends and seasonality by including calendar year and month in the model (Figure S1B in Multimedia Appendix 1) [17]. ITS analysis is a statistical method used to evaluate the impact of an intervention or event, which is introduced at a specific time point on time series data. A model with a linear function of the year was selected based on visual exploration of the trend plots and comparison with models with a quadratic form. Poisson regression models were used to estimate adjusted rate ratios (RRs) and 95% CI, comparing visit rates in the later pandemic period with the prepandemic period. The percent change in visit rates was calculated as \((\text{RR} - 1) \times 100\). ITS analyses were conducted for all visits combined and separately by visit type.

COVID-19–related care associated with each wave of COVID-19 infection might have affected the secular patterns of health care utilization. We also conducted sensitivity analyses of health care utilization among members without evidence of COVID-19. For the exploratory analysis, we plotted monthly visit rates overall and by visit type. We also conducted ITS analysis to estimate the adjusted percent change of visit rates in July-December 2022 compared to prepandemic levels. All analyses were conducted using SAS (version 9.4; SAS Institute Inc).

Results

Annual enrollment did not significantly change between 2017 and 2021, with the total number of enrolled person-years ranging from 10.9 to 11.5 million across the 8 VSD sites. Demographic characteristics of members enrolled in VSD sites remained relatively stable during 2017-2021 (Table 1).

The trend plot showed that the overall monthly visit rates fluctuated normally before March 2020, then decreased abruptly in March, reaching the lowest levels in April 2020, and then gradually increased afterward (Figure 1). The overall visit rates returned to prepandemic levels during 2021 and were stable after July 2021. Use trends differed by visit type (Figure 2). All in-person (inpatient, ED, and outpatient) visits decreased markedly in the beginning of the pandemic, and then gradually increased. Among the 3 in-person visit types, outpatient visits had the largest decrease at the beginning of the pandemic. Telehealth visits increased sharply after the start of the pandemic, and then slowly decreased, although they remained above the prepandemic levels. Combined outpatient and telehealth visit rates decreased in the early pandemic period and exceeded prepandemic levels in 2021.

The overall visit rate decreased from 7.82 visits per person-year in January-February 2020 to 5.72 visits per person-year in April-June 2020, a 26.4% reduction (95% CI –36.3% to –15.1%); it returned to prepandemic levels in October-December 2020 (adjusted percent change –0.4%, 95% CI –6.3% to 5.9%; Table 2). The overall visit rate increased by 5.1% (95% CI 0.6%–9.9%) to 8.36 visits per person-year in July-December 2021 compared with the same months in the prepandemic period (Table 3). The rates of inpatient, ED, and outpatient visits decreased by 24.9% (95% CI –33.3% to –15.3%), 35.5% (95% CI –43.7% to –26.2%), and 64.2% (95% CI –71.9% to –54.3%) in April-June 2020 and by 5.3% (95% CI –8.3% to –2.2%), 19.2% (95% CI –22.7% to –15.5%), and 30.5% (95% CI –34.4% to –26.4%) in October-December 2020 compared to the rates of 0.108, 0.288, and 6.42 visits per person-year in January-February 2020, respectively (Table 2). While the rates of inpatient and ED visits returned to prepandemic levels in July-December 2021, the rate of outpatient visits was still 12.5%
lower (95% CI –16.8% to –8.0%; Table 3). Telehealth visit rates increased by 218.3% (95% CI 186.4%-253.6%) from 1.01 visits per person-year in January-February 2020 to 3.17 visits per person-year in April-June 2020 (Table 2). With a slow decrease after the initial rise, the telehealth visit rate still increased by 97.5% (2.35 visits per person-year; 95% CI 86.0%-109.7%) in July-December 2021 compared to the prepandemic period, comprising approximately 42% of outpatient visits (5.63 visits per person-year; Table 3). Combined telehealth and outpatient visits increased by 5.5% (95% CI 0.7%-10.6%) and combined in-person visits decreased by 11.9% (95% CI –16.0% to –7.7%) in July-December 2021 compared to prepandemic levels (Table 3).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Enrollment (person-years), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2017 (n=10,881,765)</td>
</tr>
<tr>
<td><strong>Age group (years)</strong></td>
<td></td>
</tr>
<tr>
<td>0-5</td>
<td>689,315 (6.3)</td>
</tr>
<tr>
<td>6-17</td>
<td>1,575,104 (14.5)</td>
</tr>
<tr>
<td>18-44</td>
<td>4,015,752 (36.9)</td>
</tr>
<tr>
<td>45-64</td>
<td>2,981,324 (27.4)</td>
</tr>
<tr>
<td>65+</td>
<td>1,620,270 (14.9)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5,650,298 (51.9)</td>
</tr>
<tr>
<td>Male</td>
<td>5,230,911 (48.1)</td>
</tr>
<tr>
<td>Unknown</td>
<td>556 (0.01)</td>
</tr>
<tr>
<td><strong>Race or ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>2,756,704 (25.3)</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>1,294,988 (11.9)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>742,060 (6.8)</td>
</tr>
<tr>
<td>Non-Hispanic multiple or other</td>
<td>473,214 (4.4)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>4,756,019 (43.7)</td>
</tr>
<tr>
<td>Unknown</td>
<td>858,780 (7.9)</td>
</tr>
</tbody>
</table>

Figure 1. Monthly overall visit rate among all members and members without COVID-19 during 2017-2021. March 13, 2020: the COVID-19 pandemic was declared a national emergency; April-December 2020: early pandemic period; July-December 2021: later pandemic period.
Figure 2. Monthly visit rate by care setting among all members during 2017-2021. March 13, 2020: the COVID-19 pandemic was declared as a national emergency; April-December 2020: early pandemic period; July-December 2021: later pandemic period. ED: emergency department.
Table 2. Visit rates before and during the early COVID-19 pandemic in 2020 and adjusted percent change from prepandemic among all members.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Visit rate in 2020 (per person-year)</th>
<th>Adjusted percent change (95% CI)a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jan-Feb</td>
<td>Apr-Jun</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.82</td>
<td>5.72</td>
</tr>
<tr>
<td>Setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPb</td>
<td>.108</td>
<td>.080</td>
</tr>
<tr>
<td>EDc</td>
<td>.288</td>
<td>.185</td>
</tr>
<tr>
<td>Opd</td>
<td>6.42</td>
<td>2.28</td>
</tr>
<tr>
<td>THe</td>
<td>1.01</td>
<td>3.17</td>
</tr>
<tr>
<td>OP+TH</td>
<td>7.43</td>
<td>5.45</td>
</tr>
<tr>
<td>In-personf</td>
<td>6.81</td>
<td>2.55</td>
</tr>
</tbody>
</table>

aItalic formatting indicates if P<.05.
bIP: inpatient.
cED: emergency department.
dOP: outpatient.
eTH: telehealth.
fIn-person: IP+ED+OP.

Table 3. Visit rates during 2017-2019 and the late COVID-19 pandemic in 2021 and adjusted percent change from prepandemic among all members.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Visit rate (per person-year)</th>
<th>Adjusted rate ratio (95% CI)a</th>
<th>Adjusted percent change (95% CI)a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jul-Dec 2017</td>
<td>Jul-Dec 2018</td>
<td>Jul-Dec 2019</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.89</td>
<td>7.11</td>
<td>7.49</td>
</tr>
<tr>
<td>Setting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPb</td>
<td>0.105</td>
<td>0.104</td>
<td>0.105</td>
</tr>
<tr>
<td>EDc</td>
<td>0.259</td>
<td>0.262</td>
<td>0.271</td>
</tr>
<tr>
<td>Opd</td>
<td>5.91</td>
<td>6.05</td>
<td>6.24</td>
</tr>
<tr>
<td>THe</td>
<td>0.625</td>
<td>0.694</td>
<td>0.865</td>
</tr>
<tr>
<td>OP+TH</td>
<td>6.53</td>
<td>6.74</td>
<td>7.11</td>
</tr>
<tr>
<td>In-personf</td>
<td>6.27</td>
<td>6.42</td>
<td>6.62</td>
</tr>
</tbody>
</table>

aItalic formatting indicates if P<.05.
bIP: inpatient.
cED: emergency department.
dOP: outpatient.
eTH: telehealth.
fIn-person: IP+ED+OP.

Enrolled person-years during 2017-2021 ranged from 8.8 to 9.0 million after excluding members with evidence of COVID-19. Demographic characteristics of members without COVID-19 remained similar compared to all members (Table S1 in Multimedia Appendix 2). The trends of overall monthly visit rates among members without COVID-19 and all members were similar (Figure 1). The overall visit rate among members without COVID-19 was lower than that of all members, and the difference increased during the pandemic. The overall visit rate among members without COVID-19 decreased by 5.5% (95% CI −10.3% to −0.5%) in October-December 2020 from 7.37 visits per person-year in January-February 2020 to 6.85 visits per person-year in October-December 2020; inpatient and ED visit rates decreased by 15.6% (95% CI −21.1% to −9.8%) and 28.6% (95% CI −35.0% to −21.5%) from 0.107 and 0.274 visits per person-year in January-February 2020 to 0.087 and 0.189 visits per person-year in October-December 2020, respectively (Table S2 in Multimedia Appendix 2). In July-December 2021, the overall visit rate among members without COVID-19 returned to prepandemic levels, while the inpatient and ED visit rates were still 7.5% (95% CI −9.6% to −5.3%) and 8.0% (95% CI −11.0% to −4.9%) lower, reaching 0.095 and 0.241 visits per person-year, respectively (Table S3 in Multimedia Appendix 2).
There were around 4.5 million enrolled-person-years in KPSC each year during the study period. The use trends overall and by visit type at KPSC were similar to those of the entire study population during 2017-2021 (Figures S1 and S2 in Multimedia Appendix 3). The visit rates from January 2022 to June 2023 remained stable, continuing at the late 2021 levels. The adjusted percent changes of visit rates in July-December 2022 were consistent with the estimates from the main analysis (Table S1 in Multimedia Appendix 3). The overall, inpatient, ED, and combined telehealth and outpatient visit rates among all members returned to prepandemic levels during July-December 2022. The telehealth visit rate in July-December 2022 increased by 92.3% (95% CI 60.3%–130.7%) compared to the prepandemic period. Among members without COVID-19, the inpatient and ED visit rates during July-December 2022 were still 11.1% (95% CI –14.0% to –8.0%) and 10.2% (95% CI –15.3% to –4.9%) lower than prepandemic levels, respectively.

Discussion

Principal Findings

In this large, retrospective cohort study conducted across 8 integrated health care systems with more than 11 million demographically and geographically diverse members, the reductions observed in the early pandemic period persisted into late 2020 for in-person visits, but overall use and counting all types of visits tended to equilibrate because of dramatically increased use of telehealth. Furthermore, our study showed the longer-term impact of the COVID-19 pandemic on health care utilization, especially after COVID-19 restrictions were relaxed. Specifically, during late 2021, inpatient and ED visit rates returned to prepandemic levels and overall use exceed prepandemic levels. The trends observed in the KPSC data closely mirrored those of the entire study population. The trends during the extended period from January 2022 to June 2023 were stable and appeared to be a continuation of the levels observed at the end of 2021. KPSC contributed over one-third of the data.

Telehealth care was promoted and supported through reimbursement during the COVID-19 pandemic [18], which partially resulted in a surge in telehealth visits during the COVID-19 pandemic. Although the rates decreased after an initial rise, they remained doubled in late 2021 compared to prepandemic levels. Telehealth visits increased slowly before the COVID-19 pandemic due to individual and organizational barriers, which have been discussed in previous studies [19]. The COVID-19 pandemic provided an opportunity for health care systems, providers, and patients to adopt this new approach for health care delivery [20,21]. Furthermore, studies have shown that minority and low-income groups could benefit significantly from the increased accessibility of telehealth care by eliminating barriers such as transportation costs and time constraints [22,23]. Despite the fact that telehealth may not be appropriate for certain medical conditions (eg, physical examinations, medical procedures, and emergent treatments) [24] and access to in-person care improved later in the COVID-19 pandemic, telehealth care remained a mainstay of health care delivery, comprising approximately 42% (5.63 visits per person-year) of outpatient visits in late 2021 (2.35 telehealth visits per person-year). The postpandemic use of telehealth is likely to continue growing and evolving to improve patient outcomes, access to care, and convenience of care [25]. These historical trends can help inform projections for the use of telehealth and in-person modalities for care.

COVID-19–related care added substantially to the burden on the health care system [26]. During the early stages of the COVID-19 pandemic, there was competition between the resources allocated to COVID-19–related care and those directed toward non–COVID-19–related medical needs [2]. Both types of care required significant resources, potentially straining the health care system’s capacity. Excluding members with COVID-19 and all their use during the study period did not significantly change the direction of health care utilization trends in this population; however, the visit rates decreased more among members without evidence of COVID-19 during the pandemic, especially in the inpatient and ED settings. In the later pandemic period, inpatient and ED visits among members without evidence of COVID-19 decreased by 7.5% and 8% (from 0.105 and 0.258 visits per person-year in prepandemic to 0.095 and 0.241 visits per person-year in later pandemic), respectively, and these rates did not return to prepandemic levels. While there were differences in baseline use between all members and the members without COVID-19 possibly due to differences in health status and care-seeking behavior, both DiD and ITS analyses accounted for the baseline use within each group. As we adjust to the post–COVID-19 pandemic phase, it is important to investigate the reasons behind this phenomenon and prioritize efforts that address the possible unmet needs of those patients with non–COVID-19 illness.

Knowledge of health care utilization is also important for researchers who conduct studies using observational data during the pandemic period. For example, vaccine safety studies commonly identified outcomes from in-person encounters before the COVID-19 pandemic [12]. Since telehealth use remained high during the pandemic, which contributed to the delivery of health care beyond in-person visits, telehealth visits could be an important data source for identifying adverse events in vaccine safety studies during or after the pandemic period. In addition, with the changes in use over time, secular confounding needs to be considered when designing a study involving the pandemic period and selecting appropriate unbiased comparison groups.

There are several statistical considerations for these analyses. First, this study conducted analyses at the population level. We showed that the demographic characteristics of members enrolled at VSD sites were stable over the study period. Both DiD analysis and ITS analyses compared health care utilization before and during the COVID-19 pandemic among a similar population. Hence, patient-level adjustment was not critical. Second, unmeasured secular confounders (eg, changes in care practice unrelated to COVID-19 and changes in underlying health conditions of the study population) were adjusted for through either the DiD approach or a prepandemic trend estimated from 2017 to 2019 data. Although the latter required an assumption that the trend would have continued through 2020-2021 if there was no COVID-19 pandemic, the assumption
likely held, given the moderate length of the study period. The study also had some limitations. The study aimed to assess the overall impact of the COVID-19 pandemic on health care utilization, including factors such as the population’s fear of exposure in health care settings, policy changes in physical distancing, surges of new COVID-19 cases, and COVID-19 vaccinations. To assess the individual impact of each factor would require proper measurement of these factors and a different study design and analysis. In addition, the study did not assess the impact of the COVID-19 pandemic on subgroups. The impact could differ by age, race or ethnicity, socioeconomic status, and health conditions, as has been shown in previous studies [22,27]. Finally, we included persons enrolled in health plans; therefore, our findings may not be generalizable to the uninsured population.

Conclusions
During the late COVID-19 pandemic period, overall use of health care services exceeded pre-pandemic levels as telehealth services became a mainstay of the health care system. Inpatient and ED visits returned to pre-pandemic levels, although they remained low among members without evidence of COVID-19. Our findings provide valuable information for planning longer term strategic resource allocation for patient care in the post-pandemic period and for designing observational studies involving the pandemic period.

Acknowledgments
This study was supported by the Centers for Disease Control and Prevention through the Vaccine Safety Datalink under contract 75D30122D15429. We thank Abraelle Shirley and Kerresa Morrissette for project management support. The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Data Availability
The data sets generated during this study are not available for public access. Guidelines on how to access Vaccine Safety Datalink data through a sharing program administered by the National Center for Health Statistics Research Data Center are provided on the Vaccine Safety Datalink website and are subject to change.

Conflicts of Interest
LQ reports research support from Moderna, GlaxoSmithKline, and Dynavax for unrelated studies. LSS reports research support from Moderna, GlaxoSmithKline, and Dynavax for unrelated studies. JCN received funding for service on the External Safety Advisory Board for the Moderna COVID-19 vaccine program. JGD reports research support from Janssen for an unrelated study. VH received funding from Pfizer for an unrelated study paid directly to the institution.

Multimedia Appendix 1
Illustration of the periods compared in the difference-in-difference analysis and interrupted time series analysis.

[DOC File, 107 KB - publichealth_v10i1e48159_app1.doc]

Multimedia Appendix 2
Health care utilization among members without COVID-19.

[DOC File, 87 KB - publichealth_v10i1e48159_app2.doc]

Multimedia Appendix 3

[DOC File, 378 KB - publichealth_v10i1e48159_app3.doc]

References


Abbreviations

- DiD: difference-in-difference
- ED: emergency department
- ITS: interrupted time series
- KPSC: Kaiser Permanente Southern California
- RR: rate ratio
- RRR: ratio of rate ratio
- VSD: Vaccine Safety Datalink

©Lei Qian, Lina S Sy, Vennis Hong, Sungching C Glenn, Denison S Ryan, Jennifer C Nelson, Simon J Hambidge, Bradley Crane, Ousseny Zerbo, Malini B DeSilva, Jason M Glanz, James G Donahue, Elizabeth Liles, Jonathan Duffy, Stanley Xu. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 23.01.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Impact of the COVID-19 Pandemic on People Living With Rare Diseases and Their Families: Results of a National Survey

Maurizio Macaluso1,2, MD, DrPH; Marc E Rothenberg2,3, MD, PhD; Thomas Ferkol4, MD; Pierce Kuhnell1, MS; Henry J Kaminski5, MD; David W Kimberlin6, MD; Michael Benatar7, MD, PhD; Mirna Chehade8, MD; The Principal Investigators of the Rare Diseases Clinical Research Network – Cycle 4

1Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
2Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, United States
3Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
4Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
5Department of Neurology and Rehabilitation Medicine, George Washington University, Washington, DC, United States
6Division of Pediatric Infectious Diseases, University of Alabama at Birmingham, Birmingham, AL, United States
7Department of Neurology, University of Miami, Miami, FL, United States
8Mount Sinai Center for Eosinophilic Disorders, Departments of Pediatrics and Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States
9See Acknowledgments

Corresponding Author:
Maurizio Macaluso, MD, DrPH
Division of Biostatistics and Epidemiology
Cincinnati Children's Hospital Medical Center
3333 Burnet Ave - MLC 5041
Cincinnati, OH, 45229
United States
Phone: 1 513 636 2710
Email: maurizio.macaluso@cchmc.org

Abstract

Background: With more than 103 million cases and 1.1 million deaths, the COVID-19 pandemic has had devastating consequences for the health system and the well-being of the entire US population. The Rare Diseases Clinical Research Network funded by the National Institutes of Health was strategically positioned to study the impact of the pandemic on the large, vulnerable population of people living with rare diseases (RDs).

Objective: This study was designed to describe the characteristics of COVID-19 in the RD population, determine whether patient subgroups experienced increased occurrence or severity of infection and whether the pandemic changed RD symptoms and treatment, and understand the broader impact on respondents and their families.

Methods: US residents who had an RD and were <90 years old completed a web-based survey investigating self-reported COVID-19 infection, pandemic-related changes in RD symptoms and medications, access to care, and psychological impact on self and family. We estimated the incidence of self-reported COVID-19 and compared it with that in the US population; evaluated the frequency of COVID-19 symptoms according to self-reported infection; assessed infection duration, complications and need for hospitalization; assessed the influence of the COVID-19 pandemic on RD symptoms and treatment, and whether the pandemic influenced access to care, special food and nutrition, or demand for professional psychological assistance.

Results: Between May 2, 2020, and December 15, 2020, in total, 3413 individuals completed the survey. Most were female (2212/3413, 64.81%), White (3038/3413, 89.01%), and aged ≥25 years (2646/3413, 77.53%). Overall, 80.6% (2751/3413) did not acquire COVID-19, 2.08% (71/3413) acquired it, and 16.58% (566/3413) did not know. Self-reported cases represented an annual incidence rate of 2.2% (95% CI 1.7%-2.8%). COVID-19 cases were more than twice the expected (71 vs 30.3; P <.001). COVID-19 was associated with specific symptoms (loss of taste: odds ratio [OR] 38.9, 95% CI 22.4-67.6, loss of smell: OR 30.6, 95% CI 17.7-53.1) and multiple symptoms (>9 symptoms vs none: OR 82.5, 95% CI 29-234 and 5-9: OR 44.8, 95% CI 18.7-107). Median symptom duration was 16 (IQR 9-30) days. Hospitalization (7/71, 10%) and ventilator support (4/71, 6%) were uncommon. Respondents who acquired COVID-19 reported increased occurrence and severity of RD symptoms and use or dosage of select...
medications; those who did not acquire COVID-19 reported decreased occurrence and severity of RD symptoms and use of medications; those who did not know had an intermediate pattern. The pandemic made it difficult to access care, receive treatment, get hospitalized, and caused mood changes for respondents and their families.

**Conclusions:** Self-reported COVID-19 was more frequent than expected and was associated with increased prevalence and severity of RD symptoms and greater use of medications. The pandemic negatively affected access to care and caused mood changes in the respondents and family members. Continued surveillance is necessary.

*(JMIR Public Health Surveill 2024;10:e48430) doi:10.2196/48430*

**KEYWORDS**

rare diseases; rare; chronic; COVID-19 infection; cross-sectional survey; access to care; changes in symptoms and use of medications; psychological impact on self and family; access; accessibility; cross-sectional; national; nationwide; survey; surveys; COVID-19; SARS-CoV-2; coronavirus; comorbid; comorbidity; vulnerable

**Introduction**

**Background**

The COVID-19 pandemic, caused by the SARS-CoV-2, emerged in the fall of 2019 in Wuhan, China, and rapidly spread throughout the world. In March 2020, the World Health Organization (WHO) declared the outbreak to be a pandemic, and as of August 30, 2023, the WHO has reported >770 million confirmed cases and >6.9 million confirmed deaths attributable to the disease [1,2]. More than 103 million cases and 1.1 million deaths from COVID-19 have occurred in the United States [1,3], with devastating consequences for the health system [4,5] and the well-being of the entire population [6].

The *Journal of Medical Internet Research* and affiliated JMIR publications, including *JMIR Public Health and Surveillance*, have extensively documented the global impact of the COVID-19 pandemic: a total of 99 articles have reported the results of web-based surveys and analyses of internet browsing patterns and social media postings (search terms “Rare disease” and “COVID-19,” run on July 7, 2023). For example, during the early months of the pandemic, an artificial intelligence–based analysis of >902,000 tweets was used to identify sentiments regarding COVID-19 in the United States [7], and large-scale web-based surveys were conducted in China [8,9], South Korea [10], Russia [11], and Spain [12]. Several surveys have documented a negative impact on anxiety and depression, self-reported well-being, and quality of life in the general population [11-16] and in vulnerable groups such as pregnant women [17-19], university students [20-22], older people [23,24], and health care workers [25,26].

Despite the extensive JMIR literature on the impact of the pandemic, only a single study described the experience of people living with rare diseases (RDs), which documented the impact of the pandemic on conversation patterns in the Reddit platform (subreddit r/CysticFibrosis) [27]. Indeed, the literature on the impact of the pandemic on people living with RDs is sparse even 3 years into the pandemic. A PubMed search (search terms “Rare disease” and “COVID-19,” run on July 6, 2023) identified 148 potential matches, but review of the abstracts indicated that most publications were case reports, discussion papers, or other reports that do not entail data collection from people living with RDs: only 25 articles reported results of cross-sectional surveys or longitudinal studies of patients with RD or their caregivers [28-52]. Large studies (≥50 participants) in Asia [29,37,40], Europe and the United Kingdom [30,33,36,41,49,51], the United States [45,52], and Brazil [43] reported variable impact of the pandemic, including increased risk of acquiring COVID-19 [33,37,39], increased COVID-19–related mortality [33,37], increased anxiety and depression and other psychological problems [29,30,36,41,51,52], low quality of life [41,51], difficulties with access to care [36,40,43,45,52], and difficulties complying with public health recommendations such as masking [49].

According to US laws, an RD is defined as a condition affecting <200,000 individuals at a given time [53,54]. The Genetic and Rare Disease information system of the National Center for Advancing Translational Sciences lists >10,000 RDs [55,56]. Although each condition is rare, the aggregate prevalence may amount to 30 million [55]. RDs are often difficult to diagnose [57,58], most often lack specific treatments [59], and place a significant burden on families and society [60]. Some people with RDs have primary respiratory manifestations [61], immunocompromised states [62], and chronic comorbidities including intellectual disabilities [63]. People with RDs report a lower quality of life than the general population [64]; they are vulnerable and dependent on the health care system [65].

The Rare Diseases Clinical Research Network (RDCRN) is funded by 10 of the National Institutes of Health (NIH) to promote progress in the diagnosis and treatment of RDs. The RDCRN comprises 20 Rare Disease Clinical Research Consortia and a Data Management and Coordinating Center, >200 research sites worldwide, and many patient advocacy groups that ensure that patients’ voices are heard. The RDCRN has been active for 2 decades, and >40,000 people have participated in its studies [66].

**Objectives**

The RDCRN was strategically well placed to study the impact of the pandemic among people living with RDs in the United States. We designed a cross-sectional survey to describe the characteristics of the COVID-19 infection in this population, to determine whether patient subgroups were affected more frequently or experienced increased severity, to determine whether the pandemic changed RD symptom and treatment patterns, and to understand the broader impact of the pandemic on people living with an RD and their families.
Methods

Overview
We designed a cross-sectional survey to describe the characteristics of the COVID-19 infection in the population with RD. People who lived in the United States, had an RD, and were aged <90 years (including children of any age) were eligible for the survey. There were no explicit exclusion criteria beyond failure to meet the eligibility criteria. The survey was advertised through the internet, social media, an NIH press release, and outreach by patient advocacy groups. Interested individuals were asked to access a dedicated web page and complete a REDCap (Research Electronic Data Capture; Vanderbilt University [67]) survey instrument on the web (Multimedia Appendix 1).

The data collection instrument was designed ad hoc by the authors in spring 2020, with contributions from all the principal investigators of the RDCRN consortia (included as a group author as “The Principal Investigators of the Rare Diseases Clinical Research Network - Cycle 4” refer to the collaborators list), based on their expert knowledge of RDs. The instrument comprises 8 sections:

- Section 1 includes eligibility and preliminary information criteria on whether the patient or a surrogate is interested in participating, whether the target respondent lives in the United States and has an RD, and whether the respondent has participated in RDCRN research;
- Section 2 explains the study and asks for consent offering three options: (1) respond to the survey, provide contact information, and allow contact for follow-up and opportunities for participating in RD research; (2) allow linkage of the survey results with data about the same individual provided as part of RDCRN studies; and (3) respond to the survey without any identifiers (one-time, anonymous response);
- Section 3 is restricted to individuals who agree to provide identifiers and includes name, date of birth, address, telephone number, and email address to allow future contacts;
- Section 4 includes sociodemographic variables (state of residence, age, sex, race and ethnicity), RD diagnosis including symptoms and comorbidities during the months preceding the COVID-19 epidemic in the United States, treatments received during the same period, smoking habits, and use of tetrahydrocannabinol-containing products and psychoactive drugs. It also includes questions about COVID-19–related symptoms that the respondent may have experienced before the beginning of the pandemic and any changes in RD symptoms and treatment experienced after the beginning of the pandemic in the United States (March 2020);
- Section 5 ascertains whether the respondent had a positive COVID-19 test or a clinical diagnosis of COVID-19 respiratory syndrome; and
- Sections 6 to 8 separately address respondents who did not acquire COVID-19 (section 6), acquired it (section 7), or did not know (section 8).

All these sections include a checklist of symptoms compatible with COVID-19; ask whether the pandemic interfered with access to medical care, RD treatment, and special food related to RD management; whether stay-at-home orders affected their mood or behavior; and whether the respondent or family members had to seek professional help to cope with stress and anxiety. A final question asks if the target respondent died during the interval leading to the survey. In addition to the abovementioned common questions, section 7 asks additional questions about the course of the COVID-19 infection, the need for hospitalization and special treatment (assisted ventilation), the duration of the illness, and whether the respondent received investigational drugs to treat COVID-19.

Study Size Considerations
We aimed at enrolling 5000 participants based on a precision analysis, rather than a power analysis, because the survey was not designed to test specific hypotheses. At the time of planning the survey, the WHO estimates suggested that the prevalence of COVID-19 infection may be between 1% and 2%. Thus, if the patients with RD experienced the same risk as the general population, the target sample size would provide information on approximately 50-100 COVID-19 cases. As the 95% CI of 50 observed events was 37-66 based on the Poisson distribution, our estimate of the overall prevalence of COVID-19 infection among respondents would be very precise. The precision of survey measures of more common characteristics and experiences among respondents (eg, changes in access to treatment) would be much higher. Furthermore, given a sample size of 50 cases, the 95% CI of a simple mean would be between +0.25 and –0.25 SD units from the mean, affording adequate precision of measures such as the average duration of hospitalization, and estimates of rates for any categorization of the case group into 5-10 subgroups would also yield adequate precision in describing the variation in risk across population subgroups.

Ethical Considerations
As the survey was conducted on the web, the REDCap survey instrument verified the eligibility of a prospective participant, provided an explanation of the objectives and process of data collection, verified consent, and gathered the required information. First, the instrument assessed the eligibility of the intended participant (ie, the person was living with an RD, was living in the United States, and was aged <90 years), automatically terminating data collection from ineligible individuals; ascertained whether the intended participant or a surrogate would be completing the survey; and ascertained whether the potential participant had participated in the RDCRN research. Next, it provided a comprehensive explanation of the study objectives and process and asked whether (1) the participant agreed to respond to the survey, provide contact information, and allow future contact for follow-up interviews and opportunities to participate in RD research; (2) if agreeing to (1), the participant further agreed to allow linkage of the survey results with data about the same individual provided as part of the RDCRN studies; or (3) the participant agreed to respond to the survey without any identifiers (1-time, anonymous response). The explanatory language clarified that...
no adverse events were expected except for a possible breach of confidentiality for participants who agreed to provide identifiable information and that the study team intended to minimize the risk of a confidentiality breach by keeping identifiers and survey responses in separate files and by presenting the study results in a format that would not allow the reidentification of individual patients with RD. No compensation was offered for participating in this study. The institutional review board (IRB) at Cincinnati Children’s Hospital Medical Center reviewed and approved the research protocol, all recruitment materials, and the survey instrument, including the language explaining the survey and asking for consent. The IRB granted a waiver of documentation of informed consent (IRB ID 2020-0299).

Data Analysis
In this report, we aimed to address the following outstanding questions in the field:

1. How frequent was the self-reported laboratory or clinical diagnosis of COVID-19 among people living with RDs? Did this differ from the expectation based on population rates?
2. How frequent were COVID-19–related symptoms according to self-reported COVID-19 status?
3. What were the characteristics of COVID-19 infection (duration of the infection, need for hospital admission, and complications of the infection among COVID-19 cases)?
4. Did COVID-19 status influence changes in the frequency and intensity of RD-related symptoms and in the frequency of RD-related treatment?
5. Did the COVID-19 epidemic change access to care, access to special food and nutrition, demand for professional assistance to address stress and coping among people who live with RD, and was the impact affected by COVID-19 status?

We used simple descriptive statistics (frequencies and percentages for count variables and means, medians, and ranges for continuous variables) and provided 95% CIs as a measure of the precision of the estimates for survey results. Statistical hypothesis testing was used for select analyses.

To identify participants who acquired COVID-19, we used the answer to the simple survey question, “Did you acquire COVID-19?” which allowed the answers yes, no, or do not know (Multimedia Appendix 1, survey instrument item 5.1). The 3-category answer was a pragmatic approach to identifying cases of infection because at the time we conducted the survey, access to laboratory tests remained limited.

We estimated the denominator at risk of infection considering each respondent at risk every day from the beginning of the pandemic to the day they reported acquiring COVID-19 (the outcome event) or the day they filled out the survey, whichever was earlier, as is done in retrospective cohort studies. We computed exact 95% CIs for the number of cases and the corresponding incidence rate under the assumption that the number of cases follows the Poisson distribution. We compared the monthly number of self-reported COVID-19 cases with the number expected based on data reported by the New York Times [68], computed by multiplying the national monthly infection rate by the number of respondents who were at risk of acquiring a COVID-19 infection during that month. The cases reported in the New York Times database are a mix of laboratory-confirmed cases and cases meeting the state-determined diagnostic criteria. We present the observed and expected cases in a graph and test the null hypothesis that the number of observed cases equals the expected value using an exact Poisson distribution method.

For all respondents before the beginning of the pandemic (Multimedia Appendix 1, survey instrument item 4.9) and in separate sections for the respondents who answered yes, no, or do not know to the question “Did you acquire COVID-19?” (Multimedia Appendix 1, survey instrument items 6.1, 7.3, and 8.3), the instrument asked the question, “Did you have symptoms related to COVID-19? (answer options: Yes/No),” offering a checklist of 20 symptoms (new or increased cough, fever >38.0°C, new or increased shortness of breath, sore throat, stuffy nose, runny nose, chest pain, sneezing, wheezing, headache, muscle aches, loss of taste, loss of smell, conjunctivitis or pink eye, confusion, seizures, weakness, and other).

For 65 RD-related symptoms grouped into 18 categories, the instrument asked the questions “Please check if you had any of the following symptoms before the COVID-19 pandemic began in the USA. Think about your symptoms in January-February 2020. Check all that apply” (answer options: yes or no) and “Did anything change after the beginning of the pandemic (March 2020)?” (If the previous answer was “No,” answer options: yes or no; if the previous answer was “Yes,” answer options: symptom absent, less severe, same severity, or more severe; Multimedia Appendix 1, survey instrument item 4.6).

For 91 medications and treatments grouped into 15 categories, the instrument asked the questions “Please check if you used any of the treatments before the COVID-19 pandemic began in the USA. Think about the medications you took or treatments you routinely received in January-February 2020. Check all that apply” (answer options: yes or no) and “Did anything change after the beginning of the pandemic (March 2020)?” (If the previous answer was “No,” answer options: yes or no; if the previous answer was “Yes,” answer options: treatment not used, lower dosage, same dosage, or higher dosage; Multimedia Appendix 1, survey instrument item 4.7).

We computed odds ratios (ORs) and their 95% CIs to illustrate the strength of the association between self-reported infection and specific COVID-19–related symptoms as well as the number of symptoms.

To assess changes in the prevalence or intensity of a specific COVID-19 symptom, we compared the number of patients who reported the symptom appearing during the pandemic (ie, the symptom was absent before the beginning of the pandemic but was present afterward) with the number of patients who reported the symptom disappearing (ie, the symptom was present before the beginning of the pandemic but was not present afterward) and used an exact binomial test of the null hypothesis that the 2 counts were equal (ie, the parameter of the binomial distribution was 0.5).
To assess changes in the prevalence and intensity of a specific RD symptom, we compared the number of patients who reported the symptom appearing or increasing in intensity during the pandemic with the number of patients who reported the symptom disappearing or decreasing in intensity and used the exact binomial test of the null hypothesis that the 2 counts were equal. We used the same logic and statistical test to assess the changes in the prevalence of the use or dosage of a specific medication. We used heat maps to summarize the changes in COVID-19–associated symptoms, RD symptoms, and RD treatment before and after the beginning of the pandemic.

The instrument asked the question, “Do you have other diseases or complications related to the rare disease? Check all that apply,” offering a list of 26 comorbidity categories (Multimedia Appendix 1, survey instrument item 4.8).

We computed ORs and their 95% CIs to illustrate the strength of the association between self-reported infection and specific comorbidities, reporting the exact P value for the test of the null hypothesis of no association (ie, OR=1). We regarded a finding as statistically significant if the statistical testing of a specified null hypothesis yielded a P value <.05.

Results

Enrollment of participants for this study began on May 2, 2020, and ended on December 15, 2020; a total of 3413 respondents completed the survey. Most of the surveys (2643/3413, 77.44%) were completed between May and July (Figure S1 in Multimedia Appendix 1). Respondents approximately represented the geographic distribution of the US population (Table S1 in Multimedia Appendix 1): 25.76% (870/3377) were from the Midwest (21% expected) [69], 18.86% (637/3377) from the Northeast (18% expected), 33.43% (1129/3377) from the South (38% expected), and 21.94% (741/3377) from the West (24% expected). The diagnoses studied by the RDCRN accounted for 61.27% (2091/3413); and the most common diagnoses were myasthenia gravis (594/2091, 28.41%), amyotrophic lateral sclerosis (289/2091, 13.82%), eosinophilic esophagitis or eosinophilic gastrointestinal disease below the esophagus (203/2091, 9.71%), mitochondrial disease (174/2091, 8.32%), cystic fibrosis (87/2091, 4.16%), and primary ciliary dyskinesia (67/2091, 3.2%).

Two-thirds (2212/3413, 64.81%) of the respondents were female. Only 1.93% (66/3413) of the respondents reported being exclusively Black, 6.36% (217/3413) of the respondents reported other races or >1 race, and 89.01% (3038/3413) of the respondents were White (Table S2 in Multimedia Appendix 1). Approximately 5% (156/3413, 4.57%) reported Hispanic or Latino ethnicity, whereas most (2844/3413, 83.33%) of the respondents did not report ethnicity. Most of the respondents (2646/3413, 77.53%) were adults aged ≥25 years (Table S2 in Multimedia Appendix 1). The distribution by sex, race and ethnicity, and age was similar across disease categories, with the exception that most male respondents with eosinophilic esophageal or gastrointestinal diseases were aged <25 years. More detailed information on the demographic characteristics of the respondents by RD category is provided in Table S3 in Multimedia Appendix 1.

Overall, 80.6% (2751/3413) of the respondents reported that they had not acquired the COVID-19 infection, 2.08% (71/3413) of the respondents reported having acquired it, and 16.58% (566/3413) of the respondents did not know. Of the 71 respondents who reported acquiring the infection, 51 (72%) reported having a positive laboratory test. By contrast, only 15.2% (86/566) of the respondents who did not know reported having been tested (no positive test). We estimated that the self-reported cases represent the incidence over a total of 3219 person-years at risk, with an average annual incidence rate of 2.2% (95% CI 1.7%-2.8%).

Self-reported COVID-19 cases were twice as many as would be expected on the basis of the population incidence rates (71 vs 30.3; P < .001) and were concentrated in March to April 2020 (37 vs 11 expected) and in October to December 2020 (15 vs 5 expected; Figure 1), whereas self-reported cases were only slightly higher than expected in May to September 2020 (19 vs 14.2).

Approximately all respondents who acquired COVID-19 had symptoms (66/71, 93%). For 15 (83%) of the 18 symptoms, they reported experiencing the symptom more often after the beginning of the pandemic than before. Respondents who reported not acquiring COVID-19 had symptoms less often after the beginning of the pandemic than before for 16 (89%) of the 18 symptoms. The respondents who did not know displayed an intermediate pattern (Figure 2).

Specific COVID-19–related symptoms were strongly associated with self-reported infection: loss of taste (OR 38.9, 95% CI 22.4-67.6) or smell (OR 30.6, 95% CI 17.7-53.1), high fever (OR 14.5, 95% CI 8.7-24.1), and confusion (OR 12.7, 95% CI 6.80-23.7) were the strongest correlates of self-reported infection (Table 1).

Reporting multiple COVID-19–related symptoms was strongly associated with acquiring COVID-19 (>9 symptoms vs none: OR 82.5, 95% CI 29-234 and 5-9 symptoms: OR 44.8, 95% CI 8.7-241). Those who did not know were also more likely to report COVID-19–related symptoms and multiple symptoms than those who reported not acquiring the infection (Table 1). Respondents who acquired COVID-19 had demographic characteristics like others; most of the respondents were female (53/71, 75%), non-Hispanic White (59/71, 83%), and representative of the distribution of RD diagnoses reported by all respondents (results not shown in detail). The survey investigated 36 possible RD comorbidities, and 58.54% (1998/3413) of the respondents reported ≥1 comorbidities. Those who acquired COVID-19 (OR 1.60, 95% CI 1.00-2.50) and those who did not know (OR 1.40, 95% CI 1.20-1.70) were more likely to report RD comorbidities than those who did not acquire the infection. Individual associations between RD comorbidities and self-reported COVID-19 infection were weak (Table S4 in Multimedia Appendix 1).

Respondents who acquired the infection reported a significantly increased occurrence or severity for 35% (23/65) of the RD-related symptoms from before the beginning of the COVID-19 pandemic to survey completion, no change for 52% (34/65) of the symptoms, and no significant decrease for any symptoms. Those who did not know reported significantly
increased occurrence or severity for 37% (24/65) of the symptoms, no change for 51% (33/65) of the symptoms, and a significant decrease for 8% (5/65) of the symptoms. Those who did not acquire COVID-19 had increased occurrence or severity for 17% (11/65) of the symptoms, no change in 49% (32/65) of the symptoms, and significant decreases for 31% (20/65) of the symptoms (Figure 3).

Respondents who acquired the infection reported significantly increased frequency or dosage for 5% (5/91) of the medications (azithromycin, inhaled albuterol, inhaled glucocorticoids, ibuprofen, and aspirin), whereas the frequency or dosage of other medications did not change significantly. Those who did not know did not have significant changes for most medications but had a significant decrease in frequency or dose for 3% (3/91) of the medications. Those who did not acquire COVID-19 reported no significant change for 51% (46/91) of the medications and significant decreases in use or dosage for 34% (31/91) of the medications (Figure 4).

Among those who reported acquiring the infection, the COVID-19 illness experience was not severe; symptoms lasting on average <30 days; hospitalization (7/71, 10%), and mechanical ventilator support (4/71, 6%) were relatively uncommon. However, 1 respondent in this group died, and many reported that either their RD complicated the COVID-19 illness (30/71, 44%) or that the infection worsened their baseline symptoms (39/71, 55%; Table 2).

The pandemic affected access to health care: 61% (42/69) of the respondents who acquired COVID-19 and 75.02% (2382/3175) of the other respondents experienced delays in obtaining an appointment, their appointment was done using telemedicine, or their appointment was postponed (Table 3).

In addition, access to treatment for the RD was affected. On the basis of nonmissing values, 25.99% (715/2751) of the respondents who did not acquire COVID-19, 32.7% (185/566) of the respondents who did not know, and 41% (29/71) of the respondents who acquired COVID-19 experienced delays in obtaining treatment or their therapies were interrupted during the pandemic. Access to food was uninterrupted for most of the respondents, but as many as 26% (18/70) of those who acquired COVID-19 had delays or an interruption in the supply of specialized diets (Table 3). The most pronounced disruption was in access to specialized treatment such as physical or speech therapy, with approximately 56% (32/57) of the respondents experiencing an interruption of treatment. Respondents who acquired COVID-19 reported more frequently (11/68, 16%) experiencing health events that would ordinarily require hospitalization but were managed without admission. Moreover, about 1 (13/67, 19%) of out 5 respondents reported that stay-at-home orders affected their mood or behavior in a way that required medical attention, and most of the respondents (480/2717, 17.67% to 28/71, 39%) noted that they or immediate family members sought professional support to cope with stress or anxiety, with the largest proportion (28/71, 39%) among those who acquired COVID-19 (Table 3). For all but 2 items in Table 3, the variation in responses across the 3 groups defined by self-reported COVID-19 status was statistically significant. Six deaths were reported, but 5 were among the respondents who did not acquire COVID-19.

Figure 1. National survey of the impact of the COVID-19 pandemic on people with rare diseases (May 2, 2020, to December 15, 2020; N=3413). Number of self-reported COVID-19 cases by month, compared with the numbers expected on the basis of the monthly incidence rates reported by the New York Times for the entire US population.
Figure 2. National survey of the impact of the COVID-19 pandemic on people with rare diseases (RDs; May 2, 2020, to December 15, 2020; N=3413). Heat map of pandemic-associated changes in the prevalence of COVID-19 symptoms. Statistical significance of item-specific changes based on an exact binomial test of the null hypothesis that positive changes (ie, symptom appearing or increasing in intensity and medication starting or dosage increasing during the pandemic) were equal in number to negative changes (ie, symptom disappearing or decreasing in intensity and medication discontinued or dosage decreasing during the pandemic). Respondents are categorized according to their answer to the question “Did you acquire COVID-19?”. Blue boxes indicate statistically significant decrease; grey boxes indicate statistically non-significant change; red boxes indicate statistically significant increase.
Table 1. National survey of the impact of COVID-19 on people with rare diseases (May 2, 2020, to December 15, 2020; N=3413). COVID-19–related symptoms reported by survey respondents and association with answers to the question “Did you acquire COVID-19?” Those who responded “No” are the reference group.

<table>
<thead>
<tr>
<th>Specific symptom</th>
<th>Values, n (%)</th>
<th>OR(^a) (95% CI)</th>
<th>Values, n (%)</th>
<th>OR (95% CI)</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New or increased cough</td>
<td>37 (52)</td>
<td>10.7 (6.60-17.4)</td>
<td>201 (35.5)</td>
<td>5.40 (4.40-6.70)</td>
<td>253 (9.2)</td>
</tr>
<tr>
<td>Fever &gt; 38.0 °C</td>
<td>29 (41)</td>
<td>14.5 (8.70-24.1)</td>
<td>145 (25.6)</td>
<td>7.20 (5.60-9.40)</td>
<td>125 (4.54)</td>
</tr>
<tr>
<td>New or increased shortness of breath</td>
<td>33 (46)</td>
<td>12.0 (7.30-19.5)</td>
<td>151 (26.7)</td>
<td>5.00 (4.00-6.40)</td>
<td>186 (6.76)</td>
</tr>
<tr>
<td>Sore throat</td>
<td>28 (39)</td>
<td>6.20 (3.80-10.2)</td>
<td>190 (33.6)</td>
<td>4.80 (3.90-6.00)</td>
<td>260 (9.45)</td>
</tr>
<tr>
<td>Stuffy nose</td>
<td>23 (32)</td>
<td>2.90 (1.80-4.90)</td>
<td>157 (27.7)</td>
<td>2.30 (1.90-2.90)</td>
<td>388 (14.10)</td>
</tr>
<tr>
<td>Runny nose</td>
<td>18 (25)</td>
<td>2.20 (1.30-3.80)</td>
<td>155 (27.4)</td>
<td>2.40 (2.00-3.00)</td>
<td>367 (13.34)</td>
</tr>
<tr>
<td>Chest pain</td>
<td>26 (37)</td>
<td>11.8 (7.10-19.8)</td>
<td>89 (15.7)</td>
<td>3.80 (2.90-5.10)</td>
<td>128 (4.65)</td>
</tr>
<tr>
<td>Sneezing</td>
<td>6 (8)</td>
<td>1.00 (0.40-2.40)</td>
<td>97 (17.1)</td>
<td>2.30 (1.80-3.00)</td>
<td>225 (8.18)</td>
</tr>
<tr>
<td>Wheezing</td>
<td>15 (21)</td>
<td>4.10 (2.30-7.50)</td>
<td>98 (17.3)</td>
<td>3.20 (2.50-4.20)</td>
<td>167 (6.07)</td>
</tr>
<tr>
<td>Headache</td>
<td>46 (65)</td>
<td>10.1 (6.10-16.6)</td>
<td>222 (39.2)</td>
<td>3.50 (2.90-4.30)</td>
<td>425 (15.45)</td>
</tr>
<tr>
<td>Muscle aches</td>
<td>41 (58)</td>
<td>8.00 (5.00-13.0)</td>
<td>196 (34.6)</td>
<td>3.10 (2.50-3.80)</td>
<td>400 (14.54)</td>
</tr>
<tr>
<td>Loss of taste</td>
<td>29 (41)</td>
<td>38.9 (22.4-67.6)</td>
<td>50 (8.8)</td>
<td>5.50 (3.60-8.20)</td>
<td>48 (1.74)</td>
</tr>
<tr>
<td>Loss of smell</td>
<td>27 (38)</td>
<td>30.6 (17.7-53.1)</td>
<td>50 (8.8)</td>
<td>4.80 (3.30-7.20)</td>
<td>54 (1.96)</td>
</tr>
<tr>
<td>Conjunctivitis or pink eye</td>
<td>5 (7)</td>
<td>7.10 (2.70-18.9)</td>
<td>35 (6.2)</td>
<td>6.20 (3.70-10.2)</td>
<td>29 (1.05)</td>
</tr>
<tr>
<td>Confusion</td>
<td>15 (21)</td>
<td>12.7 (6.80-23.7)</td>
<td>45 (8.0)</td>
<td>4.10 (2.70-6.10)</td>
<td>57 (2.07)</td>
</tr>
<tr>
<td>Seizures</td>
<td>1 (1)</td>
<td>0.90 (0.10-6.30)</td>
<td>8 (1.4)</td>
<td>0.90 (0.40-1.80)</td>
<td>45 (1.64)</td>
</tr>
<tr>
<td>Weakness</td>
<td>33 (46)</td>
<td>6.30 (3.90-10.2)</td>
<td>152 (26.9)</td>
<td>2.70 (2.10-3.30)</td>
<td>333 (12.1)</td>
</tr>
<tr>
<td>Other</td>
<td>20 (28)</td>
<td>10.6 (6.10-18.5)</td>
<td>50 (8.8)</td>
<td>2.60 (1.80-3.70)</td>
<td>98 (3.56)</td>
</tr>
</tbody>
</table>

Number of symptoms reported

<table>
<thead>
<tr>
<th>Number of symptoms reported</th>
<th>Values, n (%)</th>
<th>OR(^a) (95% CI)</th>
<th>Values, n (%)</th>
<th>OR (95% CI)</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>6 (8)</td>
<td>4.00 (1.30-12.5)</td>
<td>24 (4.2)</td>
<td>0.40 (0.20-0.60)</td>
<td>435 (15.81)</td>
</tr>
<tr>
<td>3-4</td>
<td>11 (15)</td>
<td>11.3 (4.10-30.8)</td>
<td>59 (10.4)</td>
<td>1.40 (1.00-1.90)</td>
<td>284 (10.32)</td>
</tr>
<tr>
<td>5-9</td>
<td>37 (52)</td>
<td>44.8 (18.7-107)</td>
<td>167 (29.5)</td>
<td>4.80 (3.80-6.00)</td>
<td>241 (8.76)</td>
</tr>
<tr>
<td>&gt;9</td>
<td>11 (15)</td>
<td>82.4 (29.0-234)</td>
<td>61 (10.8)</td>
<td>10.7 (7.00-16.4)</td>
<td>39 (1.42)</td>
</tr>
</tbody>
</table>

\(^a\) OR: odds ratio.
Figure 3. National survey of the impact of the COVID-19 pandemic on people with rare diseases (May 2, 2020, to December 15, 2020, N=3413). Heatmap of pandemic-associated changes in the prevalence and intensity of RD-associated symptoms. Statistical significance of symptom-specific changes based on an exact binomial test of the null hypothesis that positive changes (ie, symptom appearing or increasing in intensity during the pandemic) were equal in number to negative changes (ie, symptom disappearing or decreasing in intensity during the pandemic). Respondents are categorized according to their answer to the question “Did you acquire COVID-19?”. Blue boxes indicate statistically significant decrease; grey boxes indicate statistically non-significant change; red boxes indicate statistically significant increase; white boxes indicate inadequate data.
Figure 4. National survey of the impact of the COVID-19 pandemic on people with rare diseases (May 2, 2020, to December 15, 2020, N=3413). Heatmap of pandemic-associated changes in use and dosage of medications. Statistical significance of item-specific changes based on an exact binomial test of the null hypothesis that positive changes (ie, medication starting or dosage increasing during the pandemic) were equal in number to negative changes (ie, medication discontinued or dosage decreasing during the pandemic). Respondents are categorized according to their answer to the question “Did you acquire COVID-19?”. Blue boxes indicate statistically significant decrease; grey boxes indicate statistically non-significant change; red boxes indicate statistically significant increase; white boxes indicate inadequate data.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex: female, n (%)</td>
<td>53 (75)</td>
</tr>
<tr>
<td>Age (y), mean (SD; range)</td>
<td>44 (21; 1-82)</td>
</tr>
<tr>
<td>Race and ethnicity: White, non-Hispanic or Latino, n (%)</td>
<td>59 (83)</td>
</tr>
<tr>
<td>Duration of symptoms (days), median (IQR)</td>
<td>16 (9-30)</td>
</tr>
<tr>
<td>Symptoms persisting through the survey date, n (%)</td>
<td>24 (37)</td>
</tr>
<tr>
<td>Sought care at an emergency department or urgent care center, n (%)</td>
<td>28 (39)</td>
</tr>
<tr>
<td>Was admitted to a hospital, n (%)</td>
<td>7 (10)</td>
</tr>
<tr>
<td>Required supplemental oxygen, n (%)</td>
<td>12 (17)</td>
</tr>
<tr>
<td>Required intubation and mechanical ventilation, n (%)</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Received investigational drugs or participated in a clinical trial, n (%)</td>
<td>9 (13)</td>
</tr>
<tr>
<td>Rare disease complicated the COVID-19 illness, n (%)</td>
<td>30 (44)</td>
</tr>
<tr>
<td>COVID-19 illness worsened the rare disease symptoms, n (%)</td>
<td>39 (55)</td>
</tr>
</tbody>
</table>
Table 3. National survey of the impact of COVID-19 on people with rare diseases (May 2, 2020, to December 15, 2020; N=3413). Responses to survey items addressing the impact of the pandemic on survey respondents and their families, according to the answer to the question “Did you acquire COVID-19?”

<table>
<thead>
<tr>
<th>Survey item and responses</th>
<th>Acquired COVID-19 (n=71), n (%)</th>
<th>Did not know (n=566), n (%)</th>
<th>Did not acquire COVID-19 (n=2751), n (%)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Were you able to continue seeing your health care provider?</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes, without problems</td>
<td>27 (39.13)</td>
<td>130 (23.81)</td>
<td>663 (25.22)</td>
<td>.10</td>
</tr>
<tr>
<td>Yes, but experienced delays in obtaining appointment</td>
<td>3 (4.35)</td>
<td>32 (5.86)</td>
<td>157 (5.97)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes, but my appointment was done via telemedicine</td>
<td>25 (36.23)</td>
<td>241 (44.14)</td>
<td>1214 (46.18)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No, appointment was put on hold</td>
<td>14 (20.29)</td>
<td>143 (26.19)</td>
<td>595 (22.63)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Were you able to continue your treatment?</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes, without problems</td>
<td>39 (57.35)</td>
<td>352 (65.55)</td>
<td>1886 (72.51)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes, but experienced delays in obtaining treatment</td>
<td>8 (11.76)</td>
<td>82 (15.27)</td>
<td>319 (12.26)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No, treatment was interrupted</td>
<td>21 (30.88)</td>
<td>103 (19.18)</td>
<td>396 (15.22)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Were you able to maintain your diet?</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes, without problems</td>
<td>52 (74.29)</td>
<td>417 (76.51)</td>
<td>2231 (84.48)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes, but experienced delays in obtaining treatment</td>
<td>11 (15.71)</td>
<td>96 (17.61)</td>
<td>307 (11.62)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No, supply of needed food was interrupted and my diet suffered from it</td>
<td>7 (10)</td>
<td>32 (5.87)</td>
<td>103 (3.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Were you able to continue specialized treatment?</strong></td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Yes, without problems</td>
<td>31 (54.39)</td>
<td>250 (56.56)</td>
<td>1082 (51.3)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes, but experienced delays in obtaining treatment</td>
<td>1 (1.75)</td>
<td>26 (5.88)</td>
<td>179 (8.49)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No, treatment was interrupted</td>
<td>25 (43.86)</td>
<td>166 (37.56)</td>
<td>848 (40.21)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Did you experience a medical event for which you would ordinarily be hospitalized, but because of COVID-19 you were managed without hospitalization?</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>52 (76.47)</td>
<td>468 (84.02)</td>
<td>2476 (92.84)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>11 (16.18)</td>
<td>56 (10.05)</td>
<td>154 (5.77)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (7.35)</td>
<td>33 (5.92)</td>
<td>37 (1.39)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Have stay-at-home orders in your area affected your mood or behavior in a way that requires medical attention?</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>48 (71.64)</td>
<td>414 (73.53)</td>
<td>2193 (81.92)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>13 (19.4)</td>
<td>122 (21.67)</td>
<td>399 (14.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Unknown</td>
<td>6 (8.96)</td>
<td>27 (4.8)</td>
<td>85 (3.18)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Have you or members of your family sought professional support coping with stress or anxiety because of the COVID-19 pandemic?</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>43 (60.56)</td>
<td>422 (74.82)</td>
<td>2237 (82.33)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Yes</td>
<td>28 (39.44)</td>
<td>142 (25.18)</td>
<td>480 (17.67)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>Test of the null hypothesis that proportions do not vary across the 3 groups. Percentages are reported based on nonmissing values. Missing values are not reported for readability.

**Discussion**

**Principal Findings**

This study reports results from the largest survey assessing the impact of the COVID-19 pandemic on people living with RDs; >3000 individuals representing a wide range of RDs participated in the study. Women, adults, and White non-Hispanic individuals were overrepresented. Recruitment and data collection procedures may have discouraged the participation of minority groups, although the causes of their underrepresentation are probably broader. For instance, the racial and ethnic profile of the respondents living with...
myasthenia gravis was similar to that of the participants of a national registry [31], but it is possible that the registry also underrepresented minority groups. Similarly, in the National Amyotrophic Lateral Sclerosis Registry, of the estimated 25,000 prevalent cases in 2017, 64.97% (16,127/24,821) were White, 5.92% (1469/24,821) were Black. [70]. In 2012 pediatric Medicaid records, 59.3% (2872/4836) of the patients with eosinophilic esophagitis were White and 15.47% (748/4836) were Black, whereas race was missing for 11.81% (571/4836) of the patients [71].

Our survey was widely advertised and relied on outreach by patient advocacy groups. Participation of >1300 respondents with an RD not studied by the RDCRN suggests that our outreach was effective in the RD community.

Underrepresentation of minority groups may be related to the globally recognized barriers to access to RD diagnosis, specialized care and social support [72,73], and other determinants of lack of participation in research [74,75]. Spurred by these findings, the RDCRN has established a cross-consortia Diversity Committee to identify actions to support and expand the participation and engagement of research participants, family members, advocates, and research staff from underrepresented or marginalized communities.

COVID-19 infection was reported by a few respondents (71/3413, 2.08%), but the number was larger than expected. The excess cases were concentrated in the first months of the pandemic and at the end of the data collection period. Thus, individuals who experienced COVID-19 early may have been more motivated to participate in the study. Likewise, survey advertisements in late 2020 could have motivated the participation of patients with an RD who were experiencing the next surge in the epidemic in the United States. Although selection is possible, the number of self-reported infections must underestimate the incidence of COVID-19 in the RD community, as underscored by the large proportion (566/3413, 16.58%) of those who did not know if they acquired the infection. This group reported COVID-19–related symptoms far more frequently than those who did not acquire COVID-19. Given the lack of access to testing early during the pandemic, many individuals with mild COVID-19 infection may have never been diagnosed. An NIH serum study of a sample of the US population estimated that 5 undiagnosed SARS-CoV-2 infections occurred for every COVID-19 case, and that approximately 17 million undiagnosed infections occurred in the United States by mid-July 2020 [76]. These findings have been confirmed in independent assessments [77,78]. Thus, despite the uncertainty about the true number of cases, our results indicate that COVID-19 has had a large impact on the RD community.

Self-reported COVID-19 was not associated with specific RD conditions, but RD comorbidities were associated with increased odds of acquiring COVID-19. Respondents who acquired the infection described a mild disease, with few requiring hospitalization or critical care and only 1 death reported. A mild course of COVID-19 infections has been reported among patients with hereditary hemorrhagic telangiectasia in Italy [50], rare endocrine diseases in a European registry study [38], lysosomal storage diseases in London [79], and Gaucher disease in New York [42]; in a large group of patients with RD in Brazil [43]; and among people with myasthenia gravis in a subanalysis of this survey [52]. These reassuring findings are in contrast with higher COVID-19–related mortality among patients with RD in Hong Kong hospitals compared with in-patients from the general population [37] and people with neurological and neurodevelopmental diseases in a Genomics England cohort study [33]. We observed that the frequency and severity of RD-related symptoms worsened during the pandemic among those who reported acquiring COVID-19 or did not know compared with those who were uninfected, indicating that COVID-19 complicated the underlying RD.

We found a greater use of selected medications among those who acquired COVID-19, little change among those who did not know, and decrease in use and dosage of many medications among those who did not acquire COVID-19. We did not expect the uninfected respondents to report reduced medication use. The need for some medications may have normally decreased between the winter before the beginning of the pandemic and the early summer when most surveys were completed. Seasonal variation exists in the use of antibiotics and prescription drugs for asthma, cystic fibrosis, and other respiratory illnesses [80-83] but not for other medications evaluated in the survey. It is possible that the pandemic decreased access to care and reduced medication use among those who were not in immediate need, whereas respondents with confirmed or possible COVID-19 infection continued to use their medications. This interpretation is not fully supported by the survey responses, which suggest that the pandemic interfered with access to regular health care, treatment for the RD, and hospitalization less frequently with respondents who did not acquire COVID-19. Several studies involving patients with RD have reported reduced access to care [36,40,43,45,52], and of patient organizations [84-88] and providers [79,89-93] have reported increased difficulties in delivering care and services, negatively affecting patients with RD. However, some reports indicate that the rise in telehealth services has had a positive impact in the RD community [43,47,90]. Access to prescription medications for asthma and chronic obstructive pulmonary disease in Great Britain initially briefly increased and then declined to below prepandemic levels during 2020 [94].

We found that the pandemic caused greater mood changes, anxiety, and stress in both respondents and their family members to an extent that required medical attention. These effects were experienced by a significant proportion of all respondents.

**Strengths and Limitations**

This study had certain limitations. First, we could not determine whether respondents were representative of the RD population. However, this concern is mitigated by the representative geographic distribution and the large number of diagnoses reported by the respondents: 48.17% (1644/3413) reported a diagnosis that was not studied by the RDCRN, suggesting that the outreach effort was effective in the RD community at large. Demographics and self-reported outcomes did not vary considerably across RDs, suggesting that the same selection forces affected all respondents.

Second, the information collected pertains mostly to adult, White non-Hispanic people with RDs. Thus, we could not examine
racial or ethnic disparities in the impact of the COVID-19 pandemic.

Third, the main motivation for conducting this survey was that most studies examining the impact of COVID-19 would focus on the general population and not on people living with an RD. We did not have access to a suitable comparison group; therefore, this report is based only on internal comparisons. Finally, data collection was based on self-reports, and we did not have external information to validate our observations. Our inference is based on internal comparisons, the validity of which relies on the assumption that between-group differences are unbiased.

Despite these limitations, this survey had substantial strengths. This is the largest survey conducted to learn about the experience of people living with RDs during the COVID-19 pandemic. The large number of respondents allowed a precise assessment of the prevalence and changes in symptoms and medications used during the early phase of the pandemic and robust internal comparisons among respondents who reported acquiring COVID-19 or were uncertain but had symptoms (ie, may have had an infection without a clinical or laboratory diagnosis). The survey provided a necessary assessment of the impact of the pandemic on the families of people living with RDs. Most participants stated their willingness to participate in follow-up surveys and future research. Thus, this survey provides a basis for evaluating the longitudinal impact of the pandemic on respondents and their families.

Conclusions

In conclusion, the incidence of self-reported COVID-19 infection in this survey of people living with RDs was higher than expected based on population rates, and many respondents were unsure if they had acquired the infection. COVID-19–related symptoms were strongly associated with self-reported infection and with unknown infection status. Although the clinical severity of the infection was not high, self-reported COVID-19 was moderately associated with RD comorbidities and was strongly associated with increased prevalence and severity of RD-related symptoms as well as greater use and dosage of certain medications. The pandemic negatively affected access to health care, RD treatment, and hospitalization; it caused mood changes and greater anxiety for respondents and their family members, requiring medical attention for some. These effects were experienced not only by those who acquired the infection but also by those who did not acquire COVID-19 during the early months of the pandemic or were unsure about it. Continued surveillance of this population is needed to inform interventions to mitigate the impact of COVID-19 and better prepare people with RDs, the health care system, and society for a future pandemic.

Acknowledgments

All Rare Diseases Clinical Research Network consortia were supported by the network’s Data Management and Coordinating Center (DMCC; U2CTR002818). Funding support for the DMCC was provided by the National Center for Advancing Translational Sciences and the National Institute of Neurological Disorders and Stroke. This survey was supported by an administrative supplement to the DMCC award (U2CTR002818-02S1). The authors were supported in part by the following awards: MM and PK (U2CTR002818 and U2CTR002818-02S1), MC and MER (U54AI117804), TF (U54HL096458), and HJK (U54NS115054). The authors would like to thank Ms Erin Fontaine for her outstanding assistance in preparing the manuscript and related material for submission.

The contributors associated with The Principal Investigators of the Rare Diseases Clinical Research Network – Cycle 4 are as follows: Brendan Lee, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; Helen Kim, Department of Anesthesia and Perioperative Care, Center for Cerebrovascular Research University of California San Francisco, CA; Hyder Jinnah, Departments of Neurology, Human Genetics and Pediatrics, Emory University School of Medicine, Atlanta, GA; Mustafa Sahin, Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, Rosamund Stone Zander Translational Neuroscience Center, Boston Children's Hospital, Harvard Medical School, Boston, MA; Eva Morava-Kozicz, Department of Clinical Genomics, Mayo Clinic, Rochester, Minnesota; Stephanie Duggins Davis, Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC; Adeline Vanderver, Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia; Florian Eichel, Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Boston, MA; Michael Shy, Department of Neurology, Carver College of Medicine, University of Iowa, Iowa City, Iowa; Chester Whitley, Department of Pediatrics, University of Minnesota, Minneapolis, MN; Ciaran Sweeney, Department of Neurology, Columbia University Irving Medical Center, New York, NY; Robert Desnick, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; Jennifer Puck and Christopher Dvorak, Division of Pediatric Allergy, Immunology, and Blood and Marrow Transplantation, UCSF Benioff Children's Hospital, University of California San Francisco, San Francisco, CA; Elie Haddad, Department of Pediatrics and Department of Microbiology, Immunology and Infectious Diseases, CHU Sainte-Justine, University of Montreal, Montreal, QC, Canada; Andrea Gropman, Children's National Health System and The George Washington School of Medicine, Washington, DC; Peter Merkel, University of Pennsylvania, Philadelphia; Eileen King, Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH; Michael
Wagner, Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Data Availability
All participant-level data elements necessary to reproduce the analyses presented in the manuscript will be made available upon request. No direct personal identifiers of survey participants will be made available. Survey responses and computed variables used for the analyses presented in the manuscript will be made available in Excel workbooks (Microsoft Corporation) and SAS data sets (SAS Institute). Prospective users will be required to file a data access application including the objectives of data access and the intended use (including a data analysis plan). A data access committee will review the application and determine whether to grant access to the data. Users may be required to sign a data use agreement before obtaining access to the data. Data and supporting documents will be available through the Rare Diseases Clinical Research Network Data Repository (under development) as soon as the Data Repository is active. Please check the Rare Diseases Clinical Research Network website for updates.

Authors’ Contributions
The principal investigators of the Rare Diseases Clinical Research Network (author group) contributed to the survey design, reviewed the manuscript, and approved the version to be published. All named authors (MM, MER, TF, PK, HJK, DWK, MB, and MC) contributed to the survey design, design of the analysis and interpretation of the results, and critical review of the manuscript and reviewed and approved the version to be published. MM and PK directed data acquisition and management and conducted the statistical analysis of the data. MM drafted the manuscript and finalized the version to be published. MM and PK take responsibility for the integrity of the work and agree to investigate and resolve any issues arising from the review process or from the readership of the journal.

Conflicts of Interest
MER is a consultant for Pulm One, Spoon Guru, ClostraBio, Serpin Pharm, Allakos, Celldex, Nextstone One, Bristol-Myers Squibb, Astra Zeneca, Ellodi Pharma, GlaxoSmith Kline, Regeneron or Sanofi, Revolo Biotherapeutics, and Guidepoint; has an equity interest in the first 7 organizations listed; and has received royalties from reslizumab (Teva Pharmaceuticals), PEESSv2 (Mapi Research Trust), and UpToDate. MER is an inventor of patents owned by Cincinnati Children’s Hospital. TF has National Institutes of Health (NIH) grant funding (HL096458, TR003860, AI146999, and HL125241) and received support from Parion Sciences and ReCode Therapeutics for clinical drug trial and observational study. He is a member of the ReCode Therapeutics Primary Ciliary Dyskinesia Clinical Steering Committee and has served as a consultant for TranslateBio and Arrowhead Pharmaceuticals. HJK is a consultant for Roche, Merck Serono, Cabeletta Bio, and UCB Pharmaceuticals and is the Chief Executive Officer and Chief Medical Officer of ARC Biotechnology, LLC, based on US Patent 8,961,98. HJK is the principal investigator of the Rare Disease Network for Myasthenia Gravis (MGNet), funded by National Institute of Neurological Disorders & Stroke (US5 NS115054) and of Targeted Therapy for myasthenia gravis (R41 NS110331). HJK is also coinvestigator for MV2C2 antibody as a new therapeutic for myasthenia gravis (R43NS124329). DWK is a site PI on a Gilead-sponsored study of remdesivir in pediatric patients. All money goes directly to his university (University of Alabama at Birmingham). MB is a consultant for Alexion, Annexon, Arrowhead, Biogen, Denali, Novartis, Orphazyme, Roche, Sanofi, and UniQure. He has a provisional patent for determining the onset of amyotrophic lateral sclerosis. He has received research funding from the NIH and the Muscular Dystrophy Association and serves on the Board of Trustees for the Amyotrophic Lateral Sclerosis Association. MC is currently a consultant for Regeneron, Adare or Ellodi, Astra Zeneca, Sanofi, and Bristol-Myers Squibb. Previously, she was a consultant for Allakos, Shire or Takeda, and Phathom. She currently receives research funding from Regeneron, Allakos, Shire or Takeda, AstraZeneca, and Adare or Ellodi and has received funding from Danone. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Supplemental materials and survey instrument.
[DOCX File, 155 KB - publichealth_v10i1e48430_app1.docx]

References


78. Tanne JH. COVID-19: US cases are greatly underestimated, seroprevalence studies suggest. BMJ 2020 Jul 24;370:m2988. [doi: 10.1136/bmj.m2988] [Medline: 32709608]


**Abbreviations**

| IRB | institutional review board |
| NIH | National Institutes of Health |
| OR | odds ratio |
| RD | rare disease |
| RDCRN | Rare Diseases Clinical Research Network |
| REDCap | Research Electronic Data Capture |
| WHO | World Health Organization |

**Edited by A Mavragani; submitted 23.04.23; peer-reviewed by J Long, M Gong, J Cebrino; comments to author 30.06.23; revised version received 20.09.23; accepted 15.12.23; published 14.02.24.**

Please cite as:

Perspectives of European Patient Advocacy Groups on Volunteer Registries and Vaccine Trials: VACCELERATE Survey Study

Sophia Themistocleous, MSc; Christos D Argyropoulos, MSci, MSc, PhD; Paris Vogazianos, PhD; George Shiamakkides, MSc; Evgenia Noulia; Andrea Nearouch, MSc, PhD; Charalampos Filippou, MSc, PhD; Fiona A Stewart, MSc, PhD; Markela Koniordou, MRes; Ioannis Kopsidas, MD, PhD; Helena H Askling, MD, PhD; Sirkka Vene, MSc; Amandine Gagneux-Brunon, MD, PhD; Jana Baranda Prellezo, MSc; Elena Álvarez-Barco, PhD; Jon Salmanton-García, MSc, PhD; Janina Leckler, PhD; Alan J Macken, MSc; Ruth Joanna Davis, MSc; Anna Maria Azzini, MD; Charis Armeftis, MSc; Margot Hellemans, MSc; Romina Di Marzo, MD, PhD; Ole F Olesen, MD, PhD; Olena Valdenmaier, MSc; Stine Finne Jakobsen, MSc, PhD; Pontus Naucclé, MD, PhD; Odile Launay, MD; Patrick Mallon, PhD; Jordi Ochando, PhD; Pierre van Damme, MD, PhD; Evelina Tacconelli, MD, PhD; Theoklis Zaoutis, MD, MSCE, PhD; Oliver A Cornely, MD, PhD; Zoi Dorothea Pana, MD, MSc, PhD

1School of Medicine, European University Cyprus, Nicosia, Cyprus
2Department of Behavioural and Social Sciences, School of Humanities, Social and Education Sciences, European University Cyprus, Nicosia, Cyprus
3Faculty of Medicine and University Hospital Cologne, Institute of Translational Research, Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany
4Faculty of Medicine, University Hospital Cologne, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Düsseldorf (CIO ABCD) and Excellence Center for Medical Mycology (ECMM), University of Cologne, Cologne, Germany
5Collaborative Center for Clinical Epidemiology and Outcomes Research (CLEO), Athens, Greece
6Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden
7Division of Infectious Diseases, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden
8Centre d’investigation clinique-INERMER 1408, University Hospital of Saint-Etienne, Saint-Etienne, France
9Groupe Immunité Muqueuse et Agents Pathogènes (GIMAP), EA3064 – Medical School of Saint-Etienne, University of Lyon, Saint-Etienne, France
10Microbiology Section, Department of Pharmaceutical Sciences and of Health, Faculty of Pharmacy, Universidad San Pablo-Centro de Estudios Universitarios (CEU), Madrid, Spain
11Centre for Experimental Pathogen Host Research, University College Dublin School of Medicine, National University of Ireland, Dublin, Ireland
12Department of Diagnostic and Public Health, Infectious Diseases, University of Verona, Verona, Italy
13Vaccine & Infectious Disease Institute – VAXINFECTIO, Centre of Evaluation of Vaccination,, Faculty of Medicine and Health Science, Universiteit Antwerpen, Antwerp, Belgium
14European Vaccine Initiative (EVI), Heidelberg, Germany
15Centre of Excellence for Health, Immunity and Infections (CHIP), Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
16Inserm CIC 1417, I-REIVAC, University Hospital of Cochin-Broca-Hôtel-Dieu, University of Paris-Descartes, Paris, France
17Faculty of Medicine and University Hospital Cologne, Clinical Trials Centre Cologne (ZKS Köln), University of Cologne, Cologne, Germany
18German Centre for Infection Research (DZIF), Partner Site Bonn-Cologne, Cologne, Germany
19Faculty of Medicine, and University Hospital Cologne, Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany

†On behalf of the VACCELERATE Consortium

Corresponding Author:
Zoi Dorothea Pana, MD, MSc, PhD
School of Medicine
European University Cyprus
6 Diogenis Str.
Engomi
Nicosia, 2404
Cyprus
Phone: 357 94049474
Email: z.pana@euc.ac.cy
Abstract

Background: The VACCELERATE Pan-European Scientific network aims to strengthen the foundation of vaccine trial research across Europe by following the principles of equity, inclusion, and diversity. The VACCELERATE Volunteer Registry network provides access to vaccine trial sites across the European region and supports a sustainable volunteer platform for identifying potential participants for forthcoming vaccine clinical research.

Objective: The aim of this study was to approach members of patient advocacy groups (PAGs) across Europe to assess their willingness to register for the VACCELERATE Volunteer Registry and their perspectives related to participating in vaccine trials.

Methods: In an effort to understand how to increase recruitment for the VACCELERATE Volunteer Registry, a standardized survey was developed in English and translated into 8 different languages (Dutch, English, French, German, Greek, Italian, Spanish, and Swedish) by the respective National Coordinator team. The online, anonymous survey was circulated, from March 2022 to May 2022, to PAGs across 10 European countries (Belgium, Cyprus, Denmark, France, Germany, Greece, Ireland, Italy, Spain, and Sweden) to share with their members. The questionnaire constituted of multiple choice and open-ended questions evaluating information regarding participants’ perceptions on participating in vaccine trials and their willingness to become involved in the VACCELERATE Volunteer Registry.

Results: In total, 520 responses were collected and analyzed. The PAG members reported that the principal criteria influencing their decision to participate in clinical trials overall are (1) the risks involved, (2) the benefits that will be gained from their potential participation, and (3) the quality and quantity of information provided regarding the trial. The survey revealed that, out of the 520 respondents, 133 individuals across all age groups were “positive” toward registering in the VACCELERATE Volunteer Registry, with an additional 47 individuals reporting being “very positive.” Respondents from Northern European countries were 1.725 (95% CI 1.206-2.468) times more likely to be willing to participate in the VACCELERATE Volunteer Registry than respondents from Southern European countries.

Conclusions: Factors discouraging participants from joining vaccine trial registries or clinical trials primarily include concerns of the safety of novel vaccines and a lack of trust in those involved in vaccine development. These outcomes aid in identifying issues and setbacks in present registries, providing the VACCELERATE network with feedback on how to potentially increase participation and enrollment in trials across Europe. Development of European health communication strategies among diverse public communities, especially via PAGs, is the key for increasing patients’ willingness to participate in clinical studies.

Introduction

Within the VACCELERATE consortium, the VACCELERATE Volunteer Registry was assembled on the principles of equity, inclusion, and diversity in vaccine clinical research across Europe [1-3]. Patient advocacy groups (PAGs) personify all 3 principles, with the aim of increasing community participation in research [4-6]. PAG members can monitor the latest news in medical conduct, have access to hard-to-reach medical procedures and treatments, and communicate with health care professionals [5]. Patients with concomitant chronic conditions are largely excluded from three-quarters of research studies listed in ClinicalTrials.gov [7]. Hence, the mission of PAGs is to assist with attaining diversity and equal representation of populations in trials, especially with patients with comorbidities who are more prone to be severely affected by infections [8], as well as overcome other barriers, such as language and limited health education. For example, the European Patients’ Forum operates to make patient organization voices heard and contends with policy-formulating processes affecting patients across Europe [9]. Alternatively, the European Patients’ Academy on Therapeutic Innovation acts to engage patients through their Patient Expert Training Programme that operates to educate and train in the context of medicine development and research [9]. These initiatives raise patients’ awareness and health literacy and may alleviate concerns regarding clinical trial research.

Another factor is trust in the health care system and its professionals. Health care providers expressing a level of optimism and conveying good quality information to potential applicants is a key parameter to improving recruitment in trials [10]. Preceding studies have recorded the willingness of participants to be involved in COVID-19 vaccine trials; in studies in France and Jordan, 47.6% (1552/3259) and 36.1% (465/1287) of the participants, respectively, were willing to participate in COVID-19 vaccine trials; in studies in France and Jordan, 47.6% (1552/3259) and 36.1% (465/1287) of the participants, respectively, were willing to participate in COVID-19 vaccine trials in 2020 [11,12]. Another study conducted in Uganda that targeted health care workers and their supporting staff (non-health care professionals) across 3 hospitals recorded a 70.2% (n=461) willingness to participate in COVID-19 vaccine trials [13]. This raised an uncertainty whether health literacy or unfamiliarity with a health environment affects judgment, trust, and willingness to participate in COVID-19 vaccine trials in different community environments and geographical areas.

Reaching out to potential volunteers is one of many steps. Once presented with the opportunity to participate in trials, whether
that be vaccine trials or clinical trials, one is troubled by the possibility of experiencing unpleasant symptoms and the consumption of time for participation, among other concerns, as well as mistrust in pharmaceutical companies [12]. Collecting data on patients’ thoughts, opinions, and beliefs across Europe may reveal effective methods for promoting volunteer registries and clinical research equity among diverse communities [14-16]. The objectives of this study are to assess European PAG members’ perspectives regarding vaccine trial participation along with their willingness to register in the Volunteer Registry [1,10]. Both are crucial factors prioritized by the VACCELERATE consortium [17]. Overall, the outcomes of this study include the promotion of the VACCELERATE Volunteer Registry and ensuring the effective delivery of its work.

### Methods

#### Study Design and Population

A standardized survey was designed to adhere to national bioethics’ requirements for each participating country and its target population including adult patients who are members of PAGs. The survey was transferred to an online format in collaboration with a consulting company (Conread Research Bureau Ltd) and was translated to English, Dutch, English, French, German, Greek, Italian, Spanish, and Swedish by the respective National Coordinator (NC) team of VACCELERATE. A list of PAGs per country was created, along with the appropriate contact details, by the task force undertaking this project with the guidance and support of the external consulting company and the respective VACCELERATE NC team.

#### Data Collection

An official invitation letter was disseminated to PAGs via email between March 9, 2022, and May 16, 2022, together with the link to the online survey (see Multimedia Appendix 1). As the aim was to record PAG members’ willingness to register at the VACCELERATE Volunteer Registry [15,18], a description of the VACCELERATE program was provided at the beginning of the survey, followed by a question that asked whether participants were willing to join the registry. Weekly reminders were sent via email to the NC or PAGs directly to increase responsiveness and engagement. All participants could exit the survey at any time they wished without completing it. Incomplete surveys (or without a signed consent form) were excluded from the final anonymous data analysis.

The collected data from the survey (Multimedia Appendix 1) included sociodemographic characteristics, prior COVID-19 vaccination status, history of SARS-CoV-2 infection, the source participants turn to for knowledge or information on health developments, and information regarding participants’ perceptions of important criteria for participating in vaccine trials. The questions posed were also an effort to understand communities’ and patients’ concerns regarding vaccine trials. This information could assist VACCELERATE investigators in designing and recruiting for future trials.

#### Statistical Analysis

All results present the frequency at which a response was selected, in percentages. For the ordinal data, a Likert rating scale was used with the following options per characteristic:

1. Importance/relevance (1. Not at all important, 2. Somewhat not important, 3. Neither important nor unimportant, 4. Somewhat important, 5. To a great extent/extremely important)


The percentages of responses for each of these selections were calculated. A score was established by calculating the dispersal of responses across the Likert scale (eg, the average between responses 1 and 5). Analysis was performed using SPSS v26 (IBM Corp).

For bivariate analysis and multivariate analysis, we used logistic regression analysis, as we wanted to calculate the effect of the independent variable on the binary dependent variable adjusted for confounders (ie, variables present that affect the variables under study and thus not allowing the results to mirror the real association between the dependent and independent variables). In particular, logistic regression analysis was performed to ascertain the effects of all demographic variables on the likelihood that a participant would be willing to participate in the VACCELERATE Volunteer Registry.

#### Ethical Considerations

PAG members were requested to access a web link where they would find an electronic version of the survey and complete the questions honestly. To ensure all responders fully understood the objectives of the survey, the concept of anonymity, and data safety, the survey was provided in 9 different languages. The survey was approved by the Cyprus National Bioethics Committee (EEBK EIT 2021.01.118) and other bioethics committees from the participating countries according to the national recommendations (Spain, Italy). Only responses with positive informed consent were processed, and no compensation was provided to responders. All records were anonymous and stored on a password-protected computer at Conread Research Bureau Ltd.

### Results

#### General Characteristics

A total of 520 responses were recorded through PAGs across the European region, as represented by a choropleth map (Figure 1) with the highest response rates recorded from Germany (165/520, 31.7%), Cyprus (149/520, 28.7%), and Greece (76/520, 14.6%). The fewest responses were collected from Denmark, Finland, Czechia, and Croatia (all had 1 response). Responses were obtained from PAG members from Belgium, Cyprus, Denmark, France, Germany, Greece, Ireland, Italy, Spain, and Sweden; some responses were obtained from other countries depending on the location of the respondents’ current residence.
Sociodemographic data showed higher participation rates by women (323/520, 62.1%). Regarding the age distribution (<29 years, 30-39 years, 40-49 years, 50-59 years, 60-69 years, and ≥70 years), most responses were collected from those aged 50-59 years (151/520, 29%). Regarding educational status, the greatest proportion of respondents had a postgraduate or doctorate degree (200/520, 38.5%), of whom 56% (112/200) reported that they predominantly consult news media for advice on health developments, while 55.5% (111/200) seek out physicians' advice or even official international health organization websites and media (111/200, 55.5%). Individuals with a primary school education (106/168, 63.1%) mostly consult their doctors, while they tend to avoid social media (22/168, 13.1%). Last, undergraduates (89/146, 61%) rely on news media for advice on health developments.

Until May 2022, 71% (369/520) of PAG members were vaccinated with 3 doses of a COVID-19 vaccine, while 13.7% (71/520) had completed a fourth dose. Regarding the participants’ self-reported diagnosis of a COVID-19 infection, 37% (10/27; the highest percentage among participants with a self-reported disease group) of individuals with a confirmed COVID-19 diagnosis had “chronic cardio and pulmonary diseases”; however, an even higher percentage (10/21, 48%) of participants with a confirmed COVID-19 infection did not wish to state if they had an illness.

Table 1 depicts the self-reported chronic diseases based on stratification by the 2 main age groups, namely Age Group 1 and Age Group 2. The former includes 6 age subcategories (<29 years, 30-39 years, 40-49 years, 50-59 years, 60-69 years, and ≥70 years), while the latter comprises 2 age subcategories (18-59 years and ≥60 years). For Age Group 1, those aged ≥70 years had the highest percentage of “All cancer conditions” (8/38, 21%), while those aged 50 years to 59 years had the lowest percentage (6/152, 4%). The highest and lowest percentages of “Rare diseases” were present for those aged 30 years to 39 years (24/88, 27%) or 60 years to 69 years (13/106, 12.3%). Finally, “Other diseases” were present the most for those aged 50 years to 59 years (94/152, 61.8%) and the least for those aged ≥70 years (15/38, 40%). On the other hand, in Age Group 2, the highest disease category for both age subgroups (18-59 years and ≥60 years) was “Other diseases,” at 53.2% (200/376) and 50.7% (73/144), respectively.
Table 1. Participants’ self-reported chronic illnesses and diseases or conditions stratified by Age Groups 1 and 2.

<table>
<thead>
<tr>
<th>Chronic illness or disease</th>
<th>Age Group 1 (n=520)</th>
<th>Age Group 2 (n=520)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;29 years (n=40), n (%)</td>
<td>18-59 years (n=376), n (%)</td>
</tr>
<tr>
<td>All cancer conditions^a</td>
<td>2 (5)</td>
<td>10 (11.4)</td>
</tr>
<tr>
<td>Rare diseases^b</td>
<td>7 (17.5)</td>
<td>24 (27.3)</td>
</tr>
<tr>
<td>Immunosuppression^c</td>
<td>3 (7.5)</td>
<td>9 (10.2)</td>
</tr>
<tr>
<td>Chronic cardio and pulmonary diseases</td>
<td>2 (5)</td>
<td>3 (3.4)</td>
</tr>
<tr>
<td>Other diseases^d</td>
<td>22 (55)</td>
<td>40 (45.5)</td>
</tr>
<tr>
<td>I do not wish to state the chronic disease I have</td>
<td>4 (10)</td>
<td>2 (2.3)</td>
</tr>
</tbody>
</table>

^a Hematological and solid cancers.
^b A broad number of conditions.
^c Acquired HIV or genetic-immunodeficiencies.
^d Other diseases as stated by the respondents themselves included the following: rheumatoid arthritis, Parkinson disease, osteoarthritis, diabetes, celiac disease, ulcerative colitis, rheumatism, stroke, arrhythmias, hepatitis, systemic lupus erythematosus, psoriasis, ankylosing spondyloarthropathy, peripheral spondyloarthropy, multiple sclerosis, and asthma, among others.

Clinical Trial Awareness and Participation

Of the 520 respondents, 93.1% (484/520) claimed to be aware of the term “clinical trial,” of whom 22.5% (109/484) reported to have previously participated in clinical trials (Table 2). Regarding their rating of their experiences, having the ability to give feedback about the clinical trial process received the poorest rating (mean 3.29 out of 5). The highest rating was given for the level of information received while participating in the trial. None of the responses received a full rating. All 520 respondents were also asked to rate a list of criteria based on how important each criterion was in affecting their participation in clinical trials (Table 3). The criterion rated as “Extremely important” by the highest number of participants (386/520, 74.2%) was understanding the risks involved in one’s participation in such trials, followed by “Giving my consent after being fully informed on the process and have all my questions answered” (369/520, 71%). Having family or friends who had previously taken part in a vaccine clinical trial was an inadequate component for persuading one to participate or not participate in clinical trials (51/520, 9.8% voted it as “Extremely important”).

Table 2. From the 520 patient advocacy group (PAG) members who responded to the survey, 109 respondents had participated in clinical trials in the past and rated their experiences.

<table>
<thead>
<tr>
<th>Respondents who had participated in clinical trials before (n=109)</th>
<th>Rating (1 to 5), n (%)</th>
<th>Mean rating (1 to 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation of the first set of information received (n=98)</td>
<td>Very bad (1)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Information about the level of risk you took while taking part (n=100)</td>
<td>Bad (2)</td>
<td>7 (7.1)</td>
</tr>
<tr>
<td>The level of information you received (n=101)</td>
<td>Neutral (3)</td>
<td>0</td>
</tr>
<tr>
<td>Ability to give your feedback about the clinical trial process (n=96)</td>
<td>Good (4)</td>
<td>6 (6.3)</td>
</tr>
<tr>
<td>Regular communication with the medical team (n=99)</td>
<td>Very good (5)</td>
<td>7 (7.1)</td>
</tr>
<tr>
<td>Distance to/from the clinical site (n=99)</td>
<td>5 (5.1)</td>
<td>18 (18.2)</td>
</tr>
</tbody>
</table>
Participants were asked to evaluate the importance of each of the following 9 statements in their decision to take part in clinical trials: (1) understanding the purpose of the trial and how it will benefit others now and in the future, (2) understanding the risks involved, (3) understanding what the benefits will be from my participation, (4) giving my consent after being fully informed on the process and have all my questions answered, (5) being convinced that my personal data is fully protected, (6) being convinced that my participation will be free of any financial cost, (7) the attitude of the researchers/physicians, (8) getting to know/meet others that will take part in the same trial, (9) having family or friends who have previously taken part in a clinical trial.

According to the regression model analysis for each of the aforementioned statements, the respondents from Northern European countries were (1) 4.484 (95% CI 2.618-7.962) times more likely to consider statement 1 (understanding the purpose of the trial and how it will benefit others now and in the future) as important for their decision, (2) 4.425 (95% CI 2.481-7.874) times more likely to consider statement 2 (understanding the risks involved) as important for their decision, (3) 1.838 (95% CI 1.164-2.907) times more likely to consider statement 3 (understanding what the benefits will be from my participation) as important for their decision (women were 1.628 [95% CI 1.041-2.547] times more likely to consider this statement as important for their decision), (4) 3.46 (95% CI 2.00-5.99) times more likely to consider statement 4 (giving my consent after being fully informed on the process and have all my questions answered) as important for their decision, (5) 2.551 (95% CI 1.639-3.968) times more likely to consider statement 5 (to be convinced that my personal data is fully protected) as important for their decision (women were 2.116 [95% CI 1.384-3.236] times more likely to consider it as important for their decision). (6) 1.773 (95% CI 1.127-2.786) times more likely to consider statement 6 (being convinced that my participation will be free of any financial cost) as important for their decision, (7) 2.470 (95% CI 1.444-4.226) times more likely to consider statement 7 (the attitude of the researchers/physicians) as important for their decision (people younger than 70 years were 3.481 [95% CI 1.586-7.642] times more likely to consider this statement as important for their decision). Respondents younger than 70 years were (8) 3.33 (95% CI 1.35-8.20) times more likely to consider statement 8 (getting to know/meet others that will take part in the same trial) as important for their decision. Respondents from Southern European countries were (9) 2.96 (95% CI 1.90-4.63) times more likely to consider statement 9 (having family or friends who have previously taken part in a vaccine trial) as important for their decision.

**VACCELERATE Volunteer Registry**

The score for a willingness to participate in the Volunteer Registry was 4 “Positive” for 25.6% (133/520) of the participants and 5 “Very positive” for 9% (47/520) of the participants (Table 4). The highest score for a willingness to participate was 3.24 for those aged 60 years to 69 years. The perspectives of participants regarding registration (per country) in the Volunteer Registry is presented in Table 5, while motives for participating in the Volunteer Registry did not vary among the sexes (Table 6).
### Table 4. Insight into the willingness of the 520 respondents to participate in the VACCELERATE Volunteer Registry sorted by age group.

<table>
<thead>
<tr>
<th>Age Group 1</th>
<th>Willingness to participate (1-5), n (%)</th>
<th>Mean rating (1 to 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very negative (1) Negative (2) Neither negative nor positive (3) Positive (4) Very positive (5)</td>
<td></td>
</tr>
<tr>
<td>Total sample</td>
<td>32 (6.2) 72 (13.9) 236 (45.3) 133 (25.6) 47 (9.0)</td>
<td>3.18</td>
</tr>
<tr>
<td>≤29 years (n=40)</td>
<td>3 (7.5) 5 (12.5) 15 (37.5) 14 (35) 3 (7.5)</td>
<td>3.23</td>
</tr>
<tr>
<td>30-39 years (n=88)</td>
<td>8 (9.1) 12 (13.6) 36 (40.9) 27 (30.7) 5 (5.7)</td>
<td>3.10</td>
</tr>
<tr>
<td>40-49 years (n=96)</td>
<td>8 (8.3) 15 (15.6) 44 (45.8) 17 (17.7) 12 (12.5)</td>
<td>3.10</td>
</tr>
<tr>
<td>50-59 years (n=152)</td>
<td>7 (4.6) 19 (12.5) 76 (50) 36 (23.7) 14 (9.2)</td>
<td>3.20</td>
</tr>
<tr>
<td>60-69 years (n=106)</td>
<td>5 (4.7) 14 (13.2) 48 (45.3) 29 (27.4) 10 (9.4)</td>
<td>3.24</td>
</tr>
<tr>
<td>≥70 years (n=38)</td>
<td>1 (2.6) 7 (18.4) 17 (44.7) 10 (26.3) 3 (7.9)</td>
<td>3.18</td>
</tr>
</tbody>
</table>

### Table 5. Patient advocacy group (PAG) members’ (n=520) willingness to participate in the VACCELERATE Volunteer Registry, by country of residence.

<table>
<thead>
<tr>
<th>Country of residence</th>
<th>Willingness to participate (1-5), n (%)</th>
<th>Mean rating (1-5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very negative (1) Negative (2) Neither negative nor positive (3) Positive (4) Very positive (5)</td>
<td></td>
</tr>
<tr>
<td>Germany (n=165)</td>
<td>7 (4.2) 23 (13.9) 73 (44.2) 47 (28.5) 15 (9.1)</td>
<td>3.24</td>
</tr>
<tr>
<td>Cyprus (n=149)</td>
<td>15 (10.1) 24 (16.1) 67 (45) 35 (23.5) 8 (5.4)</td>
<td>2.98</td>
</tr>
<tr>
<td>Greece (n=76)</td>
<td>5 (6.6) 15 (19.7) 39 (51.3) 9 (11.9) 8 (10.5)</td>
<td>3.00</td>
</tr>
<tr>
<td>Ireland (n=36)</td>
<td>0 2 (5.6) 16 (44.4) 10 (27.8) 8 (22.2)</td>
<td>3.67</td>
</tr>
<tr>
<td>Spain (n=33)</td>
<td>3 (9.1) 3 (9.1) 15 (45.5) 10 (30.3) 2 (6.1)</td>
<td>3.15</td>
</tr>
<tr>
<td>Sweden (n=19)</td>
<td>0 1 (5.3) 10 (52.6) 6 (31.6) 2 (10.5)</td>
<td>3.47</td>
</tr>
<tr>
<td>France (n=13)</td>
<td>1 (7.7) 2 (15.4) 3 (23.1) 5 (38.5) 2 (15.4)</td>
<td>3.38</td>
</tr>
<tr>
<td>Other(^a) (n=29)</td>
<td>1 (3.5) 2 (6.9) 13 (44.8) 11 (37.9) 2 (6.9)</td>
<td>3.38</td>
</tr>
</tbody>
</table>

\(^a\)Belgium, Croatia, Czecia, Denmark, Finland, Hungary, Italy, Luxemburg, The Netherlands, and Switzerland; all with the lowest response rates.

### Table 6. Of the respondents willing to participate in the VACCELERATE Volunteer Registry (n=180), reasons given for opting in, by gender.

<table>
<thead>
<tr>
<th>Reasons</th>
<th>All respondents, n (%)</th>
<th>Men (n=77), n (%)</th>
<th>Women (n=103), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Help advance medical research</td>
<td>155 (86.1)</td>
<td>68 (88.3)</td>
<td>87 (84.5)</td>
</tr>
<tr>
<td>Receive payment upon my participation</td>
<td>26 (14.4)</td>
<td>12 (15.6)</td>
<td>14 (13.6)</td>
</tr>
<tr>
<td>I lost one of my own and want to help medical science advance to help people overcome their problems.</td>
<td>29 (16.1)</td>
<td>12 (15.6)</td>
<td>17 (16.5)</td>
</tr>
<tr>
<td>As a patient, I believe science would be greatly supported to enrich their data by my participation.</td>
<td>131 (72.8)</td>
<td>57 (74)</td>
<td>74 (71.8)</td>
</tr>
<tr>
<td>I (or someone close to me) had a pleasant experience while taking part in a clinical trial in the past.</td>
<td>14 (7.8)</td>
<td>8 (10.4)</td>
<td>6 (5.8)</td>
</tr>
<tr>
<td>I will be one of the first to get to know the medical advancements.</td>
<td>53 (29.4)</td>
<td>18 (23.4)</td>
<td>35 (34)</td>
</tr>
<tr>
<td>Other reasons</td>
<td>3 (1.7)</td>
<td>1 (1.3)</td>
<td>2 (1.9)</td>
</tr>
</tbody>
</table>

Countries in which participants were more willing to join the Volunteer Registry were Ireland (3.67/5.00), Sweden (3.47/5.00), and France (3.38/5.00; Table 5). The respondents who were both “positive” and “very positive” were then directed to a supplementary question to select a reason for choosing to join the Volunteer Registry. The most widely held reason was to “Help advance medical research” (155/180, 86.1%), and the greatest proportion of this response came from individuals aged <29 years or ≥70 years. The second most popular response, at 72.8% (131/180), was “As a patient, I believe science would be greatly supported to enrich their data by my participation.”

An additional point of interest was the option to “Receive payment upon my participation,” which was recorded by 18.8% (24/128) of those aged 18 years to 59 years, whereas for those ≥60 years of age, the percentage was significantly lower, at only 3.9% (2/52).

The 104 survey participants who were unwilling to register in the Volunteer Registry were asked for justification. The 2 most frequently reported reasons across all countries for refusing were “Do not trust pharmaceutical companies/medical researchers/private or public companies” (27/104, 26%).
Trial investigators should prioritize the incorporation of benefits; factors a trial team should be able to deliver [9]. procedures to be carried out, and volunteer support are all key medication or vaccine, description of any symptoms and Communication regarding any concerns a potential volunteer confidence in and attract the attention of potential volunteers. and accurately training the medical team to execute all aspects and stages of a clinical trial are also of high importance to instill and properly training the medical team to execute all aspects and stages of a clinical trial are also of high importance to instill

Discussion

Principal Findings

The COVID-19 pandemic has shown the significance and importance of having people voluntarily participating in the development of new vaccines. The objectives of this effort were to approach PAG members across Europe through an online survey to assess their perspectives related to participation in vaccine trials despite their diagnoses. This study was also an opportunity to promote the Volunteer Registry to more diverse communities and measure patients’ willingness to register in the registry. VACCELERATE intends to approach individuals with morbidities or comorbidities with an objective to ensure inclusiveness and research equity as well as increase literacy about vaccine and clinical trials. Ultimately, VACCELERATE hopes to make participation in well-designed clinical trials a normal part of the scientific progress that leads to better medical care for all [21].

At the time of questionnaire preparation, only 3 COVID-19 vaccination doses were advocated across Europe; however, upon survey launch, a fourth dose had been introduced, which may explain the low percentage of the later dose recorded in the survey. The survey results also demonstrate that members of all educational backgrounds seek health advice from a doctor, emphasizing the importance of communicating with and training doctors on innovative medicine and clinical trials. Recruiting and accurately training the medical team to execute all aspects and stages of a clinical trial are also of high importance to instill confidence in and attract the attention of potential volunteers. Communication regarding any concerns a potential volunteer may have, a clear breakdown of all information regarding the medication or vaccine, description of any symptoms and procedures to be carried out, and volunteer support are all key factors a trial team should be able to deliver [9].

Trial investigators should prioritize the incorporation of benefits; communicate trial benefits (access to new drugs and vaccines) to potential volunteers; and ensure patients feel safe, supported, and, above all, well informed throughout the whole duration of a trial. This can be achieved by establishing trust and addressing knowledge deficits (eg, provide informative material in the patient’s native language and using simple terms) to prevent misinformation and uphold transparency. Communicating this information and establishing a network of active participants in clinical trials may be facilitated via collaboration with PAGs [22]. Furthermore, educating potential volunteers on the value of their participation and input in vaccine trials and how the pandemic’s impact on the community can be minimized may inspire further volunteering [23,24].

The most popular reason for opting to participate in the Volunteer Registry is to help advance medical research, an opinion that has been recorded in the past as a motivating factor for participation [23]. Potential volunteers may also be driven by the offer of potential personal benefit, a view which was not strongly supported by the PAG members of this study [24]. On the other hand, unwillingness to join the registry was attributable to participants’ mistrust in pharmaceutical companies, medical researchers, private or public companies, and governmental agencies [12]. Volunteering may also be affected by personal ideologies, past experiences, ethnicity, and religion, among other factors. Based on this study’s results, a significant difference in the willingness to participate in the registry has been observed between Northern and Southern European respondents. This observation needs to be further evaluated in the future.

Feedback on past participation in clinical trials involving a certain medicine showed that PAG members had an overall positive experience. Participants reported that, in terms of the quality and quantity, the information they received about the level of risk they were taking while participating in the trial was good overall. Still, 2 important aspects were lacking: offering volunteers the opportunity to provide feedback regarding their experience in the trial and providing opportunities for regular communication with the medical team. Both of these aspects may prevent participants from joining vaccine trial registries or participating in clinical trials, as these outcomes may affect the security and trust one feels in partaking in clinical trials. Therefore, the participants’ reactions indicate there is still room for improvement when designing and developing clinical trials in which support, inclusiveness, and safety of volunteers are prioritized.

Limitations

Limitations were recognized during the execution of the study. The number of responses from women was significantly greater than that from men, raising concerns of gender disparities as well as limited generalizability. An additional factor that may have shaped these results is that a greater number of respondents had a higher level of education (38% of them had a postgraduate or doctorate degree). Finally, as most responses were collected from Germany, Greece, and Cyprus (Figure 1), there was deliberation as to whether the results can be considered representative of PAGs across Europe.
Conclusions
This study exhibited and analyzed data from a Pan-European, online survey targeting adult individuals with chronic underlying conditions (patients) from the European region who are members of PAGs.

Despite the high percentage of participants acknowledging awareness of the term “clinical trials,” few individuals affirmed they were willing to participate in the Volunteer Registry due to a lack of trust in certain bodies involved in clinical trial conduct, concerns involving safety, or even religious beliefs. The results of this online survey represent only an initial indication of patient willingness to register in the Volunteer Registry and support that PAG members value advancements in medical research and clinical trials.

The information presented aids in interpreting issues and setbacks in existing registrations so that a plan may be constructed to improve future promotion, campaigns, and approach schemes for vaccine trials, in particular for PAG members. Overall, raising public awareness of the benefits of clinical trials and improving health literacy may increase participation in vaccine clinical trials. Careful planning and more thought need to be invested in designing trials to guarantee inclusiveness, equality, and strong support networks for groups such as PAGs.

Acknowledgments
This work was supported by the European Union’s Horizon 2020 research and innovation program under grant agreement number 101037867 (VACCELERATE). We would like to express our sincere gratitude to all European Patient Advocacy Groups for their work and assistance with disseminating this survey. The survey was designed and developed in collaboration with Conread Research Bureau Ltd, Nicosia, Cyprus. The research leading to these results was conducted as part of the VACCELERATE consortium. For further information, please refer to the official website of the VACCELERATE network [24].

Data Availability
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest
PM has received honoraria or research funding to his institution from Gilead Sciences, Viiv Healthcare, MSD, and AstraZeneca. OAC reports grants or contracts from BMBF, Cidara, EU-DG RTD (101037867), F2G, Gilead, MedPace, MSD, Mundipharma, Octapharma, Pfizer, and Scynexis; consulting fees from AbbVie, AiCuris, Biocon, Cidara, Gilead, IQVIA, Janssen, Matinas, MedPace, Menarini, Moderna, Molecular Partners, MSG-ERC, Noxxon, Octapharma, Pfizer, PSI, Scynexis, and Seres; honoraria for lectures from Abbott, AbbVie, Al-Jazeera Pharmaceuticals/Hikma, Gilead, Grupo Biotoscana/United Medical/Knight, MedScape, MedUpdate, Merck/MSD, Noscendo, Pfizer, Shionogi, and streamedup!; payment for expert testimony from Cidara; participation on a Data Safety Monitoring Board or Advisory Board for Boston Strategic Partners, Cidara, IQVIA, Janssen, MedPace, PSI, Pulmocide, Shionogi, and The Prime Meridian Group; a patent at the German Patent and Trade Mark Office (DE 10 2021 113 007.7); stocks from CoRe Consulting and EasyRadiology; and other interests from Wiley. The other authors declare participation on a Data Safety Monitoring Board or Advisory Board for Boston Strategic Partners, Cidara, IQVIA, Janssen, MedPace, PSI, Pulmocide, Shionogi, and The Prime Meridian Group; a patent at the German Patent and Trade Mark Office (DE 10 2021 113 007.7); stocks from CoRe Consulting and EasyRadiology; and other interests from Wiley. The other authors declare that they have no known competing financial interests nor personal relationships that could have appeared to influence the work reported in this paper.

Multimedia Appendix 1
Invitation letter.
[DOCX File , 70 KB - publichealth_v10i1e47241_app1.docx ]

References


17. VACCELERATE Volunteer Registry. URL: https://vaccelerate.eu/volunteer-registry/ [accessed 2024-03-01]

18. MapChart. URL: https://www.mapchart.net/ [accessed 2024-04-01]

19. Attribution-ShareAlike 4.0 International (CC BY-SA 4.0 DEED). Creative Commons. URL: https://creativecommons.org/licenses/by-sa/4.0/ [accessed 2024-04-01]


Abbreviations

PAG: patient advocacy group
NC: National Coordinator
Perspectives of European Patient Advocacy Groups on Volunteer Registries and Vaccine Trials: VACCELERATE Survey Study

JMIR Public Health Surveill 2024;10:e47241

URL: https://publichealth.jmir.org/2024/1/e47241
doi:10.2196/47241
PMID:38573762
The Changing Landscape of Respiratory Viruses Contributing to Hospitalizations in Quebec, Canada: Results From an Active Hospital-Based Surveillance Study

Rodica Gilca1,2,3, MD, PhD; Rachid Amini1, MD, MSc; Sara Carazo1,2,3, MD, PhD; Radhouene Doggui1, PhD; Charles Frenette1, MD; Guy Boivin2, MD, PhD; Hugues Charest5, PhD; Jeannot Dumaresq6, MD

1Direction des risques biologiques, Institut national de santé publique du Québec, Québec, QC, Canada
2Research Center of Centre hospitalier universitaire de Québec-Université Laval, Québec, QC, Canada
3Département de médecine préventive, Université Laval, Québec, QC, Canada
4Department of Medicine, Division of Infectious Diseases, McGill University Health Center, Montreal, QC, Canada
5Laboratoire de santé publique, Institut national de santé publique du Québec, Montreal, QC, Canada
6Département of Microbiology and Infectiology, Centre intégré de santé et de services sociaux de Chaudière-Appalaches, Levis, QC, Canada

Corresponding Author:
Rodica Gilca, MD, PhD
Direction des risques biologiques
Institut national de santé publique du Québec
945 Av. Wolfe
Québec, QC, G1V5B3
Canada
Phone: 1 4186505115 ext 6278
Email: rodica.gilca@inspq.qc.ca

Abstract

Background: A comprehensive description of the combined effect of SARS-CoV-2 and respiratory viruses other than SARS-CoV-2 (ORVs) on acute respiratory infection (ARI) hospitalizations is lacking.

Objective: This study aimed to compare the viral etiology of ARI hospitalizations before the pandemic (8 prepandemic influenza seasons, 2012-13 to 2019-20) and during 3 pandemic years (periods of increased SARS-CoV-2 and ORV circulation in 2020-21, 2021-22, and 2022-23) from an active hospital-based surveillance network in Quebec, Canada.

Methods: We compared the detection of ORVs and SARS-CoV-2 during 3 pandemic years to that in 8 prepandemic influenza seasons among patients hospitalized with ARI who were tested systematically by the same multiplex polymerase chain reaction (PCR) assay during periods of intense respiratory virus (RV) circulation. The proportions of infections between prepandemic and pandemic years were compared by using appropriate statistical tests.

Results: During prepandemic influenza seasons, overall RV detection was 92.7% (1384/1493) (respiratory syncytial virus [RSV]: 721/1493, 48.3%; coinfections: 456/1493, 30.5%) in children (<18 years) and 62.8% (2723/4339) (influenza: 1742/4339, 40.1%; coinfections: 264/4339, 6.1%) in adults. Overall RV detection in children was lower during pandemic years but increased from 58.6% (17/29) in 2020-21 (all ORVs; coinfections: 7/29, 24.1%) to 90.3% (308/341) in 2021-22 (ORVs: 278/341, 82%; SARS-CoV-2: 30/341, 8.8%; coinfections: 110/341, 32.3%) and 88.9% (361/406) in 2022-23 (ORVs: 339/406, 84%; SARS-CoV-2: 22/406, 5.4%; coinfections: 128/406, 31.5%). In adults, overall RV detection was also lower during pandemic years but increased from 43.7% (333/762) in 2020-21 (ORVs: 26/762, 3.4%; SARS-CoV-2: 307/762, 40.3%; coinfections: 7/762, 0.9%) to 57.8% (731/1265) in 2021-22 (ORVs: 179/1265, 14.2%; SARS-CoV-2: 552/1265, 43.6%; coinfections: 42/1265, 3.3%) and 50.1% (746/1488) in 2022-23 (ORVs: 409/1488, 27.5%; SARS-CoV-2: 337/1488, 22.6%; coinfections: 36/1488, 2.4%). No influenza or RSV was detected in 2020-21; however, their detection increased in the 2 subsequent years but did not reach prepandemic levels. Compared to the prepandemic period, the peaks of RSV hospitalization shifted in 2021-22 (16 weeks earlier) and 2022-23 (15 weeks earlier). Moreover, the peaks of influenza hospitalization shifted in 2021-22 (17 weeks later) and 2022-23 (4 weeks earlier). Age distribution was different compared to the prepandemic period, especially during the first pandemic year.

Conclusions: Significant shifts in viral etiology, seasonality, and age distribution of ARI hospitalizations occurred during the 3 pandemic years. Changes in age distribution observed in our study may reflect modifications in the landscape of circulating respiratory viruses during the COVID-19 pandemic.
RVs and their contribution to ARI hospitalizations. During the pandemic period, SARS-CoV-2 had a low contribution to pediatric ARI hospitalizations, while it was the main contributor to adult ARI hospitalizations during the first 2 seasons and dropped below ORVs during the third pandemic season. Evolving RVs epidemiology underscores the need for increased scrutiny of ARI hospitalization etiology to inform tailored public health recommendations.

(JMIR Public Health Survell 2024;10:e40792) doi: 10.2196/40792

KEYWORDS
respiratory viruses; SARS-CoV-2; COVID-19; hospitalizations; acute respiratory infections; children; adults; coinfections; prepandemic; pandemic

Introduction

Stringent mitigation efforts, such as border closures, travel restrictions, lockdowns, social distancing, use of masks in public spaces, school and business closures, and remote work, had been implemented worldwide to reduce the transmission of SARS-CoV-2 and its impact on hospital bed capacity [1]. While the first weeks of 2020 in the Northern Hemisphere were dominated by respiratory viruses other than SARS-CoV-2 (ORVs), SARS-CoV-2 almost completely replaced seasonally circulating ORVs within several weeks [2-4]. Public health measures in response to the pandemic altered traditional seasonality of some respiratory viruses (RVs), with virtual disappearance of others during extended periods of time in different parts of the world [5-10]. Among children, mitigation efforts led to a decrease in pediatric visits and hospitalizations overall and especially in those having acute respiratory infections (ARIs) [11,12], bronchiolitis [13,14], and pediatric asthma exacerbations associated with ARIs [15]. Easing of public health measures was accompanied by a subsequent surge of ORV circulation [16-18] and gradual return to usual seasonal patterns [19,20]. An interseasonal surge in RSV hospitalizations occurred in 2021 in winter instead of the usual summer season in Australia [16] and during the summer-fall months instead of winter in Canada [20]. In 2022, an unusually late influenza season was observed in the Northern Hemisphere [20] and an earlier than usual RSV hospitalization peak occurred between October and December [19,21].

The pandemic had an impact on laboratory resources, with lower volumes of performed tests and changes in the propensity to test for ORVs, especially during the first months, as well as on health-seeking behaviors, which may complicate the interpretation of ORV surveillance. Detection of ORVs in hospitalized patients may be more informative since admission requires a certain degree of severity and the propensity to be tested for a larger panel of RVs is higher. However, because of the high demand of SARS-CoV-2 tests in hospital laboratories, testing for ORVs was reduced even in hospitalized patients during the first stages of the pandemic.

A number of reports described detection of ORVs in patients hospitalized with ARI or COVID-19 during the pandemic [9,22-26]. However, to our knowledge, no report has described the results of the systematic detection of both ORVs and SARS-CoV-2 (not only at the physician request) using a panel of multiple RVs in a multicenter network including both pandemic years and comparing them with as long as 8 prepandemic years. The characterization of the combined impact of both SARS-CoV-2 and ORVs on ARI hospitalizations during the 3 pandemic years and its comparison with prepandemic seasons may provide insightful information on the postpandemic period when SARS-CoV-2 is expected to cocirculate along with ORVs.

In Quebec, Canada, a prospective hospital-based surveillance network with systematic testing for a panel of 17 RVs in pediatric and adult patients admitted for ARI has been in place since 2012-2013 during periods with high influenza circulation [27-30]. The same network was used for surveillance during the pandemic by adding SARS-CoV-2 to the panel and by extending surveillance periods to intense RV circulation. We report here the results of the comparison of RV contribution to hospitalization during 3 pandemic seasons (2020-21 to 2022-23) and 8 prepandemic seasons.

Methods

Population and Study Design

The characteristics of the Quebec prospective hospital-based surveillance network during the prepandemic years have been described in detail elsewhere [27-30]. In brief, 4 regional hospitals (2 community and 2 academic or tertiary; all of them serving both children and adults) with a catchment area of nearly 10% of the Quebec population (approximately 8.8 million in 2023) participated in the surveillance during 8 influenza seasons since 2012-13. One of the 4 hospitals (approximately 15% of the population included in previous years) was not able to participate in 2020-21 because of challenges with hospital resources during the pandemic, and it rejoined the network in 2021-22. Two additional tertiary hospitals, 1 adult and 1 pediatric hospital, joined the network in 2021-22, for a total of 6 hospitals (approximately 15% of the Quebec population). Results from these 2 hospitals are included only in the description of virus detection per week and are not used for comparison with the prepandemic period. All patients presenting to the emergency department with ARI were systematically swabbed during high influenza activity weeks of prepandemic years or during periods with increased hospitalizations due to SARS-CoV-2 or ORVs during the pandemic (Figure 1).
Figure 1. Weekly number of patients hospitalized for acute respiratory infection (ARI) included in active hospital-based prospective surveillance during the prepandemic (2012-20) and pandemic (2020-23) seasons in Quebec, Canada. Four hospitals participated in the prepandemic period. Three hospitals participated in 2020-21 with periodic sampling by day of the week. Four hospitals participated in 2021-22 and 2022-23 with periodic sampling by day of the week. Two additional hospitals that joined in 2021-22 are not presented.

Inclusion and Exclusion Criteria

Eligible patients were those who were admitted for ≥24 hours and who met a standardized ARI definition (fever or feverishness not attributed to another illness or cough or sore throat) that was expanded in 2020-21 to include symptoms specific for COVID-19 (adjusted from the Canadian Nosocomial Infection Surveillance Program [CNISP] [31]; fever or history of fever not attributed to another illness, or cough [or exacerbation of cough] or difficulty breathing [or exacerbation of difficulty breathing], or sudden extreme fatigue, or at least two of the following symptoms: rhinorrhea or nasal congestion, sore throat, myalgia or arthralgia, or sudden anosmia or ageusia). Nurses collected demographic and clinical details from the patient or legal representative on a standardized questionnaire and reviewed patients’ charts at discharge for additional clinical information. Patients with onset of ARI symptoms after admission, those who refused to consent or were unable to consent (during the period before the ethics committee exemption), those who did not meet the ARI definition, and those who were admitted for less than 24 hours were excluded from this analysis.

Surveillance Period

For the prepandemic years, the surveillance period started when the positivity rate for influenza in respiratory specimens from the provincial sentinel laboratory surveillance was ≥15% for 2 consecutive weeks and stopped the week after this rate dropped below 15% or when the planned sample size for the season was achieved (800-1000 specimens depending on the season). The provincial laboratory surveillance included >40 laboratories across the province of Quebec with >100,000 respiratory specimens per year. Surveillance lasted from 7 to 12 weeks per season (median of 8.5 weeks) between epi-weeks 49 (earliest) and 14 (latest) (Figure 1). In 2020-21, the surveillance period (September 27, 2020 [epi-week 40] to May 29, 2021 [epi-week 21]; overall duration of 35 weeks) coincided with Québec’s second and third COVID-19 waves (caused by ancestral and Alpha SARS-CoV-2 variants) [32] (Figure 2; Multimedia Appendix 1). In 2021-22, the surveillance period (October 04, 2021 [epi-week 40] to June 18, 2022 [epi-week 24]; duration of 34 weeks) captured an unexpected interseasonal RSV surge along with the descending and ascending phases of the fourth (Delta variant), fifth (Omicron BA.1 variant), and sixth (Omicron BA.2 variant) waves. The 2 additional hospitals joined the surveillance in 2021, with the adult hospital starting at epi-week 43 and the pediatric hospital starting at epi-week 46. During 2022-23, the surveillance period lasted 32 epi-weeks (from October 02, 2022 [epi-week 40] to May 20, 2023 [epi-week 20]). Details of the epidemiologic situation in Quebec are presented in Multimedia Appendix 1. Because of challenges with hospital resources during periods of high SARS-CoV-2 circulation, the surveillance was paused during some weeks and sampling of enrollment during predetermined days of the week was adopted by some hospitals (Figure 1).
Figure 2. Proportion of respiratory virus detection among patients hospitalized for acute respiratory infection by epi-week in the hospital-based prospective surveillance network during the prepandemic (2012-20) and pandemic (2020-23) seasons in Québec, Canada. All participating hospitals are included (4 hospitals from 2012-13 to 2019-20, 3 hospitals in 2020-21, and 6 hospitals in 2021-22 and 2022-23). In order to simplify the presentation, epi-week 53 (2014-15 and 2019-20) is excluded. Weeks 40 to 20 are presented. The apparent increase in the proportion of hospitalizations due to influenza during epi-week 14 is due to the small number of hospitalized patients since only 1 season (2015-16) contributed. For consistency with other seasons, data for epi-weeks 40 to 20 during the pandemic seasons are presented, although the surveillance period was longer for some pandemic years.

hCoV: common human coronavirus; hMPV: human metapneumovirus; hPIV: human parainfluenza virus; RSV: respiratory syncytial virus.

Laboratory Analysis
Nasal specimens collected on flocked swabs from eligible patients were sent to the provincial public health laboratory (Laboratoire de Santé Publique du Québec [LSPQ]) and tested using the Luminex NxTAG Respiratory Pathogen Panel assay that detects influenza A (subtypes H3 and H1); influenza B; respiratory syncytial virus (RSV) (A and B differentiated starting in 2016-17); human parainfluenza viruses (hPIVs) 1, 2, 3, and 4; human metapneumovirus (hMPV); common human coronaviruses (hCoVs) NL63, HKU1, 229E, and OC43; enteroviruses/rhinoviruses (not differentiated); adenovirus; bocavirus; and 3 bacteria (Mycoplasma pneumoniae, Chlamydia pneumoniae, and Legionella pneumophila). Nucleic acids were purified using the bioMerieux eMAG platform, and polymerase chain reaction (PCR) products were analyzed on a Luminex Magpix system, as prescribed by the manufacturer. NxTAG assays were approved for diagnosis by Health Canada.

This assay was systematically used during all prepandemic years and during all pandemic years in the hospitals included in the main analysis (comparison between prepandemic and pandemic seasons). Additional assays used by hospitals and contributing only to descriptive results (Figure 2) were as follows: (1) BioFire Respiratory Panel 2.1 (RP2.1) used for ORV testing by local laboratories (considered in the descriptive analysis for patients for whom specimens were not available to be tested by Luminex NxTAG); (2) in-house multiplex reverse transcription PCR (MRVP) detecting influenza A and B; hPIVs 1, 2, and 3; adenovirus; rhinovirus; enterovirus; hCoVs 229E and OC43; RSV; and hMPV (used by the adult center added to the surveillance starting at epi-week 43 in 2021) [33]; and (3) in-house PCR using LightMix Modular Assays according to the manufacturer’s recommendations [34] to detect influenza A and B, RSV, hCoV (not differentiated), hMPV, adenovirus, hPIV (not differentiated), enteroviruses/rhinoviruses (not differentiated), and SARS-CoV-2 (used by the pediatric center added to the surveillance starting at epi-week 46 in 2021). Throughout the pandemic years, SARS-CoV-2 was detected at local laboratories by using commercially available diagnostic tests.

Statistical Analysis
Proportions were compared by using the chi-square or Fisher exact test when appropriate. Mean values were compared by using the Wilcoxon test. The Cochran-Armitage trend test was used to assess the linear trend of proportions across age categories. Statistical significance was set at \( P < .05 \). Statistical analyses were conducted using SAS version 9.4 (SAS Institute). A similar hospitalization rate and viral etiology distribution was assumed for days with and without enrollment during weeks with only 3 enrollment days.

Ethical Considerations
Institutional Review Board approval was obtained from all participating hospitals (Hôpital régional de Rimouski [number: CCER 11-12-13], Hôpital de Chicoutimi [number: 2011-032], Hôpital de la Cité-de-la-Santé – Laval [number: 06.02.02/2011-2012], and Centre hospitalier universitaire...
régional de Trois-Rivières [number: 2011-016-00]) for the first 3 years, and a signed informed consent form, including the possibility of a secondary analysis, was used. A waiver was obtained for the following years when the project was conducted as sentinel surveillance mandated by the Ministry of Health from the Research Ethics Board of the Centre hospitalier universitaire de Québec-Université Laval (2019-4455), and it was considered exempt from the requirement for ethics approval. During pandemic years, surveillance was performed within the legal mandate of the National Director of Public Health of Quebec under the Public Health Act and did not require research ethics committee review. This retrospective analysis used deidentified data. No compensation was provided to patients.

**Results**

**Surveillance Participants**

Overall, 15,199 patients potentially eligible for surveillance were approached (6412 during the prepandemic period, 1454 in 2020-21, 3124 in 2021-22, and 4209 in 2022-23) (Multimedia Appendix 2). Patients missed by nurses (n=25) or those with samples not received by LSPQ or samples of insufficient volume (n=1213) were comparable to those included in the main analysis with respect to age (mean age 61 vs 62 years; \( P=.97 \)) and sex (47% vs 49% female; \( P=.17 \)). A total of 10,550 patients hospitalized for community-acquired ARI were included in the analysis: 5832 (1493 children aged 0-17 years and 4339 adults) during the 8 prepandemic influenza seasons, 791 (29 children and 762 adults) during the 2020-21 season, 1606 (341 children and 1265 adults) during the 2021-22 season, and 1894 (406 children and 1488 adults) during the 2022-23 season (Table 1).

When comparing the age distribution of hospitalized patients from the hospitals that participated in the surveillance since the beginning, the proportion of children was significantly lower in 2020-21 compared to prepandemic seasons (29/791, 3.7% vs 1493/5832, 25.6%; \( P<.001 \)), and it increased and almost reached the levels observed during prepandemic seasons in the following 2 pandemic years (around 21% in both the second [341/1606, 21.2%] and third [406/1894, 21.4%] pandemic years; \( P<.001 \)). During both prepandemic and pandemic seasons, the proportion of young adults among patients hospitalized with ARI was very low (1%-2% for those aged 18-29 years and those aged 30-39 years), and there was a subsequent gradual increase with age to 2%-4% in those aged 40-49 years, 4%-7% in those aged 50-59 years, 13%-19% in those aged 60-69 years, 19%-25% in those aged 70-79 years, and 29%-39% in those aged \( \geq 80 \) years (\( P<.001 \) for the increase in the proportion for both prepandemic and pandemic seasons) (Figure 3). Compared to the prepandemic period, the proportion of patients aged \( \geq 60 \) years was significantly higher during the first pandemic year (653/791, 82.6% vs 3561/5832, 61.1%; \( P<.001 \)) and then decreased to varying degrees during the subsequent years (approximately 70% in 2021-22 [1123/1606, 69.9%] and 2022-23 [1341/1894, 70.8%]) but remained higher than that during the prepandemic period (\( P<.001 \)), mirroring the decrease in pediatric hospitalizations but likely also associated with age-dependent SARS-CoV-2 severity (Figure 3).
Table 1. Number and proportion of patients hospitalized for acute respiratory infection by age group and detected respiratory virus in Quebec, Canada, during prepandemic (2012-20) and pandemic (2020-23) seasons.

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>Test ed, n</th>
<th>≥1 virus positive, n (%)</th>
<th>≥1 virus positive, n (%)</th>
<th>SARS-CoV-2, with or without another virus, n (%)</th>
<th>Other virus without SARS-CoV-2, n (%)</th>
<th>Test ed, n</th>
<th>≥1 virus positive, n (%)</th>
<th>SARS-CoV-2, with or without another virus, n (%)</th>
<th>Other virus without SARS-CoV-2, n (%)</th>
<th>Test ed, n</th>
<th>≥1 virus positive, n (%)</th>
<th>SARS-CoV-2, with or without another virus, n (%)</th>
<th>Other virus without SARS-CoV-2, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-17</td>
<td>(25.6)</td>
<td>(92.7)</td>
<td>(3.7)</td>
<td>(58.6)</td>
<td>(0)</td>
<td>(21.2)</td>
<td>(90.3)</td>
<td>(8.8)</td>
<td>(1.5)</td>
<td>(21.4)</td>
<td>(88.9)</td>
<td>(5.4)</td>
<td>(83.5)</td>
</tr>
<tr>
<td>18-29</td>
<td>(1.4)</td>
<td>(63.3)</td>
<td>(1.5)</td>
<td>(25.0)</td>
<td>(0.6)</td>
<td>(10)</td>
<td>(60.0)</td>
<td>(40.0)</td>
<td>(20.0)</td>
<td>(1.0)</td>
<td>(57.9)</td>
<td>(21.1)</td>
<td>(36.8)</td>
</tr>
<tr>
<td>30-39</td>
<td>(2.2)</td>
<td>(57.7)</td>
<td>(1.6)</td>
<td>(46.2)</td>
<td>(0)</td>
<td>(26)</td>
<td>(61.5)</td>
<td>(46.2)</td>
<td>(15.4)</td>
<td>(0.9)</td>
<td>(58.8)</td>
<td>(17.6)</td>
<td>(41.2)</td>
</tr>
<tr>
<td>40-49</td>
<td>(2.7)</td>
<td>(65.0)</td>
<td>(3.7)</td>
<td>(55.2)</td>
<td>(2.2)</td>
<td>(35)</td>
<td>(65.7)</td>
<td>(54.3)</td>
<td>(11.4)</td>
<td>(1.7)</td>
<td>(65.6)</td>
<td>(25.0)</td>
<td>(40.6)</td>
</tr>
<tr>
<td>50-59</td>
<td>(7.0)</td>
<td>(65.3)</td>
<td>(7.0)</td>
<td>(50.9)</td>
<td>(4.4)</td>
<td>(71)</td>
<td>(53.5)</td>
<td>(16.9)</td>
<td>(4.2)</td>
<td>(45.6)</td>
<td>(15.2)</td>
<td>(30.4)</td>
<td></td>
</tr>
<tr>
<td>60-69</td>
<td>(12.8)</td>
<td>(60.8)</td>
<td>(18.6)</td>
<td>(32.7)</td>
<td>(29.9)</td>
<td>(48)</td>
<td>(49.2)</td>
<td>(35.4)</td>
<td>(13.8)</td>
<td>(12.9)</td>
<td>(42.0)</td>
<td>(16.3)</td>
<td>(25.7)</td>
</tr>
<tr>
<td>70-79</td>
<td>(19.1)</td>
<td>(61.0)</td>
<td>(24.5)</td>
<td>(40.2)</td>
<td>(29.9)</td>
<td>(382)</td>
<td>(52.9)</td>
<td>(38.2)</td>
<td>(14.7)</td>
<td>(23.9)</td>
<td>(44.4)</td>
<td>(20.3)</td>
<td>(24.1)</td>
</tr>
<tr>
<td>≥80</td>
<td>(29.3)</td>
<td>(64.3)</td>
<td>(39.4)</td>
<td>(49.4)</td>
<td>(44.9)</td>
<td>(481)</td>
<td>(63.6)</td>
<td>(50.1)</td>
<td>(13.8)</td>
<td>(33.9)</td>
<td>(56.6)</td>
<td>(27.7)</td>
<td>(28.9)</td>
</tr>
<tr>
<td>≥60</td>
<td>(61.1)</td>
<td>(62.5)</td>
<td>(82.6)</td>
<td>(42.0)</td>
<td>(39.4)</td>
<td>(1123)</td>
<td>(56.6)</td>
<td>(42.7)</td>
<td>(14.0)</td>
<td>(70.8)</td>
<td>(49.8)</td>
<td>(23.1)</td>
<td>(26.7)</td>
</tr>
<tr>
<td>Total</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(44.2)</td>
<td>(38.8)</td>
<td>(1606)</td>
<td>(64.7)</td>
<td>(36.2)</td>
<td>(28.5)</td>
<td>(100)</td>
<td>(58.4)</td>
<td>(19.0)</td>
<td>(39.5)</td>
</tr>
</tbody>
</table>

aP<.05 for comparison with 2012-20; Fisher exact test.
bP<.05 for comparison between the first pandemic year (2020-21) and each of the subsequent years (2021-22 and 2022-23).

Figure 3. Age distribution of patients hospitalized for acute respiratory infection in the hospital-based prospective surveillance network in the prepandemic period (peaks of influenza seasons from 2012-13 to 2019-20) and 3 pandemic years (2020-21, 2021-22, and 2022-23) in Quebec, Canada.
Viral Etiology

Overall Hospitalizations

During the prepandemic seasons, at least one respiratory virus was detected in 70.4% (4107/5832) of the patients hospitalized for ARI (Table 1). The most frequently detected viruses across all age groups were influenza (2107/5832, 36.1%), RSV (1080/5832, 18.5%), hMPV (475/5832, 8.1%), enteroviruses/rhinoviruses (465/5832, 7.9%), and hCoV (376/5832, 6.4%). Coinfections were detected in 12.3% (720/5832) of patients.

In 2020-21, the most frequently detected virus was SARS-CoV-2 (307/791, 38.8%). ORVs were detected in 5.4% (43/791) of patients, and the most frequent were enteroviruses/rhinoviruses (26/791, 3.3%), hMPV (11/791, 1.4%), and adenoviruses (10/791, 1.3%). No influenza or RSV was detected. Coinfections were detected in 1.8% (14/791) of patients; 1.0% (8/791) involved a combination of ORVs without SARS-CoV-2 and 0.8% (6/791) involved SARS-CoV-2 with ORVs.

In 2021-22, the most frequently detected viruses were SARS-CoV-2 (582/1606, 36.2%), RSV (162/1606, 10.1%), enteroviruses/rhinoviruses (153/1606, 9.5%), influenza viruses (61/1606, 3.8%; all being influenza A [H3N2]), adenoviruses (59/1606, 3.7%), and bocaviruses (55/1606, 3.4%). ORVs (excluding those with SARS-CoV-2 coinfection) were detected in 28.5% (457/1606) of patients. Coinfections were detected in 9.5% (152/1606) of patients: 7.3% (118/1606) involved a combination of ORVs without SARS-CoV-2 and 2.1% (34/1606) involved SARS-CoV-2 with ORVs. The peak of hospitalization due to RSV was detected 16 weeks earlier (epi-week 40) compared to prepandemic seasons (on average, epi-week 4) (Figure 2). The peak of hospitalization due to influenza occurred 17 weeks later in 2021-22 (epi-week 17) than the average peak at epi-week 52 during prepandemic seasons (Figure 2).

In 2022-23, the most frequently detected viruses were SARS-CoV-2 (359/1894, 19.0%), RSV (215/1894, 11.4%), influenza viruses (134/1894, 7.1%; mostly influenza A [H3N2] at 88.8% [119/134]), and enteroviruses/rhinoviruses (213/1894, 11.2%). ORVs without SARS-CoV-2 were detected in 39.5% (748/1894) of patients. Coinfections were detected in 8.7% (164/1894) of patients: 6.9% (130/1894) involved a combination of ORVs without SARS-CoV-2 and 1.8% (34/1894) involved SARS-CoV-2 with ORVs. The timing of influenza and RSV hospitalization was more aligned with historical time frames, although differences still occurred (influenza peaked 4 weeks earlier [epi-week 48 vs epi-week 52]; RSV peaked 15 weeks earlier [epi-week 41 vs epi-week 4]) (Figure 2).

Pediatric Hospitalizations

During the prepandemic period, at least one RV was detected in 92.7% (1384/1493) of children, including 30.5% (456/1493) coinfections. Pediatric hospitalizations due to RVs sharply decreased during the first pandemic year. The detection rate decreased to 58.6% (17/29) in 2020-21 (7/29, 24.1% being coinfections) and reverted to a rate comparable to that in the prepandemic period in 2021-22 (308/341, 90.3%) and 2022-23 (361/406, 88.9%). A similar trend was observed for coinfections, with a decrease to 24.1% (7/29) in 2020-21 and an increase to a rate closer to that in the prepandemic years during the 2 subsequent years (110/341, 32.3% in 2021-22 and 128/406, 31.5% in 2022-23).

Prior to the COVID-19 pandemic, the most frequently detected viruses in hospitalized children were RSV (721/1493, 48.3%) and influenza (365/1493, 24.4%) (Table 2). In 2020-21, SARS-CoV-2, RSV, and influenza were not detected in any hospitalized child, and hospitalizations with an identified virus in children were due to enteroviruses/rhinoviruses (8/29, 27.6%), adenoviruses (7/29, 24.1%), and bocaviruses (6/29, 20.7%). The proportions of detected adenoviruses and bocaviruses were significantly higher than those reported during the prepandemic seasons (Table 2), and no significant differences were detected for the rest of the viruses. In the 2 subsequent pandemic years, RSV was again the most frequently detected virus in children, while influenza and SARS-CoV-2 had a low etiological contribution to pediatric ARI hospitalizations. In 2022-23, most of the hospitalizations in children were due to ORVs without SARS-CoV-2 (278/341, 81.5%), and only 8.8% (30/341) were due to SARS-CoV-2 (including 10 coinfections with ORVs). The predominant virus was RSV (115/341, 33.7%), followed by enteroviruses/rhinoviruses (110/341, 32.3%) and bocaviruses (49/341, 14.4%). Compared to prepandemic seasons, the contribution was significantly higher for enteroviruses/rhinoviruses, hPIV 1-4, and bocaviruses, and lower for RSV, influenza, and hCoV (Table 2). In 2022-23, most of the hospitalizations were due to ORVs (339/406, 83.5%), and 5.4% (22/406) were due to SARS-CoV-2 (including 13 coinfections with ORVs). The predominant virus was again RSV (158/406, 38.9%), followed by enteroviruses/rhinoviruses (114/406, 28.1%), bocaviruses (53/406, 13.1%), hMPV (39/406, 9.6%), hCoV (38/406, 9.4%), adenoviruses (35/406, 8.6%), influenza (29/406, 7.1%), and hPIV (26/406, 6.4%). Compared to prepandemic seasons, significantly less cases of RSV and influenza and more cases of enteroviruses/rhinoviruses, bocaviruses, and hPIV were detected.

Some differences in the age distribution of children hospitalized before or during the pandemic were observed regarding some etiological viruses. For example, the average age among children hospitalized with RSV was 16 months in 2022-23, which was higher than in the prepandemic period (8 months) and in 2021-22 (7 months; P<.001). Differences in the average age of children hospitalized with enteroviruses/rhinoviruses were also observed but to a lower degree. It was higher in 2022-23 (21 months) and 2021-22 (20 months) than in the prepandemic period (17 months; P<.001).
Table 2. Children hospitalized for acute respiratory infection and detected viruses in Quebec, Canada, during prepandemic (2012-20) and pandemic (2020-23) seasons.

<table>
<thead>
<tr>
<th>Variable</th>
<th>2012-20 (4 hospitals)</th>
<th>2020-21 (3 hospitals)</th>
<th>2021-22 (4 hospitals)</th>
<th>2022-23 (4 hospitals)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of included patients</td>
<td>187 (113-322)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>At least one respiratory virus (not mutually exclusive)</td>
<td>173 (103-307)</td>
<td>1384 (92.7)</td>
<td>17 (58.9)</td>
<td>308 (90.3)</td>
</tr>
<tr>
<td>Influenza, any</td>
<td>45.5 (23-96)</td>
<td>365 (24.4)</td>
<td>0 (0)</td>
<td>24 (7.0)</td>
</tr>
<tr>
<td>Influenza A</td>
<td>34.3 (22-57)</td>
<td>274 (18.4)</td>
<td>0 (0)</td>
<td>24 (7.0)</td>
</tr>
<tr>
<td>H1N2</td>
<td>15.3 (0-32)</td>
<td>122 (8.2)</td>
<td>0 (0)</td>
<td>24 (7.0)</td>
</tr>
<tr>
<td>H1N1</td>
<td>18.5 (0-49)</td>
<td>148 (9.9)</td>
<td>0 (0)</td>
<td>3 (0.7)</td>
</tr>
<tr>
<td>A unsubtyped</td>
<td>0.5 (0-2)</td>
<td>4 (0.3)</td>
<td>0 (0)</td>
<td>1 (0.2)</td>
</tr>
<tr>
<td>Influenza B</td>
<td>114 (0-61)</td>
<td>91 (6.1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>RSV</td>
<td>90.1 (44-172)</td>
<td>721 (48.3)</td>
<td>0 (0)</td>
<td>115 (33.7)</td>
</tr>
<tr>
<td>Adenovirus</td>
<td>10.9 (4-20)</td>
<td>87 (5.8)</td>
<td>7 (24.1)</td>
<td>28 (8.2)</td>
</tr>
<tr>
<td>hMPV</td>
<td>20.4 (0-38)</td>
<td>163 (10.9)</td>
<td>0 (0)</td>
<td>40 (11.7)</td>
</tr>
<tr>
<td>hPIV 1-4</td>
<td>7.3 (0-13)</td>
<td>58 (3.9)</td>
<td>0 (0)</td>
<td>30 (8.8)</td>
</tr>
<tr>
<td>hCoV</td>
<td>19.3 (7-37)</td>
<td>154 (10.3)</td>
<td>3 (10.3)</td>
<td>14 (4.1)</td>
</tr>
<tr>
<td>Enteroviruses/rhinoviruses</td>
<td>30.0 (22-48)</td>
<td>240 (16.1)</td>
<td>8 (27.6)</td>
<td>110 (32.3)</td>
</tr>
<tr>
<td>Bocaviruses</td>
<td>16.9 (6-33)</td>
<td>135 (9.0)</td>
<td>6 (20.7)</td>
<td>49 (14.4)</td>
</tr>
<tr>
<td>SARS-CoV-2</td>
<td>N/A</td>
<td>N/A</td>
<td>0 (0)</td>
<td>30 (8.8)</td>
</tr>
</tbody>
</table>

Respiratory viruses without SARS-CoV-2

| Mono-infection | 116 (69-231) | 928 (62.2) | 10 (34.5) | 178 (52.2) |
| Coinfection, any RV without SARS-CoV-2 | 57 (27-91) | 456 (30.5) | 7 (24.1) | 100 (29.3) |

SARS-CoV-2

| Mono-infection | N/A | N/A | 0 (0) | 20 (5.9) |
| Coinfection, SARS-CoV-2 + any RV | N/A | N/A | 0 (0) | 10 (2.9) |
| Total co-infections | 57 (27-91) | 456 (30.5) | 7 (24.1) | 110 (32.2) |

Notes:
- N/A: not applicable.
- P<.05 for comparison with 2012-20; Fisher exact test.
- RSV: respiratory syncytial virus.
- hMPV: human metapneumovirus.
- hPIV: human parainfluenza virus.
- hCoV: common human coronavirus.
- RV: respiratory virus.
- SARS-CoV-2 in co-infection with adenoviruses (n=4), hMPV (n=2), and enteroviruses/rhinoviruses (n=2).
- SARS-CoV-2 in co-infection with RSV (n=4), adenoviruses (n=3), bocaviruses (n=2), hCoV (n=2), hMPV (n=1), and enteroviruses/rhinoviruses (n=1).

**Adult Hospitalizations**

During the prepandemic seasons, at least one RV was detected in 62.8% (2723/4339) of adults (264/4339, 6.1% co-infections), and influenza was the most predominant virus (1742/4339, 40.1% globally and 1742/2723, 63.9% of all detected viruses). During all pandemic years, significantly fewer adults (1810/3515, 51.5%) were positive for at least one RV (including SARS-CoV-2 and ORVs) in comparison to the prepandemic...
period (2723/4339, 62.8%). With a few exceptions, the detection rate was comparable to prepandemic years in the younger age groups (18-59 years) and was lower in those aged ≥60 years (Table 1). SARS-CoV-2 was the most frequently identified virus during all 3 pandemic years among adults hospitalized for ARI (2020-21: 307/762, 40.3%; 2021-22: 552/1265, 43.6%; and 2022-23: 337/1488, 22.6%) (Tables 1 and 3). At least one ORV was detected in 3.4% (26/762), 14.2% (179/1265), and 27.5% (409/1488) of adults in 2020-21, 2021-22, and 2022-23, respectively. Except for enteroviruses/rhinoviruses and hPIV in 2022-23, all individual viruses were detected in lower proportions during the pandemic period compared to the prepandemic period, although differences did not always reach statistical significance owing to low numbers (Table 3).

A sensitivity analysis, which excluded comparisons of the hospital that did not participate during the first pandemic year, did not reveal differences in detected trends (data not presented).

The results of viral detection in the 2 additional hospitals in 2021-22 and 2022-23 compared to the 4 hospitals included in the main analysis are presented in Multimedia Appendix 3. Differences might be explained by the different assays used for laboratory analyses, the populations (only pediatric patients in one hospital and only adults in the other hospital), and the time spans. While comparisons with prepandemic seasons are not appropriate in this context, we included results from the additional hospitals in the presentation of viral etiology by week in order to illustrate a more comprehensive impact of RVs on hospitalization (Figure 2).
Table 3. Adults hospitalized for acute respiratory infection and detected viruses in Quebec, Canada, during prepandemic (2012-20) and pandemic (2020-23) seasons.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Peak of the 2012-20 influenza season (4 hospitals)</th>
<th>2020-21 (3 hospitals)</th>
<th>2021-22 (4 hospitals)</th>
<th>2022-23 (4 hospitals)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Detection rate (N=4339), n (%)</td>
<td>Detection rate (N=762), n (%)</td>
<td>Detection rate (N=1265), n (%)</td>
<td>Detection rate (N=1488), n (%)</td>
</tr>
<tr>
<td>Number of included patients</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>At least one respiratory virus (not mutually exclusive)</td>
<td>340.4 (224-425)</td>
<td>2723 (62.8)</td>
<td>333 (43.7)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>731 (57.8)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>746 (50.1)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>731 (57.8)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>333 (43.7)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>746 (50.1)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Influenza, any</td>
<td>217.1 (132-334)</td>
<td>1742 (40.1)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>37 (2.9)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>185.4 (109-324)</td>
<td>1483 (34.2)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>37 (2.9)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>125.8 (2-316)</td>
<td>1006 (23.2)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>37 (2.9)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>57.0 (0-123)</td>
<td>456 (10.5)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>2.6 (0-8)</td>
<td>21 (0.5)</td>
<td>0 (0)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Influenza B</td>
<td>32.4 (1-146)</td>
<td>259 (6.0)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>44.9 (19-75)</td>
<td>359 (8.3)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>47 (3.7)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>RSV&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.6 (0-5)</td>
<td>13 (0.3)</td>
<td>3 (0.4)</td>
<td>31 (2.5)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>39.0 (2-103)</td>
<td>312 (7.2)</td>
<td>11 (1.4)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>20 (1.6)</td>
</tr>
<tr>
<td>Adenovirus</td>
<td>14.5 (3-25)</td>
<td>116 (2.7)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>18 (1.4)</td>
</tr>
<tr>
<td></td>
<td>27.8 (7-49)</td>
<td>222 (5.1)</td>
<td>0 (0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>17 (1.3)</td>
</tr>
<tr>
<td>Enteroviruses/rhinoviruses</td>
<td>27.5 (12-43)</td>
<td>220 (5.1)</td>
<td>18 (2.4)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>43 (3.4)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Bocaviruses</td>
<td>2.5 (0-6)</td>
<td>20 (0.5)</td>
<td>2 (0.3)</td>
<td>6 (0.5)</td>
</tr>
<tr>
<td></td>
<td>2.72 (1.9-3.5)</td>
<td>2.64 (1.1)</td>
<td>1 (0.1)</td>
<td>18 (1.4)</td>
</tr>
<tr>
<td></td>
<td>33 (12-76)</td>
<td>264 (6.1)</td>
<td>1 (0.1)</td>
<td>18 (1.4)</td>
</tr>
<tr>
<td>SARS-CoV-2</td>
<td>N/A</td>
<td>N/A</td>
<td>307 (40.3)</td>
<td>552 (43.6)</td>
</tr>
<tr>
<td></td>
<td>33 (12-76)</td>
<td>264 (6.1)</td>
<td>7 (0.9)</td>
<td>42 (3.3)</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.
<sup>b</sup>P<.05 for comparison with 2012-20; Fisher exact test.
<sup>c</sup>RSV: respiratory syncytial virus.
<sup>d</sup>hMPV: human metapneumovirus.
<sup>e</sup>hPIV: human parainfluenza virus.
<sup>f</sup>hCoV: common human coronavirus.
<sup>g</sup>RV: respiratory virus.
<sup>h</sup>SARS-CoV-2 in coinfection with hMPV (n=4), adenoviruses (n=2), enteroviruses/rhinoviruses (n=1), and bocaviruses (n=1).
<sup>i</sup>SARS-CoV-2 in coinfection with adenoviruses (n=13), hPIV (n=1), hCoV (n=3), enteroviruses/rhinoviruses (n=4), influenza A (H3N2; n=3), and RSV (n=1).
<sup>j</sup>SARS-CoV-2 in coinfection with hCoV (n=6), hMPV (n=5), enteroviruses/rhinoviruses (n=3), adenoviruses (n=2), influenza A (H3N2; n=2), hPIV (n=2), and RSV (n=1).
Discussion

Principal Findings

Our report, based on the systematic detection of RVs, describes changes in the etiology of pediatric and adult ARI hospitalizations during 3 pandemic years compared with 8 prepandemic winter seasons in the same population. To our knowledge, this is the longest time span of follow-up during both pandemic and prepandemic periods in both pediatric and adult hospitalized patients. We detected continuing changes in viral etiology and age distribution during the 3 pandemic years compared to the prepandemic period. While SARS-CoV-2 was the most frequent viral etiology of ARI hospitalizations during the 3 pandemic years, it was absent among children during the first year and scarcely detected during the second and third years. ORVs were the most important contributors to pediatric ARI hospitalization during all 3 pandemic years. In adults, SARS-CoV-2 was the most important contributor to ARI hospitalization during the first 2 pandemic years, but its relative importance gradually decreased, mirroring the increasing role of ORVs, and dropped below ORVs’ detection during the third pandemic year. The most striking differences in age distribution were detected during the first pandemic year, mainly due to a remarkable decrease in pediatric ARI hospitalization.

Interpretation of the Findings

The differences in the etiology and age distribution of ARI hospitalizations were a consequence of changes in both ORV and SARS-CoV-2 circulation and their impacts on hospitalizations, reflecting the intensity of mitigation measures and the resulting modification of immunity in the population following extended periods of the absence of some viruses (such as RSV). For instance, the circulation of ORVs was very low in 2020-21 when more stringent measures were implemented, increased during autumn 2021 following the easing of some measures, declined in December 2021 following the tightening of the measures in response to the Omicron wave, and increased again in February-March 2022 following broader travelling and school opening in Quebec (Multimedia Appendix 1) [35]. The increase in the average age of children hospitalized with RSV and enteroviruses/rhinoviruses during the pandemic period suggests that younger children had less exposure to these viruses due to the mitigation measures deployed in 2020-21. Another potential contributing factor was the evolution of SARS-CoV-2 variants. For example, during the 2020-21 season when ancestral and Alpha SARS-CoV-2 variants (affecting mostly adults) circulated, the pediatric population was spared, while during the autumn and winter of 2021-22 when Delta and then Omicron variants predominated [32], the pediatric population was more affected. Finally, COVID-19 vaccine uptake and effectiveness, and outpatient antiviral availability, which had an impact on preventing hospitalizations, varied by age group. During the second pandemic year, older patients were prioritized in the COVID-19 vaccination campaigns, and more patients with comorbidities were offered outpatient antiviral treatment. For children, however, COVID-19 vaccines were available later, and in young adults, vaccine uptake was lower than in the older population [36]. Therefore, the impact on SARS-CoV-2 hospitalizations varied by age and time period. The effectiveness of COVID-19 vaccines and their impact on the SARS-CoV-2 portion of ARI hospitalizations varied depending on the type of vaccine, the frequency of doses and delays between dose administrations, and the proportion of the population with prior COVID-19 infection, which steadily increased with the unfolding of the pandemic, allowing the development of hybrid immunity [37].

Overall, during the prepandemic influenza seasons, the 2 most frequently detected viruses among patients hospitalized with ARI were influenza and RSV. SARS-CoV-2 was the most important respiratory virus during the first pandemic year. In 2021-22 and 2022-23, its contribution decreased, while the contribution of ORVs increased. However, there were differences between the pediatric and adult populations. RVs (mostly RSV) affected more children than adults during the prepandemic winter seasons. Although the overall impact of ORVs was lower during the pandemic period, children remained mostly affected by ORVs and not by SARS-CoV-2 during all pandemic years. In adults, influenza was the most important virus during the prepandemic years, whereas SARS-CoV-2 was more important than ORVs during the first 2 pandemic years, while their roles reversed during the third pandemic year. It is of note that with the addition of the contribution of SARS-CoV-2 to ARI hospitalizations, the relative role of ORVs decreased during the pandemic years compared to the prepandemic period both in children and adults [38,39], with some minor exceptions (eg, enteroviruses/rhinoviruses) [40,41].

Nonenveloped viruses (rhinoviruses and adenoviruses) were less impacted by preventive measures aiming to halt SARS-CoV-2 transmission [40-42]. Their detection has been reported in hospitalized patients during periods of decreased detection of other RVs [43,44]. In our study, among children, the proportion of adenoviruses and enteroviruses/rhinoviruses was higher during the first pandemic year compared to the prepandemic period, and the proportion of enteroviruses/rhinoviruses remained higher during the second and third pandemic years. Among adults, the proportion of enteroviruses/rhinoviruses was higher during the third pandemic year. Since our study did not include less severe infections not leading to hospitalization, our findings reflect not only the circulation of these viruses, but also their virulence that may vary by age and season. It is of note that the proportion of enteroviruses/rhinoviruses might have been underestimated in the prepandemic period of our study because it was conducted only during the peaks of influenza circulation, and the weeks of the most intense circulation of enteroviruses/rhinoviruses may not have been captured.

The important reduction of pediatric respiratory hospitalization, virtual absence of influenza and RSV during the first pandemic year, and increased role of COVID-19 and ORVs, especially RSV, in hospitalized children during the second and third pandemic years in our study are in line with reports from some other countries and other Canadian provinces [14,23,24,45,46]. Of note, the timing of the unusual surge of RSV varied by region of the world. In Europe, Israel, and the United States, increases were reported in the Spring-Summer of 2021 [14,23,24,45]. Major shifts in the epidemiology of RSV with large-scale outbreaks were reported in New Zealand (corresponding to the
usual season) [47] and Australia (out of season) [16]. In Canada, the first province to report a surge of RSV was Quebec in August 2021, while the other provinces followed with a delay of 2 to 3 months [46]. In 2022, an earlier than usual start of the RSV season was reported in the Northern Hemisphere [17,21], with a peak of RSV-related hospitalizations between October and December [19,21]. The surge in RSV hospitalizations and the shift in the average age of children hospitalized with RSV (older in 2021-22) observed in our study are consistent with the results of other studies [48,49], and the findings might be attributable to the “immunity debt” that occurred because of the lack of exposure during the first pandemic year. However, it is not clear why important variability in the increase of the RSV hospitalization rate occurred globally, for example, it was relatively comparable to the prepandemic period in Germany but higher in the United States [19,21]. After a prolonged absence, the influenza season of 2021-22 occurred later than during prepandemic years in the Northern Hemisphere and affected mostly adults [20,50,51]. A shorter season than the typical prepandemic season was reported in Canada and Europe [51], while a longer season occurred in the United States [52]. During the third pandemic year (2022-23), an unusually early influenza season was reported in other Canadian provinces [20], Europe [53], and the United States [54], which is similar to our findings.

Evidence from across the world suggests that RV-specific seasonality is being progressively re-established [17,20,53,55]. However, it is yet unclear how the co-circulation of different viruses, now with the addition of SARS-CoV-2, will impact the occurrence and severity of ARI hospitalizations. Systematic testing for a panel of RVs allowed us to detect a high proportion of coinfection (>30%, including both SARS-CoV-2 and ORVs) in children during both the prepandemic and pandemic periods (with the exception of a somewhat lower proportion in the first pandemic year) and <10% coinfection in adults. However, our analysis did not aim to study the severity of coinfection or potential viral interference since we did not assess the expected coinfection rate, which requires additional epidemiological data. Viral interference between RVs (positive [additive or synergistic] or negative [antagonistic]) has been demonstrated at the cellular, host, and population levels before the pandemic [56]. SARS-CoV-2 could also interact with existing RVs. For example, it has been reported in a retrospective population-based cohort study that children with prior SARS-CoV-2 infection are more likely to have an RSV infection [57]. Moreover, rhinovirus infection may reduce the likelihood of SARS-CoV-2 infection according to in vitro assays [56].

Study Limitations

This study had some limitations. First, the prepandemic surveillance occurred during the peaks of influenza activity, and therefore, the relative contribution of other RVs may be underestimated as compared to the entire winter season. Surveillance periods during the pandemic years were mostly tailored to the increase in respiratory hospitalizations following increased SARS-CoV-2 or ORV circulation and were much longer. However, we believe that the comparison is still valid because it includes periods with the most strain on hospital capacity due to intensive circulation of ORVs, SARS-CoV-2, or both. Second, the ARI definition used during the pandemic years was broader and less specific for some RVs than during the prepandemic years and may have contributed to lower detection of RVs. In addition, it may have contributed to an increase in older age groups in which symptoms, such as exacerbation of difficulty breathing and sudden extreme fatigue, may be associated with other nonrespiratory conditions. On the other hand, this broad definition may have decreased the probability of missing patients in whom RVs may have contributed to the deterioration of their condition. Third, the first pandemic year was limited to only 3 hospitals, and periodic pauses and sampling of enrollment during the 2 pandemic years were necessary given the stretched resources, which limited the sample size. However, this should not have influenced the relative contribution of RVs and global comparisons, as the surveillance period during pandemic years was much longer than prepandemic seasons and sensitivity analysis did not reveal differences in trends when excluding the hospital not included in the first pandemic year. In addition, this approach allowed the maintenance of surveillance in the context of resource challenges during the pandemic. Fourth, 19% of eligible patients hospitalized for ARI were excluded from the main analysis. However, we do not believe that this impacted the validity of comparisons made in this study given that age was the main factor driving differences in RV detection and there was no difference by age and gender in missed eligible patients compared to those included in the study. Finally, our results may not necessarily be extrapolated to other regions or periods of time because of temporal and geographical differences in ORVs and SARS-CoV-2 epidemiology.

Study Strengths

The main strength of this study was the systematic testing for a broad panel of RVs for all admitted patients with symptoms of ARI during surveillance periods, by using the same diagnostic assay during all study periods. Moreover, it included both pediatric and adult populations during a total of 11 years, with the participation of at least three hospitals from different Quebec regions in each year, a broad case definition, a comprehensive detection of 17 RVs and SARS-CoV-2 potentially contributing to ARI hospitalizations, and a possibility to distinguish community-acquired infections from health care–acquired infections.

Conclusion

Important shifts in viral etiology, seasonality, and age distribution of ARI hospitalizations in children and adults were observed during the 3 pandemic years compared to the prepandemic period. While the first pandemic season was significantly different from the prepandemic winter seasons, the second and third years were more comparable in terms of both RV contribution and age distribution. The complex interplay among mitigation measures, intrinsic seasonality and secular trends of ORVs, changes in circulating SARS-CoV-2 variants and their virulence, COVID-19 vaccine uptake and effectiveness, outpatient antiviral treatments, and potential viral interference may have played roles in the differential contribution of ORVs and SARS-CoV-2 to ARI hospitalizations. Our study underscores the importance of surveillance in
understanding altered seasonal patterns of RVs and shows that the role of SARS-CoV-2 relative to ORVs is continually changing. The current situation may reflect a transition period until SARS-CoV-2 finds its ecological niche in the human population and ORVs re-establish their usual seasonal patterns. Although new SARS-CoV-2 variants may emerge and cause occasional increases in hospitalizations, in the long run, it may establish itself as another usual respiratory virus. At this point, it is difficult to foresee its role compared to ORVs; however, our study suggests that SARS-CoV-2 may continue to be the most important contributor to ARI hospitalizations in adults. Increased scrutiny of continuing changes in the etiology of ARI hospitalizations by using systematic multiplex testing approaches that allow valid comparisons, including assessment of observed and expected coinfections, is needed to inform mathematical modeling and appropriate public health recommendations.

Acknowledgments
We acknowledge France Bouchard who coordinated the surveillance and Sophie Auger who entered and cleaned the collected data throughout the years. We are extremely grateful to the front-line surveillance staff from the participating hospitals who collected and provided data in challenging circumstances during the pandemic period: Ménard Francois, Gagnon Maude, Desbiens Karine, Tremblay Chantale, Simard Patricia, Murry Carole, and Rhainds Jennifer from Chicoutimi Hospital; Estel Debois, Alexandra Bouffard, Alexandra Fortier, Alexandra Mondor, Rosalie Beaudoin, Kristina Boucher, Mélyna Carrier, Brigitte Dion, Marie-Ève Chamberland, and Lise Anne Paradis from Hôtel-Dieu de Lévis Hospital; Dolcé Patrick, Gagnon Isabelle, Lévesques Julie, Bernatchez Isabelle, and Lévesques Janie from Rimouski Hospital; Poirier André, Danylo Alexis, Tapps Danielle, Loranger Josée, and Toupin Guylaine from Trois-Rivières Hospital; Wadas Katerin, Lafleur Caroline, Ane Tres Silicia, and Tomas Fernanda from Centre universitaire de santé McGill; and Thibeault Roseline, Jacob-Wagner Marieve, Hamelin Marie-Eve, Theriault Ariane, Pelletier-Bélanger Joannie, and Côté Claudia from Centre hospitalier de l’Université Laval. We also thank Desautels Lyne, Martinneau Christine, and Ménard Joël from Laboratoire de Santé Publique du Québec who supported the surveillance despite the exponentially increasing demand on laboratories during the pandemic. We also acknowledge the support of Hany Geagea and Lauriane Padet for revising the literature. This work was supported by the Ministère de la Santé et des Services sociaux du Québec. The sponsor was not involved in the study design, data collection, result interpretation, and drafting of the manuscript.

Data Availability
The data collected in this surveillance study are the property of the Quebec Ministry of Health (Ministère de la Santé et des Services sociaux du Québec) and have been shared with the research group under the legal mandate of the National Director of Public Health of Quebec under the Public Health Act precluding data sharing with a third party. Aggregate data are available within the manuscript and the multimedia appendix files. Real-time surveillance data from HospiVir are available online [58].

Authors’ Contributions
RG, RA, and SC conceived and designed the study. RA, CF, GB, and JD collected the data. RA, SC, HC, CF, GB, and JD contributed to the data or analysis tools. RD added the results of the third year to the revised version of the manuscript and updated the literature search included in the revised version. RA, SC, and HC performed the statistical (RA and SC) and laboratory (HC) analyses. All authors contributed to the interpretation of the results. RG drafted the manuscript. All authors revised the paper and approved the final version.

Conflicts of Interest
RG, RA, and SC report that the Ministère de la Santé et des Services sociaux du Québec provided financial support to their institution for this work. RG reports personal fees from Abbie (honorary for a conference on respiratory syncytial virus burden in children unrelated to this work). The other authors have nothing to disclose.

Multimedia Appendix 1
Timeline of SARS-CoV-2 epidemiology and main mitigation measures in Quebec, Canada.

[DOCX File, 42 KB - publichealth_v10i1e40792_app1.docx]

Multimedia Appendix 2
Surveillance flowchart by period in the hospitals participating in surveillance during the prepandemic (2012-2020) and pandemic (2020-2023) seasons in Québec, Canada.

[DOCX File, 25 KB - publichealth_v10i1e40792_app2.docx]

Multimedia Appendix 3
Results of viral detection in patients hospitalized for acute respiratory infections in 2021-22 and 2022-23 (4 hospitals participating during the prepandemic period and 2 additional hospitals) in Québec, Canada.

References


19. RSV-NET Interactive Dashboard. CDC. URL: [https://www.cdc.gov/rsv/research/rsv-net/dashboard.html](https://www.cdc.gov/rsv/research/rsv-net/dashboard.html) [accessed 2024-01-29]


32. Données sur les variants du SRAS-CoV-2 au Québec. Institut national de santé publique du Québec. URL: [https://www.inspq.qc.ca/covid-19/donnees/variants](https://www.inspq.qc.ca/covid-19/donnees/variants) [accessed 2024-04-05]


36. Données de vaccination contre la COVID-19 au Québec. Institut national de santé publique du Québec. URL: [https://mobile.inspq.qc.ca/covid-19/donnees/vaccination](https://mobile.inspq.qc.ca/covid-19/donnees/vaccination) [accessed 2024-04-05]


39. RESP-NET Interactive Dashboard. CDC. URL: [https://www.cdc.gov/surveillance/resp-net/dashboard.html](https://www.cdc.gov/surveillance/resp-net/dashboard.html) [accessed 2024-02-12]


**Abbreviations**

- **ARI**: acute respiratory infection
- **hCoV**: common human coronavirus
- **hMPV**: human metapneumovirus
- **hPIV**: human parainfluenza virus
- **LSPQ**: Laboratoire de Santé Publique du Québec
ORV: respiratory virus other than SARS-CoV-2
RSV: respiratory syncytial virus
RV: respiratory virus
Updated Surveillance Metrics and History of the COVID-19 Pandemic (2020-2023) in Latin America and the Caribbean: Longitudinal Trend Analysis

Lori Ann Post, PhD; Scott A Wu, BA; Alan G Soetikno, BS; Egon A Ozer, MD, PhD; Yingxuan Liu, MA; Sarah B Welch, MPH; Claudia Hawkins, MD, MPH; Charles B Moss, PhD; Robert L Murphy, MD; Maryann Mason, PhD; Robert J Havey, MD; Alexander L Lundberg, PhD, MSc

Buehler Center for Health Policy and Economics, Robert J Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States

Department of Emergency Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Department of Medicine, Division of Infectious Diseases, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Center for Pathogen Genomics and Microbial Evolution, Robert J Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States

Center for Global Communicable and Emerging Infectious Diseases, Robert J Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States

Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States

Robert J Havey, MD Institute for Global Health, Northwestern University, Chicago, IL, United States

Department of Medicine, General Internal Medicine and Geriatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Lori Ann Post, PhD
Buehler Center for Health Policy and Economics
Robert J Havey, MD Institute for Global Health
Northwestern University
420 E Superior
Chicago, IL, 60611
United States
Phone: 1 312 503 1706
Email: lori.post@northwestern.edu

Abstract

Background: In May 2020, the World Health Organization (WHO) declared Latin America and the Caribbean (LAC) the epicenter of the COVID-19 pandemic, with over 40% of worldwide COVID-19–related deaths at the time. This high disease burden was a result of the unique circumstances in LAC.

Objective: This study aimed to (1) measure whether the pandemic was expanding or contracting in LAC when the WHO declared the end of COVID-19 as a public health emergency of international concern on May 5, 2023; (2) use dynamic and genomic surveillance methods to describe the history of the pandemic in the region and situate the window of the WHO declaration within the broader history; and (3) provide, with a focus on prevention policies, a historical context for the course of the pandemic in the region.

Methods: In addition to updates of traditional surveillance data and dynamic panel estimates from the original study, we used data on sequenced SARS-CoV-2 variants from the Global Initiative on Sharing All Influenza Data (GISAID) to identify the appearance and duration of variants of concern (VOCs). We used Nextclade nomenclature to collect clade designations from sequences and Pangolin nomenclature for lineage designations of SARS-CoV-2. Additionally, we conducted a 1-sided t test for whether the regional weekly speed (rate of novel COVID-19 transmission) was greater than an outbreak threshold of 10. We ran the test iteratively with 6 months of data across the period from August 2020 to May 2023.

Results: The speed of pandemic spread for the region had remained below the outbreak threshold for 6 months by the time of the WHO declaration. Acceleration and jerk were also low and stable. Although the 1- and 7-day persistence coefficients remained statistically significant for the 120-day period ending on the week of May 5, 2023, the coefficients were relatively modest in
magnitude (0.457 and 0.491, respectively). Furthermore, the shift parameters for either of the 2 most recent weeks around May 5, 2023, did not indicate any change in this clustering effect of cases on future cases. From December 2021 onward, Omicron was the predominant VOC in sequenced viral samples. The rolling \( t \) test of speed=10 became entirely insignificant from January 2023 onward.

**Conclusions:** Although COVID-19 continues to circulate in LAC, surveillance data suggest COVID-19 is endemic in the region and no longer reaches the threshold of the pandemic definition. However, the region experienced a high COVID-19 burden in the early stages of the pandemic, and prevention policies should be an immediate focus in future pandemics. Ahead of vaccination development, these policies can include widespread testing of individuals and an epidemiological task force with a contact-tracing system.

(JMIR Public Health Surveill 2024;10:e44398) doi:10.2196/44398

**KEYWORDS**

SARS-CoV-2; COVID-19; Latin America; Caribbean; pandemic; surveillance; COVID-19 transmission; speed; acceleration; deceleration; jerk; dynamic panel; generalized method of moments; GMM; Arellano-Bond; 7-day lag; epidemiological; pandemic; genomic; transmission

**Introduction**

**Background**

COVID-19, the disease caused by the virus SARS-CoV-2, was first detected in Wuhan, China, in the fall of 2019 [1-5]. The first case of COVID-19 in Latin America and the Caribbean (LAC) [6] occurred in São Paulo, Brazil, and involved a Brazilian national traveling home from Northern Italy [7]. Our research team conducted an analysis of the pandemic in LAC 1 year into the pandemic [8]. This study provides 2 additional years of updated surveillance and epidemiological analysis for the region.

We have adopted the World Bank’s definition of LAC, which is based on economic development and geographical proximity, encompassing Argentina, Belize, Bolivia, Brazil, Colombia, Chile, Costa Rica, Ecuador, El Salvador, Guatemala, Guyana, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Suriname, Uruguay, and Venezuela in the Latin America region and Antigua and Barbuda, Aruba, the Bahamas, Barbados, Bermuda, the British Virgin Islands, the Cayman Islands, Cuba, Curacao, Dominica, the Dominican Republic, Grenada, Haiti, Jamaica, Puerto Rico (the United States), St Barthelemy, St Kitts and Nevis, St Lucia, St Vincent and the Grenadines, Sint Maarten (the Netherlands), Trinidad and Tobago, Turks and Caicos Islands, and the US Virgin Islands in the Caribbean region [6].

The World Health Organization (WHO) and Director-General Ghebreyesus declared the end of COVID-19 as a public health emergency of international concern on May 5, 2023 [9-11], based on the recommendation of the COVID-19 Emergency Committee [11]. To that end, we compare how the pandemic was progressing before and after the declaration. The “public health emergency of international concern” status is not a definitive statement of whether a pandemic exists or whether one has ended. The aim of this study was to provide a statistical context to determine whether COVID-19 had transitioned from pandemic to endemic status in LAC at the point of the declaration, a time when many governments worldwide had dropped emergency policy responses around COVID-19.

Epidemiological terms, such as “pandemic,” “epidemic,” “outbreak,” and “endemic,” are used to describe the occurrence and spread of diseases [2,12]. The distinctions between these terms lie in their scope and geographic extent. The terms “pandemic,” “epidemic,” “outbreak,” and “endemic” do not describe the severity of a disease; instead, they describe its prevalence [13]. An epidemic refers to a sudden increase in the number of disease cases in a specific population or region. If the epidemic spreads across several countries or continents, it becomes a pandemic. An outbreak, in contrast, describes a sudden increase in a concentrated setting, usually involving a more limited geographic area than an epidemic. The word “endemic” refers to the constant presence of a disease in a particular geographic region or population, with no sudden increases in case volume [14-16].

Public health surveillance is the “ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning and evaluation of public health practice” [17]. Surveillance not only explains the burden due to a disease but also generates research questions and guides researchers on topics that require further investigation [18-32]. Surveillance allows us to compare the burden of a disease between geographical regions and to understand which regions are most impacted. The impact can be measured through absolute numbers and rates of how many people contract a disease, how many die, and what the affiliated costs are.

However, traditional public health surveillance has inherent limitations, which this study addresses. For example, surveillance predominantly offers a retrospective view, presenting a static snapshot of past events [18-32]. In the middle of a burgeoning pandemic, policy makers and public health practitioners also need to understand what is about to happen. For example, forward-looking metrics are helpful to understand the trajectory of an outbreak’s growth, the shift from linear to exponential expansion, or disparities in the impact across regions. To that end, we have developed enhanced surveillance metrics that capture the dynamic nature of a pandemic and identify imminent growth, notably pinpointing a region’s position along the epidemiological outbreak curve. Although many large cities, states, and countries conducted helpful COVID-19 disease projection models to inform leaders how to
“flatten the curve” so as not to overwhelm health systems, these one-off studies cost millions and only gave projections for a relatively short time frame [33-49]. Moreover, these studies were outdated by the time of their publication and were not relevant to future variants of concern (VOCs) with varied rates of transmission, severe disease, and reproduction. Our enhanced surveillance metrics provide predictive models in an ongoing systematic fashion consistent with surveillance systems. Our enhanced metrics also incorporate dynamic indicators of the pandemic speed at national, regional, and worldwide levels. These metrics gauge the acceleration of the speed from 1 week to the next and predict new cases based on prior dynamic panels across months of daily infections, effectively forecasting the progression of the outbreak. This predictive ability identifies the “echoing forward” of cases. Our metrics have undergone rigorous testing and validation in prior publications that this study updates [8,50-57].

For purposes of this study, standard surveillance metrics explain what has already happened in LAC, while enhanced surveillance metrics show what is about to happen by predicting new cases based on past infections or where along an epidemiological curve a country may sit. We use both types of metrics to inform the possible declaration of the end of the pandemic.

Objectives

This study had 3 objectives. First, we aimed to measure whether there was an expansion or contraction in the pandemic in LAC when the WHO declared the end of the COVID-19 public health emergency of international concern on May 5, 2023. At both the region and country level, we used advanced surveillance and analytical techniques to describe the status of the pandemic in a 2-week window around the WHO declaration. From a public health perspective, we needed to know whether the rate of new COVID-19 cases was increasing, decreasing, or stable from week to week and whether any changes in the transmission rate indicated an acceleration or deceleration of the pandemic. Statistical insignificance represents a significant finding, as this can signal the epidemiological “end” to the pandemic if the rate of new cases is 0 (or extremely low) and stable, meaning that the number of new cases is neither accelerating nor decelerating.

Second, we used dynamic and genomic surveillance methods to describe the history of the pandemic in the region and situate the time window around the WHO declaration within the broader history. We included the ratio of COVID-19 deaths to the number of transmissions as a proxy for the mortality risk from infection at the population level. We also included a historical record of genomic surveillance from sequenced viral specimens to identify the appearance and spread of VOCs in the region.

Third, we aimed to provide a historical context for the course of the pandemic in LAC. We addressed several questions: How did countries respond to the pandemic? How did the region fare in terms of the disease burden? What social, economic, and political factors shaped the course of COVID-19 in the region? This context can provide important lessons for disease prevention and mitigation in future pandemics.

Regional analyses are valuable because multiple factors, not limited to climate change, governance, economic trade routes, public health infrastructure, and human migration, can all shape the risk profile of COVID-19 transmissions [58-62]. In this study, we discussed the role of these factors in the course of the pandemic in LAC. In May 2020, the WHO declared LAC the epicenter of the pandemic, accounting for over 40% of all worldwide COVID-19 deaths at the time [63]. This high disease burden was a result of the unique circumstances in LAC, and these circumstances can inform preparedness for future pandemics and waves of transmissions.

Methods

Study Design and Data Collection

We conducted trend analyses with longitudinal COVID-19 data from Our World in Data (OWID) [64] to provide updates of traditional surveillance data and dynamic panel estimates from the original studies by Post and colleagues [8,55,56,65-67]. For the LAC region, the data comprised an unbalanced panel of 43 countries and territories, from August 21, 2020, to May 12, 2023. Please see the Introduction section for the complete list of countries. Since several countries worldwide switched from daily to weekly reports at various points in 2023, so we used a cubic spline to interpolate daily new cases and deaths if any country had 4 consecutive periods of nonzero new cases interspersed by 6 days of 0 new cases.

Additionally, we used data on sequenced SARS-CoV-2 variants from the Global Initiative on Sharing All Influenza Data (GISAID) to identify the appearance and duration of VOCs [68-72]. We used Nextclade nomenclature [73] to collect clade designations from sequences and Pangolin nomenclature for lineage designations of SARS-CoV-2 [74,75]. Metadata for the study period were collected on June 22, 2023. To avoid a low frequency or potentially erroneous samples, the data set was further filtered to exclude months with fewer than 100 available samples, variant groups with fewer than 5 samples in a month, and variant groups representing less than 0.5% of the total samples in a month. The final data set consisted of 184,386 total samples available on GISAID [69-72].

Traditional surveillance metrics include the speed of spread of the pandemic, defined as the rate of new COVID-19 cases per 100,000 people. Enhanced metrics add acceleration, jerk, and 1- and 7-day persistence measures. Acceleration is the change in speed from one unit of time to the next. This measure can identify whether the rate of transmission is increasing (positive acceleration), decreasing (negative), or stable (0). Jerk is the change in acceleration from one time unit to the next. Its name is lifted from physics nomenclature. A large jerk can signal explosive growth in transmission rates. The 1- and 7-day persistence measures provide the impact of the 1- and 7-day lag speed on the current speed. Thus, these measures capture how COVID-19 cases echo-forward to cases either 1 or 7 days later. They are derived from an Arellano-Bond dynamic panel data model [76], which follows the form

\[ y_{it} = \beta y_{it-1} + \beta X_{it} + \alpha_i + u_{it}, \]

where the dependent variable is speed, the independent variables include weekend and recent week indicators, \( \alpha_i \) is a country
fixed effect, and $u_i$ is the idiosyncratic error term. See the initial study for details [8].

**Statistical Analysis**

Lastly, we analyzed the potential “statistical end” to the pandemic with (1) dynamic panel estimates for shifts in the pandemic [40,42] and (2) a 1-sided $t$ test for whether the mean speed was equal to or greater than the outbreak threshold of 10 [77]. We ran the latter test on a rolling 6-month window of the weekly regional speed, and we plotted the $P$ values from the $t$ test over time. All statistical analyses were conducted in R version 4.2.1 (R Foundation for Statistical Computing) with the plm package (version 2.6-2) [65,66].

**Ethical Considerations**

All data used in this study are publicly available and contain no identifiable, private information. Therefore, the study does not constitute research with human subjects, as defined by 45CFR46:102, and Institutional Review Board review was unsolicited. However, the authors note that anonymized COVID-19 data surveillance systems can generate local and global ethical questions beyond the scope of this study [78].

**Results**

**Dynamic Panel Estimates**

Table 1 presents dynamic panel estimates for the most recent time window. The Wald test for regression was significant ($P<.001$), and the Sargan test failed to reject the validity of the overidentification restrictions ($P=.99$). Although the 1- and 7-day lag coefficients were statistically significant as a function of the large sample size, suggesting a cluster effect in which cases on a given day impacted cases 1 and 7 days later, the coefficients were modest in magnitude (0.457 and 0.491, respectively). Furthermore, the shift parameters for the most recent week were significant and negative, meaning the clustering effect had become smaller in the week after the WHO declaration (but the shift parameter was positive and similar in magnitude for the prior week).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-day lag coefficient</td>
<td>0.457</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>7-day lag coefficient</td>
<td>0.491</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Shift parameter week of April 28</td>
<td>0.212</td>
<td>.03</td>
</tr>
<tr>
<td>Shift parameter week of May 5</td>
<td>−0.274</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Weekend</td>
<td>0.981</td>
<td>.08</td>
</tr>
</tbody>
</table>

aLAC: Latin America and the Caribbean.
bWald test: $\chi^2=951.65$ ($P\leq2.22e-16$); Sargan test: $\chi^2=37$ ($P=.99$).

**Static Surveillance Metrics**

Static surveillance metrics for the week of April 28, 2023, are provided in Table 2. The same metrics for the week of May 5, 2023, are provided in Table S1 in Multimedia Appendix 1. Most countries had a small number of new COVID-19 cases. The exceptions were Brazil, Costa Rica, Panama, Puerto Rico, and Uruguay. By far, the highest transmission rate was observed in Puerto Rico, where the speed was 144 in the week of April 28 and 182 in the following week. The next highest speed over the 2 weeks was just under 19 in Brazil, and the speed for the country dipped to 12 in the following week. For all other countries and territories, the speed was below the threshold considered a low transmission rate by the Centers for Disease Control and Prevention (CDC) [79]. Most rates fell well below the informal threshold of 10 cases per week per 100,000 population [8,50-57]. Specifically, low transmission is considered no more than 10 cases per 100,000 people per week, moderate transmission is 10-50 cases per 100,000 people per week, and substantial transmission is 50-100 cases per 100,000 people per week [79,80].

### Table 1. Arellano-Bond dynamic panel estimates of the number of daily COVID-19 infections reported by LAC\(^a\) countries from April 28 to May 12, 2023.\(^b\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-day lag coefficient</td>
<td>0.457</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>7-day lag coefficient</td>
<td>0.491</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Shift parameter week of April 28</td>
<td>0.212</td>
<td>.03</td>
</tr>
<tr>
<td>Shift parameter week of May 5</td>
<td>−0.274</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Weekend</td>
<td>0.981</td>
<td>.08</td>
</tr>
</tbody>
</table>

\(^a\)LAC: Latin America and the Caribbean.
\(^b\)Wald test: $\chi^2=951.65$ ($P\leq2.22e-16$); Sargan test: $\chi^2=37$ ($P=.99$).
<table>
<thead>
<tr>
<th>Country</th>
<th>New COVID-19 cases, n</th>
<th>Cumulative COVID-19 cases, n</th>
<th>7-day Moving average of new cases</th>
<th>Weekly transmission rate/100,000 individuals</th>
<th>New deaths, n</th>
<th>Cumulative deaths, n</th>
<th>7-day Moving average of deaths</th>
<th>Death rate/100,000 individuals</th>
<th>Conditional death rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>0</td>
<td>10,044,957</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Aruba</td>
<td>10</td>
<td>44,114</td>
<td>10</td>
<td>0</td>
<td>287</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Barbados</td>
<td>0</td>
<td>107,466</td>
<td>0</td>
<td>0</td>
<td>588</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Belize</td>
<td>0</td>
<td>70,782</td>
<td>0</td>
<td>0</td>
<td>668</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Bermuda</td>
<td>0</td>
<td>18,860</td>
<td>0</td>
<td>0</td>
<td>165</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Bolivia</td>
<td>61</td>
<td>1,197,239</td>
<td>60.57</td>
<td>3.48</td>
<td>22,377</td>
<td>0.14</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Brazil</td>
<td>5827</td>
<td>37,449,418</td>
<td>5953.43</td>
<td>18.95</td>
<td>701,494</td>
<td>39.86</td>
<td>0.15</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Chile</td>
<td>208</td>
<td>5,283,908</td>
<td>273</td>
<td>7.44</td>
<td>61,384</td>
<td>8.71</td>
<td>0.28</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Colombia</td>
<td>89</td>
<td>6,364,636</td>
<td>89.29</td>
<td>1.20</td>
<td>142,713</td>
<td>1</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>139</td>
<td>1,228,659</td>
<td>151.29</td>
<td>18.81</td>
<td>9351</td>
<td>1.71</td>
<td>0.23</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Cuba</td>
<td>44</td>
<td>1,113,088</td>
<td>37.29</td>
<td>0</td>
<td>8530</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Curacao</td>
<td>2</td>
<td>45,798</td>
<td>1.71</td>
<td>6.69</td>
<td>301</td>
<td>0</td>
<td>0.56</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>7</td>
<td>661,045</td>
<td>5.71</td>
<td>0</td>
<td>4384</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Ecuador</td>
<td>126</td>
<td>1,061,766</td>
<td>108</td>
<td>4.89</td>
<td>36,019</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>Grenada</td>
<td>0</td>
<td>19,693</td>
<td>0</td>
<td>0</td>
<td>238</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Guatemala</td>
<td>85</td>
<td>1,248,171</td>
<td>97.14</td>
<td>0</td>
<td>20,189</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Guyana</td>
<td>1</td>
<td>73,162</td>
<td>3</td>
<td>0</td>
<td>1298</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Haiti</td>
<td>0</td>
<td>34,228</td>
<td>0.29</td>
<td>0</td>
<td>860</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>Honduras</td>
<td>0</td>
<td>472,533</td>
<td>0</td>
<td>0</td>
<td>11,112</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Jamaica</td>
<td>9</td>
<td>154,786</td>
<td>7.43</td>
<td>2.24</td>
<td>3536</td>
<td>0</td>
<td>0.28</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Mexico</td>
<td>1182</td>
<td>7,590,885</td>
<td>1264.14</td>
<td>6.49</td>
<td>333,990</td>
<td>9.43</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>1</td>
<td>15,697</td>
<td>1</td>
<td>0.10</td>
<td>245</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Panama</td>
<td>86</td>
<td>1,036,733</td>
<td>89</td>
<td>13.67</td>
<td>8621</td>
<td>0.43</td>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Paraguay</td>
<td>0</td>
<td>735,759</td>
<td>0</td>
<td>0</td>
<td>19,880</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.03</td>
</tr>
<tr>
<td>Peru</td>
<td>280</td>
<td>4,501,130</td>
<td>262.29</td>
<td>5.77</td>
<td>220,122</td>
<td>7</td>
<td>0.10</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>670</td>
<td>1,224,366</td>
<td>606.71</td>
<td>144.29</td>
<td>5901</td>
<td>1.57</td>
<td>0.33</td>
<td>0</td>
<td>0.33</td>
</tr>
<tr>
<td>St Barthelemy</td>
<td>1</td>
<td>5486</td>
<td>0.14</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sint Maarten</td>
<td>0</td>
<td>11,030</td>
<td>0</td>
<td>0</td>
<td>92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>St Kitts and Nevis</td>
<td>0</td>
<td>6598</td>
<td>0</td>
<td>0</td>
<td>46</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>St Lucia</td>
<td>0</td>
<td>30,052</td>
<td>0</td>
<td>0</td>
<td>409</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>St Vincent and Grenadines</td>
<td>1</td>
<td>9611</td>
<td>1</td>
<td>0</td>
<td>124</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Suriname</td>
<td>0</td>
<td>82,495</td>
<td>0</td>
<td>0</td>
<td>1404</td>
<td>0</td>
<td>1.22</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>7</td>
<td>191,350</td>
<td>-1.57</td>
<td>3.01</td>
<td>4387</td>
<td>0</td>
<td>-0.21</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Turks and Caicos Islands</td>
<td>4</td>
<td>6565</td>
<td>4.57</td>
<td>0</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>US Virgin Islands</td>
<td>2</td>
<td>24,919</td>
<td>2</td>
<td>0</td>
<td>130</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
</tbody>
</table>
We noted, however, some sparsity in data reports. To temper the episodic reporting that became more frequent in 2023, we shifted from daily transmission rates and daily speed to weekly transmission rates and weekly speed in contrast to our initial study. A second benefit of shifting from daily to weekly reporting was that the CDC established weekly thresholds of what constitutes a low, moderate, and substantial outbreak, as noted earlier. For example, having 0 new cases in Argentina over the 2 weeks is clearly a product of a failure to report data. Although Puerto Rico appears to have been in a large outbreak around the WHO declaration, island territories often vacillate between high and low transmission rates [81]. The outbreaks in the other 4 countries were relatively small. Based on the definition of a pandemic or an outbreak in several countries, the data indicated a shift from pandemic to endemic COVID-19 in most of LAC, while it was an epidemic in Puerto Rico.

The surveillance metrics demonstrated little to no change before and after the WHO declaration. Without question, Brazil had the most cases of COVID-19 transmissions and deaths, but this rank is partly a function of population size. Thus, a better measure is the number of COVID-19 cases and deaths per 100,000 people. Moreover, death is often a better proxy for the state of an outbreak than transmissions because deaths are less likely to be undercounted [82]. Undercounting may be due to poor public health infrastructure, a switch to home antigen testing, or a dearth of polymerase chain reaction (PCR) testing or other resources. Although Brazil reported 0.02 deaths per confirmed infection, several countries had higher rates. Peru had the highest rate at 0.05, followed by Mexico at 0.04 and Ecuador, Haiti, and Paraguay at 0.03.

Table 3 and Table S2 in Multimedia Appendix 1 contain enhanced dynamic surveillance metrics for the 2 weeks before and after May 5, 2023. Again, the speed was low for every country except Brazil, Costa Rica, Panama, Puerto Rico, and Uruguay. Acceleration and jerk were both either small or negative for every country and territory. The 7-day persistence effect on the speed was also small except for the countries in outbreak status. These metrics suggest the pandemic may have indeed ended for the region. Only a single territory was in a substantial outbreak and 4 others in relatively modest outbreaks; thus, epidemiologically, COVID-19 would be considered an epidemic in Puerto Rico and an end-stage epidemic in Brazil, Costa Rica, Panama, and Uruguay, with mild outbreaks in these countries. The average speed for the LAC region was below the outbreak threshold. Since COVID-19 cases in LAC have leveled off, with the exception of Puerto Rico, indicating that the disease is stable or constant in the region, it is more accurate to say that COVID-19 in LAC has shifted to being endemic [13,83,84]. Note that the numbers in Table 3 and Table S2 in Multimedia Appendix 1 are not calculated as day-over-day averages across the week, as they are in Table 2 and Table S1 in Multimedia Appendix 1. Thus, the magnitudes of the speed are consistent but not identical to those matched across tables.

Table 4 compares the 7-day persistence effect on the speed for the top 5 countries around the 2 weeks of the WHO declaration. These ranks largely reflect the speed in the countries with outbreaks in the prior tables. However, Chile appeared at rank 5 in the week of April 28, 2023. This level of persistence would be cause for alarm if Chile did not see a substantial drop, as it did, in the subsequent week.

<table>
<thead>
<tr>
<th>Country</th>
<th>Weekly speed</th>
<th>Weekly acceleration</th>
<th>Weekly jerk</th>
<th>7-day Persistence effect on speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Aruba</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Barbados</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.23</td>
</tr>
<tr>
<td>Belize</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bermuda</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bolivia</td>
<td>3.47</td>
<td>0.02</td>
<td>-0.01</td>
<td>2.55</td>
</tr>
<tr>
<td>Brazil</td>
<td>19.35</td>
<td>-0.24</td>
<td>0.07</td>
<td>17</td>
</tr>
<tr>
<td>Chile</td>
<td>9.74</td>
<td>-0.80</td>
<td>0.02</td>
<td>12.09</td>
</tr>
<tr>
<td>Colombia</td>
<td>1.20</td>
<td>0.01</td>
<td>0</td>
<td>0.77</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>20.44</td>
<td>-0.43</td>
<td>-0.07</td>
<td>17.59</td>
</tr>
<tr>
<td>Cuba</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.10</td>
</tr>
<tr>
<td>Curacao</td>
<td>5.90</td>
<td>0.24</td>
<td>0.02</td>
<td>3.47</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>Ecuador</td>
<td>4.20</td>
<td>0.23</td>
<td>0</td>
<td>2.42</td>
</tr>
<tr>
<td>Grenada</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.88</td>
</tr>
<tr>
<td>Guatemala</td>
<td>0.60</td>
<td>0</td>
<td>0</td>
<td>0.43</td>
</tr>
<tr>
<td>Guyana</td>
<td>0.51</td>
<td>0</td>
<td>0</td>
<td>0.16</td>
</tr>
<tr>
<td>Haiti</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Honduras</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>Jamaica</td>
<td>1.86</td>
<td>0.10</td>
<td>0.02</td>
<td>1.32</td>
</tr>
<tr>
<td>Mexico</td>
<td>6.94</td>
<td>-0.12</td>
<td>-0.03</td>
<td>5.60</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>0.09</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>Panama</td>
<td>14.12</td>
<td>-0.05</td>
<td>-0.04</td>
<td>11.86</td>
</tr>
<tr>
<td>Paraguay</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Peru</td>
<td>5.40</td>
<td>0.05</td>
<td>0.06</td>
<td>4.94</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>130.59</td>
<td>4.03</td>
<td>0.40</td>
<td>85.52</td>
</tr>
<tr>
<td>St Barthelemy</td>
<td>1.30</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sint Maarten</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>St Kitts and Nevis</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>St Lucia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>St Vincent and Grenadines</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.95</td>
</tr>
<tr>
<td>Suriname</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>-0.76</td>
<td>0.60</td>
<td>0.47</td>
<td>2.29</td>
</tr>
<tr>
<td>Turks and Caicos Islands</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>US Virgin Islands</td>
<td>1.72</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Uruguay</td>
<td>18.94</td>
<td>-0.40</td>
<td>0.03</td>
<td>17.12</td>
</tr>
<tr>
<td>Venezuela</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>Argentina</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*LAC: Latin America and the Caribbean.
Table 4. LAC<sup>a</sup> countries with the highest 7-day persistence estimate in the weeks of April 28 and May 5, 2023.

<table>
<thead>
<tr>
<th>Country</th>
<th>7-day Persistence</th>
<th>Country</th>
<th>7-day Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puerto Rico</td>
<td>91.89</td>
<td>Puerto Rico</td>
<td>91.89</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>17.59</td>
<td>Costa Rica</td>
<td>14.38</td>
</tr>
<tr>
<td>Uruguay</td>
<td>17.12</td>
<td>Brazil</td>
<td>13.62</td>
</tr>
<tr>
<td>Brazil</td>
<td>17.00</td>
<td>Uruguay</td>
<td>13.33</td>
</tr>
<tr>
<td>Chile</td>
<td>12.09</td>
<td>Panama</td>
<td>9.94</td>
</tr>
</tbody>
</table>

<sup>a</sup>LAC: Latin America and the Caribbean.

Figure 1 plots the regional speed, acceleration, jerk, and 7-day persistence metrics from August 21, 2020, to May 12, 2023. The dashed gray line denotes the informal CDC outbreak threshold speed of 10. The region was in a nearly continuous state of outbreak until September 2021. Several months later, the Omicron variant drove a substantial outbreak, culminating in a speed of 66 novel cases per week per 100,000 population. A second, smaller wave followed, which brought a peak speed of 20 in July 2022. A final outbreak barely eclipsed the outbreak threshold speed of 10 in December 2023. The region has seen the speed decrease, stabilize, and remain well below the outbreak threshold ever since.

Figure 1. Timeline of speed, acceleration, jerk, and 7-day persistence in LAC from 2020-2023. LAC: Latin America and the Caribbean.

Figure 2 plots variant groups as a proportion of all viral specimens collected and sequenced in the region (and made available through GISAID) each month. The early, nearly continuous state of outbreak in the region spanned the ancestral variant through the temporary dominance of the Delta variant. Interestingly, the region also had a period in which the majority of viral specimens returned as the Gamma variant. Compared to other worldwide regions, LAC had a greater diversity of variants. Later outbreaks were driven by the Omicron variant. LAC, like much of the rest of the world, saw a surge in cases amid the heightened transmissibility of Omicron [85]. Another potential indication of the end to the pandemic is the continued dominance of the Omicron variant. Although the region saw a mixture of variants prior to the arrival of Omicron in November 2021, viral sequences have almost exclusively returned as Omicron and its subvariants ever since.

Figure 2 plots variant groups as a proportion of all viral specimens collected and sequenced in the region (and made available through GISAID) each month. The early, nearly continuous state of outbreak in the region spanned the ancestral variant through the temporary dominance of the Delta variant. Interestingly, the region also had a period in which the majority of viral specimens returned as the Gamma variant. Compared to other worldwide regions, LAC had a greater diversity of variants. Later outbreaks were driven by the Omicron variant. LAC, like much of the rest of the world, saw a surge in cases amid the heightened transmissibility of Omicron [85]. Another potential indication of the end to the pandemic is the continued dominance of the Omicron variant. Although the region saw a mixture of variants prior to the arrival of Omicron in November 2021, viral sequences have almost exclusively returned as Omicron and its subvariants ever since.
Figure 2. VOC groups as a proportion of all sequenced SARS-CoV-2 specimens over time in LAC from March 2020 to May 2023. LAC: Latin America and the Caribbean; VOC: variant of concern.

Figure 3 plots $P$ values from a series of 1-sided $t$ tests of whether the speed for the region was equal to or greater than the outbreak threshold of 10. These tests were conducted on a rolling 6-month window of the weekly regional speed. The dashed gray line denotes the least restrictive conventional significance level threshold of $\alpha=0.10$. The test immediately rejected the null hypothesis in favor of the alternative and continued to so until the end of 2021. The test statistic became totally insignificant for a brief period in December 2021, before regaining significance in February 2022. This second period of significance was driven by the large Omicron outbreak. The test lost significance again in August 2022 and has been totally insignificant from December 2022 onward. This more recent lack of statistical significance is consistent with the end to the pandemic in the region, as the test clearly failed to reject the null hypothesis of the outbreak threshold speed.

Figure 4 provides a timeline of the onset of COVID-19 in LAC, as well as vaccination programs and major events that shaped the course of the pandemic in the region, such as economic, educational, and political impacts.
Discussion

Principal Findings

The results indicate that COVID-19 was well contained in LAC at the time of the WHO declaration. Both traditional and enhanced surveillance metrics indicated low and stable transmission rates around the time of the declaration. Nearly every SARS-CoV-2 test specimen returned as the Omicron variant from November 2021 onward. Finally, the rolling $t$ test for regional outbreak status was entirely insignificant from December 2022 onward. Collectively, these results suggest a transition from pandemic to endemic in the region. However, these more recent trends do not reflect the full history of COVID-19 in LAC. The region had a unique set of risk factors and responses to COVID-19, which shaped the trajectory of transmission rates, especially in the early stages of the pandemic.
Although many countries in LAC reacted swiftly and preemptively to protect citizens and contain the spread of COVID-19, the region struggled to contain the levels of infection, mainly due to the prevalence of informal economies and the limitations of health infrastructure and social protection systems [86-90]. The informal sector of the economy refers to unrecorded economic activity, which would contribute to the gross domestic product (GDP) and tax revenue, if recorded. The sector contributes around a third of all economic activity in low-to-middle-income countries, but the sector poses a challenge for public health prevention efforts because its workers are often unknown and unregulated by governments [91]. LAC also faces an elevated risk profile for pandemics because of several factors, including climate change and human migration [92,93]. (We note, however, that the relationship between weather and COVID-19 transmissions can be complex, and not every product of climate change necessarily raises the risk of transmission [58]. Recent work indicates prevention strategies should be instituted in the summer and fully implemented in the winter [59].)

In May 2020, LAC was declared the epicenter of the pandemic by the WHO, with over 40% of the world’s COVID-19 deaths and more than 6.5 million confirmed cases at the time [63]. Just 1 year later, in May 2021, the region recorded its one-millionth COVID-19 death, with almost 90% of the deaths occurring in 5 countries: Brazil (44.3%), Mexico (22.1%), Colombia (8.3%), Argentina (7.3%), and Peru (6.7%) [94]. At that time, of the 153.5 million COVID-19 vaccines administered in the Americas, only 21.6% came through LAC [94,95]. Through 2022 and 2023, difficulties in production, access, and equitable distribution among different countries within LAC limited COVID-19 vaccination success and exacerbated hesitancy and rejection [96]. Despite the relatively late arrival of COVID-19 to LAC, the region suffered disproportionately. Several factors emerged as key determinants of the region’s hardship, including pandemic unpreparedness, fragile health care systems, socioeconomic inequalities, and poor governmental support [97,98]. Ultimately, LAC suffered the largest death toll from COVID-19 across developing regions and 1 of the sharpest declines in economic activity, with expected long-lasting socioeconomic consequences for the region [99].

LAC was economically vulnerable when the pandemic hit, due to limited fiscal resilience after the 2008 worldwide recession [86]. In 2020, the region experienced a historic 7% contraction in the GDP, with many countries seeing over a 10% economic decline [86,100]. This contraction was the largest in the region’s history [101]. Although modest growth was observed in 2021 and 2022 (6.9% and 3.5%, respectively), the International Monetary Fund projected a slowdown to 1.7% in 2023, largely due to inflation caused by interest rate hikes and fiscal policies [100]. Challenges included worsening inequality, higher poverty rates, and impacts on labor income, particularly among the extremely poor [102-106]. Major economic challenges anticipated for LAC in the years ahead include lingering effects of the pandemic in the setting of increased energy and food prices, which are mostly due to Russia’s invasion of Ukraine and China’s economic slowdown [100]. Even as the region rebounded, poverty rates remained high and continued to rise in 2022 [107,108]. Labor income, accounting for 68% of income among impoverished households, saw significant drops, particularly among the extremely poor [109].

LAC was especially burdened by sociopolitical unrest during the pandemic, in large part due to preexisting increases in authoritarian practices, weak democratic institutions, politicized judicial systems, and high levels of organized and unorganized crime and violence [100]. The region experienced a disproportionate number of political and public health threats, most notably the assassination of Haiti’s president in a coup d’état, followed by a devastating earthquake that effectively eliminated any remnants of the public health infrastructure [110,111]. COVID-19 was eclipsed by the outbreak of violence, cholera, and extreme poverty [112]. Other political turmoil included significant demonstrations in Brazil repeatedly protesting former President Jair Bolsonaro and his far-right administration’s handling of the pandemic, which included promotion of unscientific theories and pandemic denialism [113-116]. Intense protests also occurred in Cuba, Chile, Bolivia, and Colombia, with the average growth rate of the number of protests in LAC marked at 5.84% compared to a worldwide rate of 1.15%, according to the United Nations Development Program [97,117]. Given previously described socioeconomic issues, protests in LAC were also markedly intersectional, with examples including regular violent demonstrations in Colombia over both pandemic-related restrictions and police brutality and unpopular tax reform at the time [118]. Overall, in the region, as a result of unpopular policies and social unrest, a widespread anti-incumbent electoral trend was observed, starting in 2020 [100].

**Policies Implemented to Control and Mitigate the Transmission of COVID-19**

Much of the LAC COVID-19 response was spearheaded by the WHO’s Pan American Health Organization (PAHO). PAHO efforts included building the COVID-19 Genomic Surveillance Regional Network, which provided the region a way to track the evolution of COVID-19 and monitor new variants and pathogens with pandemic potential [119]. PAHO also deployed extra COVID-19 support personnel and equipment to weaker regional health systems, notably in Venezuela, Bolivia, Paraguay, Guyana, Nicaragua, Honduras, Guatemala, Haiti, and some of the smaller Caribbean island states [120]. Throughout the pandemic, PAHO emphasized a multisectoral approach that included school closures, teleworking, hospital readiness assessments, and increased testing capabilities [120-124].

Specific countries’ public health responses to the early pandemic included Chile implementing a nightly curfew, with many large cities soon establishing mandatory quarantine measures [125]. After the first case was reported in Costa Rica on March 6, 2020, the country’s National Emergency Commission enabled systematic and interinstitutional mobilization of resources, as well as the activation of emergency operation centers [125]. After community transmission was identified on March 26, 2020, in Mexico, the country introduced mandatory restrictions that included suspending classes in all public and private schools, enforcing stay-at-home orders, cancelling public and private events, banning meetings of more than 100 people, and...
self-quarantining among disease contacts for at least 15 days [125]. Upon the declaration of a state of national emergency in Peru on March 15, 2020, strict controls were placed on citizens’ movements, with only trips related to the purchase of food and medicines allowed and only workers from critical sectors permitted to commute [125].

The WHO encouraged countries in LAC to deploy a phased vaccine introduction that prioritized frontline and health care workers, as well as older adults and people with underlying health conditions [126]. Vaccine introduction and administration, which began in the region at the end of December 2020, varied widely by country both initially and throughout the pandemic despite various organizational efforts to ensure equitable access [126,127]. Internal vaccine research and development in LAC was minimal compared to the rest of the world, with no development spearheaded by any LAC nations [128]. The largest scientific contribution, as measured by published COVID-19 vaccine research manuscripts, was made by Brazil, which contributed one-third of the region’s papers regarding vaccine efficacy and safety [128]. Vaccination campaigns were a critical tool in the pandemic because vaccines were highly effective in reducing the probability of death, and population level fatality rates are partly a function of the percentage of the population vaccinated [62,129,130].

In October 2022, PAHO’s director reassured the region that disease trends insinuated a potential transition of the pandemic from an acute phase to a phase of sustained control [100]. Through 2022, the United States donated over 70 million COVID-19 vaccines to LAC directly and via the COVID-19 Vaccines Global Access (COVAX) partnership, with top recipients including Mexico (16.9 million doses), Guatemala (8.5 million doses), and Colombia (7 million doses) [100,131,132]. Larger agreements and agencies such as COVAX relied on a variety of vaccine types, mainly including messenger RNA (mRNA)–based vaccines (Pfizer-BioNTech, Moderna) and viral vector vaccines (Oxford University-AstraZeneca, Johnson & Johnson, Sputnik V) [133-135]. The lowest vaccine rates were observed in Haiti, with just over 2% of its population fully vaccinated by December 2022 [100]. Additional countries with vaccination rates below 40% included Grenada, Jamaica, St Lucia, and St Vincent and the Grenadines [100].

As the pandemic continued, new light was shed on the effect of LAC migrant crises on public health and vaccination policies [136]. Venezuela was deeply impacted [137,138]. By August 2022, more than 6.8 million Venezuelans fled the country due to political, social, and economic turmoil in what is now considered the largest international displacement in LAC in contemporary history [136]. Most Venezuelan refugees remained in LAC, with Argentina, Brazil, Chile, Colombia, Ecuador, and Peru having absorbed 75% of the migrants [139]. In addition to placing severe socioeconomic strain on recipient countries, the massive refugee influx also increased communicable disease spread and created new public health obstacles. For COVID-19, these challenges included ensuring access to vaccinations [136]. For example, most countries required government-issued identification to access COVID-19 vaccinations, which almost no refugees possessed upon arrival [136,140]. Colombia was 1 of few countries to eventually extend COVID-19 vaccine access to migrants with irregular immigration status, including those without documentation, and did so in October 2021 [136]. Meanwhile, many LAC leaders, including Colombian President Iván Duque Márquez and Venezuelan President Nicolás Maduro, blamed the constant flux of migrants throughout LAC for upsurges in COVID-19 cases [137]. In addition to questionable vaccine access, 75% of displaced Venezuelans in Peru and 39% in Colombia believed they would not be able to access health care if they became ill with COVID-19 [136].

At the end of the pandemic, to the best of our knowledge, barring Haiti due to the paucity of testing and reporting [141,142], the highest COVID-19 mortality rate in the region was in Peru, followed by Chile, Brazil, Trinidad and Tobago, and Argentina [100]. Despite accounting for just 25% of the world’s total COVID-19 case volume, LAC accounted for 43% of total deaths [119]. Major gaps in technical capacity, medical care coordination, and information technology infrastructure (many areas do not have infrastructure to support digital health systems and still rely on a pen-and-paper medical record system) were retrospectively highlighted as key reasons why LAC experienced such catastrophic spread and burden of COVID-19 [143,144]. In addition to these improvements, the region will focus on comprehensive disease surveillance strategies and methods to mitigate collateral damage on public health and social services during similar future crises [98].

**Limitations**

COVID-19 data had become less frequently reported around the world by the time the WHO declared an end to the pandemic [145]. Additionally, more people began to use at-home tests as the pandemic evolved [146]. Some countries, such as Argentina, appear to have stopped reporting novel COVID-19 cases altogether. Viral specimen tests for VOCs in GISAID are also dependent on testing and sequencing capacity, which varied by country across the region. Because the enhanced surveillance metrics of the speed, acceleration, jerk, and 7-day persistence are based on rates, not total counts, statistical bias caused by countries dropping in or out of the sample is mitigated, but to the extent that a nonincluded country is unrepresentative of the region in the disease burden, the omission of a country or territory can still influence historical data comparisons. The dynamic panel estimates are based on a 120-day window, which further limits intertemporal biases caused by changes in reporting. Still, data availability is always a limitation. Continued reports on COVID-19 transmissions after mid-May 2023 would provide additional insight to whether COVID-19 had by then truly shifted from pandemic to endemic status. This team did not believe transmissions data were sufficiently available to draw firm conclusions beyond that time point.

**Conclusion**

The concern about potential resurgence of the virus is valid. As long as COVID-19 continues to spread and mutate, the possibility of new variants emerging remains. Variants could potentially be more transmissible, resistant to vaccines, or cause more severe illness. This underlines the importance of continued vigilance, vaccination efforts, and worldwide cooperation to control the spread of the virus [54].
LAC experienced a high COVID-19 disease burden in the early stages of the pandemic. To prepare for future pandemics, one of the most important lessons from across the world may lie in the ways to reduce the disease burden ahead of the dissemination of vaccines and treatment modalities [147]. On that front, rapid and widespread testing of individuals, coupled with an epidemiological task force and contact-tracing system, is an effective first line of defense [148]. Lockdown policies have also proven effective [149], perhaps in part because international trade is positively associated with COVID-19 transmissions [61]. (As with climate change and transmissions, the relationship between containment policies and transmissions can be complex, and stricter containment may not always imply fewer transmissions [150,151].) Finally, new technologies will always renew the potential to improve early surveillance methods [152-154].

One bright point for LAC and other regions has been an advance in regional cooperation to tackle the public health threat posed by novel pathogens [155,156]. Novel indicators of pandemic preparedness at the regional level might further identify countries in relative need of support [157]. Measures of general governance, for example, are positively correlated with vaccination rates [60]. Recent work also suggests ways to reduce not only the death caused by novel pathogens but also the likelihood of their appearance, such as better surveillance of human and wildlife interaction and reductions in air pollution [158,159]. Continued cooperation will be necessary to reduce both the likelihood and impact of future pandemics [160,161].

Acknowledgments
Research reported in this publication was generously supported by the Robert J Havey, MD Institute for Global Health’s “Global Innovation Challenge Award” at Northwestern University, Feinberg School of Medicine. This study was also supported by Feed the Future through the US Agency for International Development, under the terms of contract 7200L1800003, and the Feed the Future Innovation Lab for Collaborative Research on Sustainable Intensification (grant AID-OAA-L-14-00006). The opinions expressed herein are those of the authors and do not necessarily reflect the views of the US Agency for International Development or the authors’ organizations.

We gratefully acknowledge all data contributors (ie, the authors and their originating laboratories responsible for obtaining the specimens and their submitting laboratories for generating the genetic sequence and metadata and sharing via the Global Initiative on Sharing All Influenza Data [GISAID] Initiative) on which this research is based.

Data Availability
The data in this study were obtained from 2 publicly available sources. Our World in Data provided data on COVID-19 transmissions and death rates [162]. The Global Initiative on Sharing All Influenza Data (GISAID) provided data on sequenced SARS-CoV-2 specimens.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary tables.

[DOCX File, 24 KB - publichealth_v10i1e44398_app1.docx ]

References


66. Global Initiative on Sharing All Influenza Data (GISAID). URL: https://www.gisaid.org/ [accessed 2023-09-01]


146. Stein R. As the pandemic ebbs, an influential COVID tracker shuts down. npr. 2023 Feb 10. URL: https://tinyurl.com/yurl.com/bdeafukz


Abbreviations

CDC: Centers for Disease Control and Prevention
COVAX: COVID-19 Vaccines Global Access
GDP: gross domestic product
GISAID: Global Initiative on Sharing All Influenza Data
LAC: Latin America and the Caribbean
mRNA: messenger RNA
PAHO: Pan American Health Organization
VOC: variant of concern
WHO: World Health Organization

©Lori Ann Post, Scott A Wu, Alan G Soetikno, Egon A Ozer, Yingxuan Liu, Sarah B Welch, Claudia Hawkins, Charles B Moss, Robert L Murphy, Maryann Mason, Robert J Havey, Alexander L Lundberg. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 17.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
The Impact of Wireless Emergency Alerts on a Floating Population in Seoul, South Korea: Panel Data Analysis

Sungwook Yoon, PhD; Hyungsu Lim, PhD; Sungho Park, PhD

1Korea Information Society Development Institute, Jincheon, Republic of Korea
2Department of Information Systems, Business Statistics, and Operations Management, School of Business and Management, Hong Kong University of Science and Technology, Clear Water Bay, China (Hong Kong)
3Business School, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Sungho Park, PhD
Business School
Seoul National University
1 Gwanak-ro, Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 2 880 6949
Email: spark104@snu.ac.kr

Abstract

Background: Wireless emergency alerts (WEAs), which deliver disaster information directly to individuals’ mobile phones, have been widely used to provide information related to COVID-19 and to encourage compliance with social distancing guidelines during the COVID-19 pandemic. The floating population refers to the number of people temporarily staying in a specific area, and this demographic data can be a useful indicator to understand the level of social distancing people are complying with during the COVID-19 pandemic.

Objective: This study aimed to empirically analyze the impact of WEAs on the floating population where WEAs were transmitted in the early stages of the COVID-19 pandemic. As most WEA messages focus on compliance with the government’s social distancing guidelines, one of the goals of transmitting WEAs during the COVID-19 pandemic is to control the floating population at an appropriate level.

Methods: We investigated the empirical impact of WEAs on the floating population across 25 districts in Seoul by estimating a panel regression model at the district-hour level with a series of fixed effects. The main independent variables were the number of instant WEAs, the daily cumulative number of WEAs, the total cumulative number of WEAs, and information extracted from WEAs by natural language processing at the district-hour level. The data set provided a highly informative empirical setting as WEAs were sent by different local governments with various identifiable district-hour-level data.

Results: The estimates of the impact of WEAs on the floating population were significantly negative (−0.013, P=.02 to −0.014, P=.01) across all specifications, implying that an additional WEA issuance reduced the floating population by 1.3% (=100(1−e−0.013)) to 1.4% (=100(1−e−0.014)). Although the coefficients of DCN (the daily cumulative number of WEAs) were also negative (−0.0034, P=.34 to −0.0052, P=.15) across all models, they were not significant. The impact of WEAs on the floating population doubled (−0.025, P=.02 to −0.033, P=.005) when the first 82 days of observations were used as subsamples to reduce the possibility of people blocking WEAs.

Conclusions: Our results suggest that issuing WEAs and distributing information related to COVID-19 to a specific district was associated with a decrease in the floating population of that district. Furthermore, among the various types of information in the WEAs, location information was the only significant type of information that was related to a decrease in the floating population. This study makes important contributions. First, this study measured the impact of WEAs in a highly informative empirical setting. Second, this study adds to the existing literature on the mechanisms by which WEAs can affect public response. Lastly, this study has important implications for making optimal WEAs and suggests that location information should be included.

(JMIR Public Health Surveill 2024;10:e43554) doi:10.2196/43554
KEYWORDS
COVID-19; empirical identification; floating population; social distancing; wireless emergency alert

Introduction

The frequency of disasters and the scale of the damage have increased dramatically over the decades. Natural disasters related to climate, weather, or water hazards have increased more than 5-fold over the past 50 years. Such incidents have resulted in approximately 115 casualties and US $202 million in damage per day [1]. Likewise, man-made disasters, including criminal attacks (eg, shooter incidents and terrorist attacks), human-based technological incidents (eg, building or bridge collapses), and events related to the manufacture, transportation, storage, and use of hazardous materials, have also increased significantly over the past 50 years. Moreover, the number of occurrences was estimated to have increased 6-fold or more, whereas the amount of damage increased by approximately 189 times [2].

Along with a significant number of disasters, the importance of wireless emergency alerts (WEAs) to rapidly deliver disaster information is emphasized. Many countries, including the United States, Japan, and South Korea, have adopted WEAs as a disaster information delivery system channel. This is because WEAs operate as a cell broadcast service (CBS) system that can send messages to multiple mobile phone users in a defined area simultaneously. Information is delivered directly to an individual’s cell phone; hence, it can be easily accessed through an individual’s mobile phone without the hassle of searching.

WEAs have been actively used to prevent the spread of COVID-19 since early 2020 [3]. In the United States, Utah and Massachusetts have issued WEAs reminding the populace of safety rules related to COVID-19 [4,5]. In New Zealand, WEAs were sent to mobile phone users across the country in the name of the National Emergency Management Agency [6]. Taiwan also used a cell-based mass texting system known as the Public Warning CBS to share information on COVID-19 [7].

One of the most efficient ways to respond to a disaster situation is to evacuate people from the danger zone to a safer location. Therefore, it is necessary to monitor changes in the flow of public movement, including the floating population, among the various public responses that occur during such disasters. Along with the growing frequency of disasters, there is increasing demand for countries or local governments to examine the impact of WEAs, as it could be one of the solutions to prevent severe damage from emergent events. WEAs are an information delivery channel optimized for this purpose, enabling people in a specific area to receive disaster notifications on their cell phones and prompting them to evacuate. Therefore, analyzing the relationship between the issuance of WEAs and the size of the floating population in a given area is the most intuitive and effective way to measure the policy performance of sending WEAs, which provides important implications for developing future policies. However, little is known about the impact of WEAs on people’s behavior because granular data sets of the focal outcome variable at various regions and short-time levels are unlikely to be available (eg, the real-time population at a specific time in a particular region). Only a few studies have examined the impact of WEAs, and they focused on either the impact of a single region [8] or simply compared the delivering and nondelivering regions [3].

We empirically investigated the associations between WEAs and the floating population in 25 districts in Seoul, the largest metropolis in South Korea, in the early stages of the COVID-19 pandemic for 13 months, from January 2020 to January 2021. The floating population denotes the transient influx of individuals within a defined area, and this metric serves as a pivotal indicator, offering perspectives on the adherence to social distancing protocols during the COVID-19 pandemic. As most WEA messages focus on compliance with the government’s social distancing guidelines, we can safely say that the goal of transmitting WEAs during the COVID-19 pandemic is to control the floating population at an appropriate level.

Methods

Data Set

In May 2005, the South Korean government established a nationwide WEA transmission system for the first time in the world, and since 2013, the WEA reception function has been mandatory for all mobile phones. WEAs operate as a CBS system, a method of simultaneously sending messages to multiple mobile phone users in a defined area. CBS is a one-to-many geo-targeted and geo-fenced messaging service, unlike SMS.

Since the establishment of the WEA transmission system, only the central government has had the authority to send WEAs, but this authority was extended to local governments in 2019 with an amendment to the relevant law. Since then, local governments have sent WEAs related to natural disasters, such as storms, floods, conflagrations, and other issues in their jurisdictions. WEA sending standards and the composition of contents are the subjects of the authority’s autonomy. Unlike WEAs sent by the central government, those from the local governments are sent only to their jurisdictional areas.

We have collected 6442 WEAs sent by each local government in the designated district (Gu—administrative district in Seoul, and there are 25 of them) in Seoul for 13 months from January 2020 to January 2021. Detailed WEAs were extracted from the “Public Data Portal” built by the Ministry of the Interior and Safety of the South Korean government. The empirical period covers the early stage of the COVID-19 pandemic in South Korea.

To investigate what kind of information should be included in the WEAs, we extracted information from the WEAs using natural language processing. The first information extracted from WEAs was whether or not the WEAs contained location information. The second indicated whether the WEAs had up-to-date information. The third indicated whether the WEAs had information on hand (eg, number of confirmed cases). The
final one indicated whether the WEAs had disease information (eg, detailed information on COVID-19).

For the dependent variable, we used the estimated floating population provided by SK Telecom, which has the largest number of subscribers in South Korea, at the district-hour level during the same period. We summarized the materials that aid in the understanding of our dependent and independent variables in Multimedia Appendix 1.

**Empirical Model**

We specified our empirical model at the region-hour level as follows:

\[
\text{Dependent variable} = \beta_0 + \beta_1 N_{it} + \beta_2 DCN_{it} + \beta_3 TCN_{it} + \beta_4 X_{it} + \sum \gamma_j \text{Fixed effects} + \epsilon_{it}
\]

where \( \text{Dependent variable} \) denotes the dependent variable in district \( i \) at time \( t \); \( N_{it} \) indicates the number of WEAs in district \( i \) at time \( t \); \( DCN_{it} \) indicates the daily cumulative number of WEAs in district \( i \) until time \( t \) (excluding the number of WEAs at time \( t \)); \( TCN_{it} \) indicates the total cumulative number of WEAs in district \( i \) until time \( t \) since January 1, 2020; \( X_{it} \) indicates the information extracted from WEAs with natural language processing; and \( \sum \gamma_j \) indicates the group of fixed effects that includes district-fixed effects, time-fixed effects, and joint district and time-fixed effects. To control the effect of unobservable factors (eg, daily work commute), we included these granular factors, weekday, and hour-fixed effects. This specification alleviated the concern about controlling unobservable factors that could affect the floating population.

The dependent variable was the floating population at time \( t \) in district \( i \). We specified continuous variables in logarithms since they have a skewed distribution (eg, \( \ln N_{it} \) and \( \ln TCN_{it} \)). As appropriate, we added 1 to the variables to avoid logarithms of 0. Table 1 shows the distribution of \( N \) and \( DCN \) during the day. The first column represents 24 hours. For instance, 0 indicates a period from midnight to 1 AM. The second and third columns represent the average \( N \) (in gray) and \( DCN \) (in light gray) across 25 districts. As shown in Table 1, \( DCN \) increased over time during a day as \( N \) accumulated. On average, people received 0.65 WEAs daily, and the peak time was 6 PM. Tables 2 and 3 report the summary statistics and correlations of the variables. Most of the correlations are significant (\( P<.001 \)) except for a correlation between floating population and number of WEAs with order information, as shown in Table 3. The correlations between each variable were mostly at a modest level, and considering our sample size (the number of observations in the analysis is 236,775), we concluded that each variable had different variations. We used the ordinary least squares method to estimate the unknown parameters of our empirical model.

Various types of information were sent through WEAs, and natural language processing techniques were used to extract and identify them. Sentences in WEAs were stemmed through a morphology analyzer named “Mecab” and “Okt tagger” in KoNLPY, an open-source Korean language processing package. After classifying common nouns, proper nouns, adjectives, and verbs through the morphology analyzer, informative words were extracted by specifying a list for the exclusion of stopwords and out-of-vocabulary words in advance. The type of information included in the WEAs was identified by classifying the extracted words according to a predetermined semantic standard (named entity recognition; NER). We used the NER application programming interface (API) of the AI API/Data service provided by the Electronics and Telecommunications Research Institute, Korea; this API is known to have a high performance of 89.4% in Korean NER [9].
Table 1. N (number of wireless emergency alerts [WEAs]) and DCN (daily cumulative number of WEAs) at hour level for WEAs issued in 25 districts of Seoul between January 2020 and January 2021.

<table>
<thead>
<tr>
<th>Hour</th>
<th>N</th>
<th>DCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>9</td>
<td>0.030</td>
<td>0.008</td>
</tr>
<tr>
<td>10</td>
<td>0.058</td>
<td>0.038</td>
</tr>
<tr>
<td>11</td>
<td>0.052</td>
<td>0.097</td>
</tr>
<tr>
<td>12</td>
<td>0.018</td>
<td>0.149</td>
</tr>
<tr>
<td>13</td>
<td>0.036</td>
<td>0.166</td>
</tr>
<tr>
<td>14</td>
<td>0.042</td>
<td>0.203</td>
</tr>
<tr>
<td>15</td>
<td>0.047</td>
<td>0.244</td>
</tr>
<tr>
<td>16</td>
<td>0.056</td>
<td>0.291</td>
</tr>
<tr>
<td>17</td>
<td>0.088</td>
<td>0.347</td>
</tr>
<tr>
<td>18</td>
<td>0.122</td>
<td>0.435</td>
</tr>
<tr>
<td>19</td>
<td>0.047</td>
<td>0.557</td>
</tr>
<tr>
<td>20</td>
<td>0.025</td>
<td>0.604</td>
</tr>
<tr>
<td>21</td>
<td>0.016</td>
<td>0.629</td>
</tr>
<tr>
<td>22</td>
<td>0.003</td>
<td>0.645</td>
</tr>
<tr>
<td>23</td>
<td>0.000</td>
<td>0.649</td>
</tr>
</tbody>
</table>

Table 2. Descriptive statistics of the floating population and wireless emergency alerts (WEAs) across 25 districts in Seoul between January 2020 and January 2021.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Median</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Floating population</td>
<td>319,000</td>
<td>327,382</td>
<td>94,020</td>
<td>1,044,150</td>
</tr>
<tr>
<td>2. Number of WEAs (N&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>0</td>
<td>0.027</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>3. Number of daily cumulative WEAs (DCN&lt;sup&gt;b&lt;/sup&gt;)</td>
<td>0</td>
<td>0.211</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>4. Number of total cumulative WEAs (TCN&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>54</td>
<td>84.210</td>
<td>0</td>
<td>495</td>
</tr>
<tr>
<td>5. Number of WEAs with location information</td>
<td>0</td>
<td>0.017</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>6. Number of WEAs with date information</td>
<td>0</td>
<td>0.007</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>7. Number of WEAs with order information</td>
<td>0</td>
<td>0.040</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>8. Number of WEAs with disease information</td>
<td>0</td>
<td>0.005</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

<sup>a</sup>N: number of WEAs.

<sup>b</sup>DCN: the daily cumulative number of WEAs in a specific district.

<sup>c</sup>TCN: the total cumulative number of WEAs in a specific district since January 1, 2020, the beginning of the observation period.
Table 3. Correlation matrix of the floating population and wireless emergency alerts (WEAs) across 25 districts in Seoul between January 2020 and January 2021.

<table>
<thead>
<tr>
<th>Correlationa</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Log (floating population)</td>
<td>1.000</td>
<td>0.016</td>
<td>0.028</td>
<td>0.153</td>
<td>0.005b</td>
<td>0.029</td>
<td>0.001c</td>
<td>0.015</td>
</tr>
<tr>
<td>(2) Number of WEAs (N)</td>
<td>0.016</td>
<td>1.000</td>
<td>0.042</td>
<td>0.106</td>
<td>0.564</td>
<td>0.437</td>
<td>0.805</td>
<td>0.400</td>
</tr>
<tr>
<td>(3) Number of daily cumulative WEAs (DCN)</td>
<td>0.028</td>
<td>0.042</td>
<td>1.000</td>
<td>0.265</td>
<td>0.014</td>
<td>0.025</td>
<td>0.028</td>
<td>0.005b</td>
</tr>
<tr>
<td>(4) Number of total cumulative WEAs (TCN)</td>
<td>0.153</td>
<td>0.106</td>
<td>0.265</td>
<td>1.000</td>
<td>0.033</td>
<td>0.042</td>
<td>0.103</td>
<td>0.023</td>
</tr>
<tr>
<td>(5) Number of WEAs with location information</td>
<td>0.005b</td>
<td>0.564</td>
<td>0.014</td>
<td>0.033</td>
<td>1.000</td>
<td>0.216</td>
<td>0.469</td>
<td>0.173</td>
</tr>
<tr>
<td>(6) Number of WEAs with date information</td>
<td>0.029</td>
<td>0.437</td>
<td>0.025</td>
<td>0.042</td>
<td>0.216</td>
<td>1.000</td>
<td>0.285</td>
<td>0.195</td>
</tr>
<tr>
<td>(7) Number of WEAs with order information</td>
<td>0.001c</td>
<td>0.805</td>
<td>0.028</td>
<td>0.103</td>
<td>0.469</td>
<td>0.285</td>
<td>1.000</td>
<td>0.284</td>
</tr>
<tr>
<td>(8) Number of WEAs with disease information</td>
<td>0.015</td>
<td>0.400</td>
<td>0.005b</td>
<td>0.023</td>
<td>0.173</td>
<td>0.195</td>
<td>0.284</td>
<td>1.000</td>
</tr>
</tbody>
</table>

a All correlations are significant (P < .001) if there is no note.
b Correlation is significant (P = .01).
c Correlation is insignificant (P = .65).
d N: number of WEAs.
e DCN: the daily cumulative number of WEAs in a specific district.
f TCN: the total cumulative number of WEAs in a specific district since January 1, 2020, the beginning of the observation period.

Ethical Considerations

To compile the data set used in this study, publicly available information was collected through the Ministry of the Interior and Safety of the Republic of Korea. We received confirmation of exemption from the Public Institutional Bioethics Committee designated by the Ministry of Health and Welfare of the Republic of Korea (P01-202309-01-043).

Results

Parameter estimates of our empirical model are shown in Table 4. We included district, weekday, hour, and joint district-hour fixed effects in all regressions.

As shown in the first row of Table 4 (N), the estimates were significantly negative across all models, implying that when people received WEAs, the floating population decreased. Based on the estimate in column 2, one increase in WEA was related to a reduction of the floating population by 1.3% (100(1 – e−0.013)). Other specifications yielded similar results, ranging from 1.3% (100(1 – e−0.013)) to 1.4% (100(1 – e−0.014)). Although the coefficients of DCN were also negative across all models and related to a reduction of the floating population from 0.34% (100(1 – e−0.0034)) to 0.52% (100(1 – e−0.0052)), they were insignificant. To check the impact of cumulative WEAs, we incorporated the TCN into the main specification, as shown in column 2 of Table 4. The results show that cumulative WEAs were also negatively correlated with the floating population. We replaced TCN with TCC (total number of confirmed cases of COVID-19) and Day (daily time trend) to check the robustness and summarize the results in columns 3 and 4 of Table 4, respectively. We used the variables related to TCN, TCC, and Day separately due to their high correlation (0.963 between log (TCN) and log (TCC); 0.902 between log (TCN) and log (Day); and 0.899 between log (TCC) and log (Day)). As shown in columns 3 and 4, TCC and Day had negative impacts on the floating population, like TCN.

We conducted several robustness checks with subsamples and different specifications. The results of the robustness checks were consistent with our main findings. Subsequently, we performed 2 falsification tests, and we found that our results did not hold when using incorrect variables in terms of district and timing. Detailed information and results of the robustness checks and falsification tests are summarized in Multimedia Appendix 2.
Table 4. Parameter estimates of the main empirical model: panel regression model using wireless emergency alerts (WEAs) and the floating population at the district-hour level with a series of fixed effects.

<table>
<thead>
<tr>
<th></th>
<th>(1) Base</th>
<th>(2) With TCN^a</th>
<th>(3) With TCC^b</th>
<th>(4) With day</th>
</tr>
</thead>
<tbody>
<tr>
<td>N^c</td>
<td>−0.014 (0.0053; P=0.01)</td>
<td>−0.013 (0.0050; P=0.02)</td>
<td>−0.013 (0.0048; P=0.01)</td>
<td>−0.013 (0.0050; P=0.02)</td>
</tr>
<tr>
<td>DCN^d</td>
<td>−0.0052</td>
<td>−0.0034</td>
<td>−0.0039</td>
<td>−0.0037</td>
</tr>
<tr>
<td>(0.0035; P=0.15)</td>
<td>(0.0035; P=0.34)</td>
<td>(0.0034; P=0.26)</td>
<td>(0.0035; P=0.30)</td>
<td></td>
</tr>
<tr>
<td>Log (TCN)</td>
<td>N/A^e</td>
<td>−0.0019 (0.0005; P=0.002)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Log (TCC)</td>
<td>N/A</td>
<td>N/A</td>
<td>−0.0012 (0.0006; P=0.06)</td>
<td>N/A</td>
</tr>
<tr>
<td>Log (Day)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>−0.0033 (0.0010; P=0.004)</td>
</tr>
<tr>
<td>Region FE^f</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Weekday FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hour FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Region x weekday x hour FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>236,775</td>
<td>236,775</td>
<td>236,775</td>
<td>236,775</td>
</tr>
<tr>
<td>R^2</td>
<td>0.804</td>
<td>0.804</td>
<td>0.804</td>
<td>0.804</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.804</td>
<td>0.804</td>
<td>0.804</td>
<td>0.804</td>
</tr>
</tbody>
</table>

^aTCN denotes the total cumulative number of WEAs in a specific district since January 1, 2020, the beginning of the observation period.
^bTCC denotes the total number of confirmed cases of COVID-19.
^cN: number of WEAs.
^dDCN denotes the daily cumulative number of WEAs in a specific district.
^eNot applicable.
^fFE: fixed effect.

Discussion

Principal Results

This study empirically measured the association between the details of issuing WEAs and the granular region-hour-level estimated floating population changes in 25 districts in Seoul, South Korea. We used variations in the number and content of WEAs independently managed by local authorities to empirically measure the WEAs’ impact. By estimating a panel-data model with a series of fixed effects, we found robust evidence that issuing WEAs was related to the reduction of the floating population in districts of Seoul. A collection of robustness and falsification tests showed that our findings were robust. Furthermore, we found that location was the only significant type of information among various types extracted from WEAs that was associated with a decrease in the floating population.

The information necessary for disaster response must be delivered promptly and accurately to overcome disaster situations. WEAs, which deliver disaster information directly to the recipients’ mobile phones, have become a major means of providing disaster information. One of the most efficient ways to respond to a disaster situation is to evacuate people from dangerous areas to safe locations. Therefore, a major objective of issuing WEAs is to control or reduce the floating population in hazardous areas to an appropriate level. These policy objectives gained further prominence during the COVID-19 pandemic, as the primary content of WEAs at the peak of the pandemic emphasized practicing social distancing and avoiding densely populated areas. Hence, an intuitive approach to measuring the impact of WEAs is to examine the relationship between the number of WEAs sent and the size of the floating population in a given area. Our empirical data set was ideal for this study due to the following reasons: (1) The data set provided a highly informative empirical setting to measure the impact of WEAs, as different local governments sent WEAs with various region-hour-level data in districts of Seoul, which are identifiable; (2) The data set contained granular information on floating populations with various region-hour–level data, which were compatible with the WEA data set; (3) WEAs are mandatory for mobile users; and lastly, (4) the empirical data set fully covered the COVID-19 pandemic in South Korea after the first confirmed case was reported, which provides us with a clear impact in a long-enough range. Overall, our empirical data were ideal for studying our research question.

This study makes several contributions. First, this study offers new empirical evidence on the impact of WEAs, advancing previous research that mainly relied on hypothetical scenario surveys due to limited WEAs and challenges in identifying key variables [10-12]. Earlier studies, particularly those examining COVID-19, faced difficulties in robust statistical analysis and data collection reflecting public responses [3,13]. Unlike previous work that focused on a few regions or solely on the...
number of WEAs, this study stands out by analyzing variations in the number and content of WEAs across different regions due to the regions’ independent authorities issuing them [8].

Second, this study provides a plausible mechanism for how WEAs can reduce other focal outcomes, such as the numbers of infections and deaths [3], which remain unclear. Previous studies revealed that population density has a strong correlation with the number of infections and deaths from COVID-19 [14-17]. Transmitting WEAs reduces the floating population in the area where WEAs are sent and consequently reduces the confirmed COVID-19 cases and deaths.

Lastly, this study provides important implications for developing a policy on making optimal WEAs to prevent national disasters or pandemics, which suggests that location information should be highlighted rather than other information (eg, date and general information).

Limitations
This study has some limitations due to technical issues with the WEA system and the difficulty in obtaining data sets. This study observed the relationship between the transmission history of WEAs in districts and the change in the floating population. However, we could not confirm whether the people who received WEAs moved to a safe place, which makes it hard to argue for a causal relationship. To supplement this limitation, we require an additional data set that can check the WEA confirmation time and movement history through personal-level mobile phone log records.

We used the estimated floating population provided by SK Telecom as a dependent variable. However, it is worth noting that while SK Telecom has the largest number of subscribers, with a market share of 44.3% as of 2021 [18], the floating population they estimated and provided may have differed from the actual floating population. We conducted an additional analysis using the floating population provided by the Seoul local government as the dependent variable, and the results are summarized in Multimedia Appendix 2. The implications of this study remained consistent even when analyzed with an additional data set.

This study focused on WEAs and the floating population of 25 districts in Seoul, the capital and largest city of South Korea. For a more generalized conclusion about the effectiveness of WEAs, it is necessary to expand and analyze the data set to include more regions in future studies.

WEAs can be blocked by changing mobile phone settings; in that case, the impact of transmitting WEAs on behavioral responses may vary. Nevertheless, due to technical constraints, the government cannot collect data on the rate of WEA blocking. The WEA sending system should be developed so that the WEA reception or blocking rate can be aggregated and used for policy formulation.

Conclusions
This study analyzed the impact of WEAs sent in the 25 districts of Seoul on changes in the floating population in those areas, using the fact that the content of the WEAs sent during the COVID-19 pandemic was focused on compliance with social distancing guidelines. The results showed that issuing WEAs and distributing information related to COVID-19 to a specific district in Seoul, South Korea, was related to the reduction of the floating population of that district. Furthermore, among the 4 types of information (location, date, order, and disease) in the WEAs, location information was the only significant type that was associated with a decrease in the floating population. Contrasting with previous studies that could not generalize findings by using hypothetical survey results or analyzing WEAs and the floating population in a single region, this study empirically elucidates the impact of WEAs on public behavior by examining the number and content of region-hour–level messages sent by 25 districts with independent dispatch authority, along with the corresponding changes in the floating population.

Our findings provided a plausible mechanism by which WEAs could effectively suppress infectious diseases by providing clues that transmitting WEAs was related to the reduction of the floating population during the COVID-19 pandemic. Our findings focused on a correlational relationship rather than a causal one, as we were unable to observe whether people who received WEAs moved to safer places.

We hope that this research will spur further investigation into the sophisticated measurement of WEA impact and the establishment of disaster response systems.

Acknowledgments
This study was edited in the form of an academic article by modifying and supplementing a portion of the Korea Information Society Development Institute Research Report titled “The Effect of Utilizing Public Data in Disaster: An Empirical Analysis.” This study was supported by the Institute of Management Research at Seoul National University.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Summary of the number of wireless emergency alerts across districts (Table S1) and monthly number of wireless emergency alerts across districts (Table S2).

[DOCX File, 33 KB - publichealth_v10i1e43554_app1.docx]
References
9. AI API/DATA. URL: https://aiopen.etri.re.kr/ [accessed 2024-03-01]

Abbreviations
API: application programming interface
CBS: cell broadcast service
NER: named entity recognition
WEA: wireless emergency alert

https://publichealth.jmir.org/2024/1/e43554
Assessing SARS-CoV-2 Testing Adherence in a University Town: Recurrent Event Modeling Analysis

Yury E García1*, PhD; Alec J Schmidt1*, MSc; Leslie Solis2*, MS; Marí a L Daza-Torres1*, PhD; J Cricelio Montesinos-López1*, PhD; Brad H Pollock1*, PhD; Miriam Nuño1*, PhD

1Department of Public Health Sciences, University of California, Davis, CA, United States
2Clinical and Translational Science Center, University of California, Davis, CA, United States
*all authors contributed equally

Corresponding Author:
Miriam Nuño, PhD
Department of Public Health Sciences
University of California
1 Shields Avenue
Davis, CA, 95616
United States
Phone: 1 5056999560
Email: mnuno@ucdavis.edu

Abstract

Background: Healthy Davis Together was a program launched in September 2020 in the city of Davis, California, to mitigate the spread of COVID-19 and facilitate the return to normalcy. The program involved multiple interventions, including free saliva-based asymptomatic testing, targeted communication campaigns, education efforts, and distribution of personal protective equipment, community partnerships, and investments in the local economy.

Objective: This study identified demographic characteristics of individuals that underwent testing and assessed adherence to testing over time in a community pandemic-response program launched in a college town in California, United States.

Methods: This study outlines overall testing engagement, identifies demographic characteristics of participants, and evaluates testing participation changes over 4 periods of the COVID-19 pandemic, distinguished by the dominant variants Delta and Omicron. Additionally, a recurrent model is employed to explore testing patterns based on the participants’ frequency, timing, and demographic characteristics.

Results: A total of 770,165 tests were performed between November 18, 2020, and June 30, 2022, among 89,924 (41.1% of total population) residents of Yolo County, with significant participation from racially or ethnically diverse participants and across age groups. Most positive cases (6351 of total) and highest daily participation (895 per 100,000 population) were during the Omicron period. There were some gender and age-related differences in the pattern of recurrent COVID-19 testing. Men were slightly less likely (hazard ratio [HR] 0.969, 95% CI 0.943-0.996) to be retested and more likely (HR 1.104, 95% CI 1.075-1.134) to stop testing altogether than women. People aged between 20 and 34 years were less likely to be retested (HR 0.861, 95% CI 0.828-0.895) and more likely to stop testing altogether (HR 2.617, 95% CI 2.538-2.699). However, older age groups were less likely to stop testing, especially those aged between 65-74 years and 75-84 years, than those aged between 0 and 19 years. The likelihood of stopping testing was lower (HR 0.93, 95% CI 0.889-0.976) for the Asian group and higher for the Hispanic or Latino (HR 1.185, 95% CI 1.148-1.223) and Black or African American (HR 1.198, 95% CI 1.054-1.350) groups than the White group.

Conclusions: The unique features of a pandemic response program that supported community-wide access to free asymptomatic testing provide a unique opportunity to evaluate adherence to testing recommendations and testing trends over time. Identification of individual and group-level factors associated with testing behaviors can provide insights for identifying potential areas of improvement in future testing initiatives.

(JMIR Public Health Surveill 2024;10:e48784) doi:10.2196/48784

KEYWORDS
Healthy Davis Together; COVID-19; COVID-19 surveillance program; community surveillance; HDT: HYT; Healthy Yolo Together; SARS-CoV-2; severe acute respiratory syndrome coronavirus 2; coronavirus; demographic; demographics; testing;
adherence; compliance; USA; United States; response program; response programs; engagement; participation; infectious; trend; trends; community based; surveillance; public health; infection control; PCR; polymerase chain reaction; RT-qPCR; reverse transcription quantitative polymerase chain reaction; viral; virus; viruses

**Introduction**

The COVID-19 pandemic presented major challenges and demanded effective public health responses involving sustained behavior change, stay-at-home restrictions, face coverings, testing, and vaccination. Public health guidelines are a critical first line of a pandemic response but rely on compliance with evolving recommendations and restrictions. Trust and compliance are driven by complex factors, including individuals’ beliefs [1], risk perception [2,3], trust in government [4], and demographic factors [5-7] that may vary throughout a pandemic, making containment particularly challenging.

Widespread testing and timely diagnosis are critical for pandemic control and preparedness. The deployment of clinical testing on a massive scale (“mass testing” [8]) was one of the essential control measures for curtailing the burden of the COVID-19 pandemic, particularly during its early phases [9]. Early detection of cases can limit viral transmission by isolation, quarantine, and contact tracing [10-12]. New preventive and surveillance mechanisms emerged throughout the pandemic, including vaccine programs, wastewater surveillance [13], and at-home COVID-19 tests [14]. Both the virus itself and protective behaviors changed throughout the waves of the pandemic [1], complicating long-term interventions. Government and public health officials implemented health contingencies that were adaptable and flexible to reduce socioeconomic and long-term health burden fatigue [15,16].

Understanding how people’s adherence to preventive measures changes over time can guide policy makers as they amend strategies to revitalize public health strategies in future outbreaks. Studies of previous natural disasters suggest that people’s perceptions of risk and their responses to such risk may vary between individuals. Gender, age, socioeconomic status, personal experience of a natural hazard, and trust in authorities affect people’s responses to catastrophic events [6,17,18]. For the COVID-19 pandemic, a growing body of research reported factors associated with adherence and nonadherence to COVID-19 preventive measures [1,19-21]. Specifically, age [5], gender [6,7], higher education level [22], marital status, and having children [23] are all associated with adherence measures.

Healthy Davis Together (HDT) [24] was a program implemented in the city of Davis, California, as an effort to mitigate the spread of COVID-19 and facilitate the return to normalcy. Launched in September 2020, HDT focused on curtailing the pandemic burden among Davis residents and workers. In July of 2021, the program was expanded to support residents of the wider Yolo County in a rebranded effort, Healthy Yolo Together (HYT). It employed a multipronged approach, including engaging the community, implementing effective testing and contact tracing measures, promoting health education, and fostering collaboration with stakeholders.

HDT implemented comprehensive interventions that combined disease control measures and the promotion of health-conscious behaviors. Polymerase chain reaction (PCR) testing accessibility was enhanced with the establishment of new testing locations offering voluntary and complimentary SARS-CoV-2 testing. HDT engaged over 200 public health ambassador students dedicated to advocating for healthy behaviors. Mass communication campaigns were promoted across diverse media platforms to encourage testing and the adoption of health-promoting behavior. The program provided incentives to promote health-promoting behaviors. Collaborations were established with local businesses to implement and adjust safety protocols. The program fostered stronger partnerships among the university, city, county, and private and community organizations. In addition, HDT implemented citywide subsewer shed–level wastewater monitoring for early detection. The program maintained an analytical team tasked with real-time data analysis to pinpoint communities or specific locations requiring increased testing implementation. This analytical approach complemented the program’s health behavior strategies.

In this study, we thoroughly examine testing behavior within the HDT program, emphasizing patterns of testing throughout. We use regression modeling for recurrent events to characterize testing behavior within the HDT program. This method enables us to examine individual testing events and current testing over time, providing insights about the participants’ testing patterns and testing adherence throughout the program’s duration.

**Methods**

**Data Description**

Testing participation was aggregated and subdivided into distinct periods according to the dominant variants Delta and Omicron: pre-Delta (November 18, 2020, to June 13, 2021), Delta (June 14, 2021, to December 20, 2021), Omicron (December 21, 2021, to March 15, 2022), and post-Omicron (March 16, 2022, to June 30, 2022). These periods were based on the most common disease control measures and the promotion of health-conscious behaviors. Polymerase chain reaction (PCR) testing accessibility was enhanced with the establishment of new testing locations offering voluntary and complimentary SARS-CoV-2 testing. HDT engaged over 200 public health ambassador students dedicated to advocating for healthy behaviors. Mass communication campaigns were promoted across diverse media platforms to encourage testing and the adoption of health-promoting behavior. The program provided incentives to promote health-promoting behaviors. Collaborations were established with local businesses to implement and adjust safety protocols. The program fostered stronger partnerships among the university, city, county, and private and community organizations. In addition, HDT implemented citywide subsewer shed–level wastewater monitoring for early detection. The program maintained an analytical team tasked with real-time data analysis to pinpoint communities or specific locations requiring increased testing implementation. This analytical approach complemented the program’s health behavior strategies.

In this study, we thoroughly examine testing behavior within the HDT program, emphasizing patterns of testing throughout. We use regression modeling for recurrent events to characterize testing behavior within the HDT program. This method enables us to examine individual testing events and current testing over time, providing insights about the participants’ testing patterns and testing adherence throughout the program’s duration.

**Methods**

**Data Description**

Testing participation was aggregated and subdivided into distinct periods according to the dominant variants Delta and Omicron: pre-Delta (November 18, 2020, to June 13, 2021), Delta (June 14, 2021, to December 20, 2021), Omicron (December 21, 2021, to March 15, 2022), and post-Omicron (March 16, 2022, to June 30, 2022). These periods were based on the most common variant reported by the California Department of Public Health [25] (Figure 1). Hereafter, the Delta B.1.617.2, Omicron B.1.1.529, and subsequent Omicron BA.2, BA.3, BA.4,266, and BA.5 variants will be referred to as Delta, Omicron, and post-Omicron, respectively. During the pre-Delta period, HDT primarily served the city of Davis. Subsequently, at the onset of the Delta variant in July 2021, testing access was extended to encompass Yolo County as a whole.
Figure 1. Tests performed per 100,000 population (light blue) and positive test rates (dark blue) daily from November 18, 2020, to June 30, 2022. Free asymptomatic testing was accessible to Yolo County residents in July 2021. Vertical lines mark 4 distinct periods characterized by dominant SARS-CoV-2 variants.

Yolo County has an estimated population of 218,774 individuals, with 106,002 males and 112,772 females. Among them, approximately 68,640 reside in the city of Davis. Throughout the program, 770,165 tests were administered to 89,924 unique individuals, including 53,869 residents of the city of Davis. A significant number of participants in Davis were linked to the University of California, Davis. The university enforced weekly testing mandates starting in the fall of 2020 for those who were physically present on campus.

HDT program participants were given a unique ID, which allowed us to distinguish between new and returning individuals. An individual can be classified as a unique participant for a specific period (eg, Delta) if they underwent testing at least once during that period, regardless of whether they were also tested during another period. However, for the overall program analysis, each participant was counted only once, regardless of how many times they were tested across multiple periods. New participants are individuals who underwent testing for the first time. Total tests refer to the overall number of tests conducted within a specific period or in total. Positive tests denote the total count of tests yielding a positive result, distinct from participants with a positive result, which indicates the number of individuals who had at least one positive test during a specified period.

Ethical Considerations
This study used deidentified data and received an exemption from institutional review by the University of California Davis Office of Research.

Diagnostic Tests
Saliva tests were conducted using an innovative high-throughput quantitative reverse transcriptase–PCR method [24]. Most tests (98.4%) comprised of saliva tests while the remaining 1.6% consisted of BinaxNOW (Abbott) rapid antigen tests administered to symptomatic individuals. Employing saliva samples for quantitative reverse transcription PCR tests enhances the feasibility of widespread asymptomatic testing in the community, as it incurs lower monetary and labor costs with only a slight decrease in sensitivity [26].

Statistical Models
We used regression modeling for recurrent events to analyze testing behavior, given that participants underwent multiple tests at irregular intervals. This modeling framework facilitates the analysis and interpretation of testing recurrence, providing insights into factors influencing testing adherence, frequency, and consistency among participants. Additionally, trends, fluctuations, and potential predictors of testing behavior can be identified, offering insights for optimizing testing strategies, resource allocation, and public health interventions to promote regular and sustained testing participation. The reReg library in R (R Foundation for Statistical Computing) was used to analyze the temporal dynamics of testing behavior [27,28].

The regression Cox model for recurrent events extends the traditional Cox proportional hazards model, which is commonly used for analyzing time-to-event data in survival analysis, to handle recurrent event data. Additionally, this modeling framework enables the examination of covariate effects on the rate of event occurrence while properly addressing the correlation among recurrent events within the same individual. This model, also known as the Andersen-Gill model, can be represented as follows:

$$\lambda_i(t) = \lambda_0(t) \times \exp (\beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip}) \quad (1)$$

where $\lambda_i(t)$ represents the hazard rate for individual $i$ at time $t$. $\lambda_0(t)$ is the baseline hazard rate at time $t$. $\beta_1, \beta_2, ..., \beta_p$ are the regression coefficients associated with the covariates $X_{i1}, X_{i2}, ..., X_{ip}$.

https://publichealth.jmir.org/2024/1/e48784
of the baseline hazard rate and the covariates, with the regression coefficients determining the impact of the covariates on the hazard rate. We assumed noninformative censoring, suggesting that the reasons people decided to take part in testing during the program are not related to why they might eventually stop testing altogether.

**Mean Cumulative Function**

The Mean Cumulative Function (MCF) denotes the expected cumulative count of tests up to a specific time point, considering both the occurrence and timing of test events. The MCF was calculated by age and racial or ethnic groups. We also apply the terminal event process Cox model to examine the occurrence of complete cessation of testing while accounting for covariates.

**Data Preprocessing**

For this study, individuals who undergo COVID-19 testing at least once a week are selected. Each selected individual is treated as a single event, even if they receive multiple tests within the same week. This study includes 61,363 participants who have undergone testing on at least 2 occasions (2 different weeks). The event for each participant begins with their first test and ends during the subsequent week of another test. A test occurrence is marked as 1, while the status remains 0 to signify ongoing participation. The final testing instance for each participant is identified, and the event of the subsequent week is marked as 0 to indicate a censoring event.

**Results**

**Description of Testing Trends**

HDT conducted 770,165 tests among 89,924 residents of Yolo County (approximately 41.1% of the Yolo population, 218,793), of which 53,869 individuals reported a zip code in the city of Davis (approximately 78.4% of Davis population, 68,710). Testing coverage was most significant in the city of Davis, where the program originated, and efforts were first focused. Figure 1 illustrates the total daily tests per 100,000 population and positive test rates observed in Yolo County throughout this study. On June 28, 2021, the positivity rate surged, likely because there were fewer tests reported on that day. Overall, 12,626 tests were positive among 11,545 unique individuals, for an aggregated test positivity rate of 1.64% and a case rate of 13%. The average daily number of tests conducted per 100,000 population was 422 (SD 210) during the pre-Delta period, 746 (SD 449) during Delta, 895 (SD 518) during Omicron, and 435 (SD 254) during the post-Omicron period. A summary of tests and positive cases per study period is reported in Table 1.

**Table 1. Summary of testing participation for each period of this study and overall.**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Pre-Delta</th>
<th>Delta</th>
<th>Omicron</th>
<th>Post-Omicron</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of the testing period (days)</td>
<td>208</td>
<td>190</td>
<td>85</td>
<td>107</td>
<td>590</td>
</tr>
<tr>
<td>Total tests, n (%)</td>
<td>191,904 (24.9)</td>
<td>309,928 (40.2)</td>
<td>166,482 (21.6)</td>
<td>101,851 (13.2)</td>
<td>770,165 (100)</td>
</tr>
<tr>
<td>Average tests per day (SD)</td>
<td>922.6 (460)</td>
<td>1631.2 (983)</td>
<td>1958.6 (1333)</td>
<td>951.9 (556)</td>
<td>1305.4 (919)</td>
</tr>
<tr>
<td>Average tests per 100,000 mean population per day (SD)</td>
<td>421.7 (210)</td>
<td>745.6 (449)</td>
<td>895.3 (518)</td>
<td>435.1 (254)</td>
<td>596.7 (420)</td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall participants</td>
<td>39,819</td>
<td>58,920</td>
<td>46,545</td>
<td>23,566</td>
<td>89,924</td>
</tr>
<tr>
<td>New participants, n (%)</td>
<td>39,819 (100)</td>
<td>33,814 (57.4)</td>
<td>13,503 (29)</td>
<td>2788 (11.8)</td>
<td>89,924</td>
</tr>
<tr>
<td><strong>Cases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total positive tests</td>
<td>1197</td>
<td>2817</td>
<td>6351</td>
<td>2261</td>
<td>12,626</td>
</tr>
<tr>
<td>Test positivity rate (%)</td>
<td>0.62</td>
<td>0.91</td>
<td>3.81</td>
<td>2.22</td>
<td>1.64</td>
</tr>
<tr>
<td>Percentage of all positives, %</td>
<td>9.5</td>
<td>22.3</td>
<td>50.3</td>
<td>17.9</td>
<td>100</td>
</tr>
<tr>
<td>Positives among participants, n</td>
<td>1138</td>
<td>2577</td>
<td>5862</td>
<td>2142</td>
<td>11,719</td>
</tr>
<tr>
<td>Positives among new participants, n (%)</td>
<td>1138 (100)</td>
<td>2570 (99.7)</td>
<td>5757 (98.2)</td>
<td>2080 (97.1)</td>
<td>11,545 (98.5)</td>
</tr>
</tbody>
</table>

Total tests conducted, participation, recruitment, and positive test tallies across pandemic periods are displayed in Table 1. For instance, during pre-Delta, 191,904 tests were conducted, which corresponds to 24.9% (n=770,165) of the total tests conducted in the entire program. During Delta, 2570 positive cases were detected in new participants, corresponding to 22.3% (n=11,545) of the total of unique people who resulted as positive.

Changes in policy at the state and local level, social expectations and behavior, the emergence of new strains, and differences in the intensity of testing lead to differing participation rates across pandemic periods. In total, HDT or HYT served 89,924 unique individuals across the entire program. A total of 39,819 (44.3%) participated in testing during pre-Delta, 58,920 (65.5%) during Delta, 46,545 (51.8%) during Omicron, and 23,566 (26.2%) during the post-Omicron period. Further, 61,363 individuals, equivalent to 68.2% of all participants, took part in the testing process multiple times. A longer period of days did not mean a higher number of participants. The highest participation of people was during the Delta period.

Omicron had the highest number of positive cases despite being the shortest period in the program. Out of all positive cases,
6351 (50.3%) were observed during Omicron. On average, 1958 (SD 1333; \( \sim 895, \text{SD 518, tests per 100,000} \)) daily tests were performed at a test positivity rate of 3.81% (Table 1). The enrollment of new participants decreased with each subsequent period: 39,819 individuals were engaged in the first stage of the program (pre-Delta), 33,814 during Delta, 13,503 during Omicron, and 2788 during post-Omicron. Most positive tests (a total of 2080, 97.1%) were from new participants (Table 1).

HDT or HYT collaborated with multiple school districts in Yolo County to provide free weekly testing to students. Excluding participants under 19, there were 67,083 participants from the community; among them, 8774 (13%) individuals tested positive, and 422 (4.8%) had positive results in two or more weeks. Further, 265,000 (34.4%) underwent testing once, while participation ranged from 1 to 83 weeks, with a median of 7.5 (IQR 7) weeks. Figure 2A illustrates uptake in testing by individuals, showing the weeks they underwent testing and information on when everyone was initially tested. It is noticeable that at the beginning of each testing wave, there is typically a higher flux of new participants signing up each week. It also enables the observation of participants’ ongoing engagement throughout the program. Figure 2B illustrates the weeks in which an individual received at least one positive test result, showing a notable increase in positive cases during the Omicron wave. Figure 2C describes key measures implemented during the Healthy Davis Program, with color bars representing the seasons: winter (blue), spring (green), summer (yellow), and fall (orange).

Testing rates between male and female residents varied across 4 distinct periods and overall. Female participants underwent more testing compared to males. Participants aged 18 or younger were less involved in testing during the pre-Delta period but increased their participation during the Delta period. Individuals aged 19-34 years underwent more testing in the pre-Delta period when compared to the Delta and post-Omicron periods. Participants who identified as Hispanics or Latinos had lower participation rates in post-Omicron (Table S1 in Multimedia Appendix 1 [29]).
Demographic Characteristics of Test Participants

A total of 48,670 participants identified as female, making up 45.9% of the 106,034 estimated female population in Yolo County. Meanwhile, 39,524 were male, representing 35% of the 112,759 estimated male population in the county. Testing in all race or ethnic groups and age groups younger than 85 years exceeded 54,698 (25%) of Yolo County’s population (Figure 3). The lowest testing participation rate was observed among individuals aged older than 85 years. This is not unexpected, given that testing for participants aged 85 years and older was compulsory at congregate living facilities funded by Medicare or Medicaid, as part of several state-level testing initiatives. The highest overall test positivity rates were observed among individuals of American Indian or Alaska Native descent (7.6%), Native Hawaiian or Other Pacific Islander (5.6%), Latinos (4.5%), and people aged 20-44 years (20-34 years, 6.5%; 35-44 years, 6.9%).

Figure 3. Rates of tests conducted and positive cases by race or ethnicity and age in years for Yolo County. The light-blue bars represent the total population size for each group based on the Census Bureau 2020 data. Dark-blue bars correspond to the proportion of participants within each demographic group. Red bars correspond to participants with a positive test result.

Adherence to Testing

In a sample of 61,363 participants that tested during this study period, there were 652,232 tests, with an average number of 10.62 (SD 13) tests per individual, in a median follow-up time of 40 (IQR 41.5) weeks. The average number of tests observed per week (MCF) by age and racial or ethnic groups is described in Figure 4A. Participants between the ages of 0 and 19 years demonstrated more consistent engagement in testing compared to other age groups, maintaining a higher level of involvement throughout this study’s period. Participants identifying as multiracial exhibited higher levels of engagement in testing compared to those identifying as Hispanic, Latino, American Indian, or Alaska Native (Figure 4B). Throughout this study period, the average testing patterns remained consistent across various age and racial or ethnic groups.

Figure 4 illustrates the baseline cumulative rate function as the expected number of tests conducted up to a particular time point, independent of any covariates included in the model (Figure 5A). We also describe the baseline cumulative hazard function in recurrent tests, representing the expected cumulative number of tests occurring up to a specific time point, without considering any covariates or risk factors (Figure 5B). It provides a baseline estimate of testing occurrence over time.
Figure 5. Baseline cumulative hazard function ((A) expected cumulative tests occurring up to a specific time) and baseline cumulative rate function ((B) expected tests up to a specific time).

**Adherence to Testing**

The recurrent event analysis showed that male participants (HR 0.969, 95% CI 0.943-0.996) and individuals 85 years and older (HR 0.781, 95% CI 0.598-1.020) experienced lower adherence to testing compared to females and younger age groups (0-19 years), respectively (Table 2). Hispanic or Latino participants (HR 0.741, 95% CI 0.712-0.772) and American Indian or Alaska Native individuals (HR 0.735, 95% CI 0.567-0.953) experienced lower testing rates compared to individuals from other racial or ethnic groups.

**Table 2.** Covariate adjusted adherence to testing analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard ratio</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (reference: female)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.969</td>
<td>0.943-0.996</td>
<td>.002</td>
</tr>
<tr>
<td>Age (years; reference: 0-19 years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td>0.861</td>
<td>0.828-0.895</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>35-44</td>
<td>0.800</td>
<td>0.765-0.836</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>45-54</td>
<td>0.840</td>
<td>0.806-0.876</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>55-64</td>
<td>0.870</td>
<td>0.829-0.913</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>65-74</td>
<td>0.782</td>
<td>0.750-0.816</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>75-84</td>
<td>0.768</td>
<td>0.714-0.826</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>85+</td>
<td>0.781</td>
<td>0.598-1.020</td>
<td>.01</td>
</tr>
<tr>
<td>Race and ethnicity (reference: White)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>0.735</td>
<td>0.567-0.953</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Asian</td>
<td>0.912</td>
<td>0.873-0.953</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0.865</td>
<td>0.748-1.001</td>
<td>.008</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>0.741</td>
<td>0.712-0.772</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Multiracial</td>
<td>0.988</td>
<td>0.928-1.052</td>
<td>.56</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>0.965</td>
<td>0.698-1.334</td>
<td>.70</td>
</tr>
</tbody>
</table>

**Discontinuation in Testing Participation**

Male participants showed a significantly higher risk of discontinuing testing compared to females (HR 1.10, 95% CI 1.08-1.13), see Table 3. Individuals aged 20-34 years (HR 2.62, 95% CI 2.54-2.70) exhibited the highest hazard of discontinuing testing, followed by those aged 35-44 years (HR 1.51, 95% CI 1.47-1.56) and 45-54 years (HR 1.31, 95% CI 1.27-1.35), compared to those aged 0-19 years. Participants identifying as multiracial (HR 1.52, 95% CI 1.20-1.94), American Indian or Alaska Native (HR 1.20, 95% CI 1.05-1.36), and Hispanic or Latino (HR 1.19, 95% CI 1.15-1.22) were more prone to discontinuing testing compared to White participants. These findings underscore the importance of targeted outreach strategies tailored to engage and retain the demographic groups at risk of stopping testing. Addressing barriers such as accessibility, convenience, and perceived need for testing among younger age groups and specific racial or ethnic groups may help mitigate the discontinuation of testing and continue surveillance of SARS-CoV-2 transmission within communities.
### Table 3. Covariate adjusted testing discontinuation analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hazard ratio</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex (reference: female)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.104</td>
<td>1.075-1.134</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Age (years; reference: 0-19 years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td>2.617</td>
<td>2.538-2.699</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>35-44</td>
<td>1.511</td>
<td>1.466-1.556</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>45-54</td>
<td>1.306</td>
<td>1.266-1.347</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>55-64</td>
<td>1.2</td>
<td>1.165-1.235</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>65-74</td>
<td>0.874</td>
<td>0.847-0.901</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>75-84</td>
<td>0.846</td>
<td>0.780-0.917</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>85+</td>
<td>1.185</td>
<td>0.936-1.501</td>
<td>.21</td>
</tr>
<tr>
<td><strong>Race and ethnicity (reference: White)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>1.157</td>
<td>0.964-1.390</td>
<td>.27</td>
</tr>
<tr>
<td>Asian</td>
<td>0.932</td>
<td>0.889-0.976</td>
<td>.002</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1.198</td>
<td>1.054-1.360</td>
<td>.02</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>1.185</td>
<td>1.148-1.223</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Multiracial</td>
<td>1.523</td>
<td>1.198-1.935</td>
<td>.01</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>0.969</td>
<td>0.908-1.034</td>
<td>.33</td>
</tr>
</tbody>
</table>

### Discussion

HDT, and later HYT, was a community COVID-19 surveillance program that focused on free asymptomatic and symptomatic testing to catch cases early and provide resources and information for transmission reduction strategies. The primary focus of this study was to assess how well the program reached and maintained coverage and equity goals by tallying the demographic characteristics of participants and changes in testing participation. The HDT or HYT program used a high-throughput method to administer and process large volumes of tests, which allowed it to reach a substantial portion of the city of Davis and Yolo County’s population, covering all who willingly underwent testing. Extra testing campaigns and expanded access to free testing were adapted over time to ensure they reached the most vulnerable and underserved populations.

Although men typically face higher vulnerability to the health impacts of COVID-19 [30,31], the testing program saw a higher participation rate among women across all adult age groups except those aged 85 years or older. Additionally, men showed a slightly lower likelihood of retesting and a higher tendency to discontinue testing altogether compared to women. These results are consistent with research from prior infectious disease outbreaks indicating that women tend to be more cautious than men in the context of an epidemic [32,33] and more likely to adopt preventive behaviors and adhere to public health guidelines [34-36]. Additionally, studies have shown that women reported higher levels of fear regarding the coronavirus than men during the periods that HDT or HYT was active [7,37]. Gender differences in risk perception may stem from deeply ingrained gender roles or disparities in trust toward authority figures and institutions [2,38].

COVID-19 testing adherence varies across different age groups. Younger individuals (20-34 years) are less likely to retest and more likely to stop testing, while those aged 35-64 years have a higher probability of discontinuing testing. However, older age groups (65-74 years and 75-84 years) are less likely to stop testing compared to those aged 0-19 years. This pattern may be influenced by the free weekly testing provided to students aged 0-19 years by HDT or HYT in collaboration with multiple Yolo County school districts, and the mandatory testing for Medicare or Medicaid-funded congregate living facility residents supported by state-level testing programs.

The adherence to and cessation of testing also exhibit variations among various ethnic groups. Communities such as Hispanic or Latino and American Indian or Alaska Native tend to test less frequently compared to the White demographic. On the other hand, the Native Hawaiian or Other Pacific Islander group tends to test more frequently. There are also disparities in stopping testing, with Hispanic or Latino and Black or African American individuals more likely to cease testing compared to White individuals, while Asian individuals are less likely to do so. American Indian or Alaska Native and Multiracial groups show no significant difference. These variations could be due to differences in testing behaviors, access to testing facilities, or other community-specific factors. It is noteworthy that HDT or HYT has been conducting targeted outreach to heavily impacted communities and using data analysis to strategically allocate resources where increased testing is needed.

Monitoring the spread of a disease and accurately identifying its burden through a voluntary surveillance program relies heavily on maintaining consistent user cooperation and continuing recruitment of new users. However, disaster fatigue is a natural response during extended public health crises such
as the COVID-19 pandemic that can reduce participation over time and stymy new recruitment efforts [15]. Nonetheless, HDT or HYT demonstrated the program’s effectiveness in retaining participants across a broad spectrum of age and race or ethnic demographics through and engaging new individuals in testing throughout its duration. The highest burden of infection arrived with the dominance of the Omicron variant in California, during which the testing program accelerated testing and outreach efforts to achieve the highest daily testing rate of the surveillance program. An overall increase in positive cases from Omicron is consistent with the drastic increase in virulence and avoidance of both immunity and contemporary prophylaxis [39-41]. Individuals who were already experienced with the system may have perceived increasing risk with the new variant, and opted for more testing; additionally, many workplaces and schools were requesting negative COVID-19 tests or proof of vaccination to return to the facility for a return to a sense of normalcy.

There was a significant decline in testing participation following the Omicron period. Motivation to engage in preventive behaviors against COVID-19 may have decreased because people became fatigued or burned out due to excessive and repeated exposure to similar messages about COVID-19 over time [15,16,21]. The increased availability, credibility, and convenience of other surveillance methods such as at-home COVID-19 tests [14], coupled with the impact of vaccines on risk perception [21], may have reduced the desire to continue with centralized surveillance testing. Lack of trust in government, misinformation, resistance, and conspiracy theories continue to pose a challenge for health authorities in maintaining or revitalizing public support for community-level surveillance of COVID-19, though it is unclear how this has affected participation in HDT or HYT.

In long-term crises such as the COVID-19 pandemic, public health authorities may constantly face new challenges in maintaining preventive strategies due to changes in the virus and social behavior. The comprehensive data collected during COVID-19 testing campaigns offers an opportunity to gain insight into people’s adherence and practices toward public health measures, which can inform the design of future strategies to educate communities about the benefits of engaging in preventive practices during new emerging infectious disease outbreaks.

The HDT or HYT program provides a unique setting to analyze people’s behavior since noncompliance with prevention measures may be due to reasons other than limitations in access or availability of tests. However, the nature of this data does not allow us to conclude the reasons for the observed differences. Nonetheless, the patterns observed in this analysis are consistent with findings reported in previous studies using different data sources, such as survey information.

The pandemic has highlighted the significance of social dynamics in disease control. Mathematical models played a crucial role in monitoring the behavior of SARS-CoV-2. However, access to information that accounts for social behavior and people’s adherence to health policies was extremely limited as was access to granular population-based incidence data. Analysis of acceptance and response is a starting point not only to improve future public health interventions but also to generate information that may be useful for the designing of statistical and mathematical models to study the dynamics of infectious diseases.

This analysis examines what factors affect acceptance to testing, using data from large-scale community PCR testing. While this type of data provides valuable insights, it falls short in capturing individual-level behavioral characteristics that may have affected the uptake of testing or vaccination resources. Collecting such data in future studies is crucial to optimizing the effectiveness of population-based interventions like testing, as it can help us identify the specific factors that influence people’s decisions to participate and inform the design of targeted interventions to improve uptake.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Demographic participation in testing.

References


Now Is the Time to Strengthen Government-Academic Data Infrastructures to Jump-Start Future Public Health Crisis Response

Jian-Sin Lee, MSc; Allison R B Tyler, PhD; Tiffany Christine Veinot, MLS, PhD; Elizabeth Yakel, PhD

School of Information, University of Michigan, Ann Arbor, MI, United States
UK Data Archive, University of Essex, Colchester, United Kingdom
Department of Health Behavior and Health Education, School of Public Health, University of Michigan, Ann Arbor, MI, United States
Department of Learning Health Sciences, School of Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Jian-Sin Lee, MSc
School of Information
University of Michigan
105 S State St
Ann Arbor, MI, 48109-1285
United States
Phone: 1 734 389 9552
Email: jianslee@umich.edu

Abstract
During public health crises, the significance of rapid data sharing cannot be overstated. In attempts to accelerate COVID-19 pandemic responses, discussions within society and scholarly research have focused on data sharing among health care providers, across government departments at different levels, and on an international scale. A lesser-addressed yet equally important approach to sharing data during the COVID-19 pandemic and other crises involves cross-sector collaboration between government entities and academic researchers. Specifically, this refers to dedicated projects in which a government entity shares public health data with an academic research team for data analysis to receive data insights to inform policy. In this viewpoint, we identify and outline documented data sharing challenges in the context of COVID-19 and other public health crises, as well as broader crisis scenarios encompassing natural disasters and humanitarian emergencies. We then argue that government-academic data collaborations have the potential to alleviate these challenges, which should place them at the forefront of future research attention. In particular, for researchers, data collaborations with government entities should be considered part of the social infrastructure that bolsters their research efforts toward public health crisis response. Looking ahead, we propose a shift from ad hoc, intermittent collaborations to cultivating robust and enduring partnerships. Thus, we need to move beyond viewing government-academic data interactions as 1-time sharing events. Additionally, given the scarcity of scholarly exploration in this domain, we advocate for further investigation into the real-world practices and experiences related to sharing data from government sources with researchers during public health crises.

(JMIR Public Health Surveill 2024;10:e51880) doi:10.2196/51880

KEYWORDS
COVID-19; crisis response; cross-sector collaboration; data infrastructures; data science; data sharing; pandemic; public health

Introduction
Although the world appears to be recovering from the intense impact of the COVID-19 pandemic, it is imperative to recognize that the far-reaching effects of this disease continue to endure across the globe. Its significant ramifications, such as economic fallout, disruptions in education, mental health challenges, and racial and socioeconomic health inequities, have left indelible marks on humanity, showcasing the vulnerabilities of human society when confronted with a global public health crisis.

Looking back, we should now ask: “What actions could have been done better to save more lives and reduce the aforementioned negative effects?” Moving forward, it is crucial to reflect on the valuable lessons that appear when we pose this question. Such reflection may prepare us better for future public health crises.

Among any effort to effectively manage the spread of the COVID-19 pandemic, the rapid sharing of data stands out as a key strategy. However, instances of failure have been exposed, signaling opportunities for improvement in this critical facet of
the pandemic response. Just 1 year into the COVID-19 outbreak, the British Medical Association’s chair of council voiced criticism, suggesting that “our devastating mortality figures could in part be a result of the failure of the government to properly and openly share data, communicate accurately, and act swiftly” [1]. He stressed the problematic absence of transparency concerning the actual availability of personal protective equipment supplies and the formulation of decisions on restrictions and tiers (ie, a classification system indicating local COVID-19 alert levels adopted in the United Kingdom) without also revealing the underlying statistics that supported them [1]. The US Centers for Disease Control and Prevention (CDC) itself faced criticism for its perceived “slow and siloed approach to sharing data,” which resulted in overly optimistic evaluations of the evolving vaccine effectiveness against the delta variant and contributed to the country’s falling behind in addressing that new viral mutation [2].

Government entities, a primary source of public health data, bear the responsibility for ensuring timely availability and facilitating the subsequent use of such data. To expedite the pandemic response efforts, social debates and academic studies have centered on enhancing data sharing among health care providers [3], across government departments at different levels [4], and on an international scale [5]. However, a data sharing approach that receives less attention but carries equal significance during public health crises is cross-sector data sharing collaborations between government entities and academic researchers. Specifically, this involves dedicated projects wherein a government entity shares the public health data it aggregates with an academic research team for further data analysis. The resulting data insights are then leveraged to inform policy making.

In this viewpoint, we outline the common challenges in data sharing during COVID-19, as well as other public health and general crises such as natural disasters and humanitarian emergencies. We then argue that government-academic data collaboration has the potential to alleviate these challenges, making it a topic worthy of deeper scholarly exploration. We aim to initiate a constructive discussion on effective strategies to foster this kind of cross-sector collaboration, thus paving the way for more robust and resilient responses to future public health crises.

The Significance of Data Sharing During Crises Cannot Be Overstated

In times of crisis, the saying “speed is everything” resonates more profoundly than ever. The landscape of a crisis is defined by difficulties related to uncertainty, urgency, and the relentless pursuit of solutions. Among these difficulties, data emerge as a beacon of understanding—a key resource that can illuminate facts, accelerate actions, and prevent unnecessary public panic. The significance of data sharing cannot be overstated during such moments when rapid access to comprehensive, accurate, and timely data becomes the linchpin of effective crisis management. The lessons learned from previous natural disasters and humanitarian crises have reinforced the importance of rapid data sharing [6-8]. This demand originated from not only frontline responders, decision makers, and those directly impacted but also those capable of and willing to contribute to crisis resolution, such as academic researchers. Similarly, even before the arrival of the unprecedented COVID-19 pandemic, a collection of public health crises highlighted the critical role of sharing data to safeguard global well-being. The outcomes of not sharing data or failing to integrate and coordinate data access during public health emergencies were studied, as seen during the 2003 outbreak of SARS [9,10]. In the past, data sharing was limited by technological constraints or underdeveloped legal frameworks. However, the necessity of data exchange was evident. Today, with advancements in data science knowledge and techniques, data sharing holds the potential for even greater advances that may improve the responses to public health crises.

Several characteristics collectively set COVID-19 apart from preceding crisis events: its expansive geographical reach, rapid transmission, prolonged duration, and far-reaching socioeconomic consequences. Moreover, the exceptional need for rapid data dissemination across jurisdictional and organizational boundaries during this crisis also added to its distinctiveness [11]. As stated by epidemiologist Dr Maria van Kerkhove, the COVID-19 technical lead at the World Health Organization (WHO), “When there’s so little information on a novel pathogen, any information that you can get your hands on is absolutely critical” [12]. During the COVID-19 pandemic, data sharing has played a crucial role in enhancing decision makers’ contextual understanding of the pandemic, which was deemed a key aspect of “public health situational awareness” by the US government during the pandemic [13]. Ideally, such heightened awareness would have facilitated more rapid responses, thus enabling the implementation of appropriate measures, such as resource allocation and public health interventions. When successful, data sharing also enables the examination of health disparities by providing access to demographic data, which was particularly significant in COVID-19 due to the disproportionate impact on marginalized socioeconomic, racial, and ethnic groups worldwide [14,15].

During the early stages of COVID-19, we did witness efforts to share data, such as information about potential treatments and disease spread. However, in many instances, the underlying data were of varying or low quality, as evidenced by the pandemic-related preprints that emerged [16]. In addition, considering the unpredictability of the COVID-19 pandemic, rapid data exchange necessitated effective coordination particularly among various stakeholders engaged in the “learning health system” cycle [17], which involved assembling, analyzing, and transforming data into knowledge and then performance, all within a limited time frame. Different actors’ data needs also varied across pandemic phases, such as whether a situation occurred suddenly or insidiously or whether consequences rapidly abated or lingered. Given the complexity of crisis data–sharing relationships, in the next section, we examine documented challenges in data sharing during crises that provide us with a better understanding of this issue.
Data Sharing Challenges During Public Health Crises

Public health crises have the potential for global ramifications; however, addressing them necessitates a localized approach [18]. During the COVID-19 outbreak, there existed attempts to share key data points (eg, viral genomes and methods of transmission) at the international, national, and local levels [19]. However, a difficulty lay in effectively acquiring public health insights due to the complexity of integrating disparate data across geopolitical boundaries and jurisdictional levels.

In the United States, the COVID-19 response was impaired due to the public health data infrastructures' inability to effectively share data across and within jurisdictions. As widely reported, the efficiency and timeliness of data sharing, supported by these data infrastructures, have not always been satisfactory. The desired data are “scattered” across unconnected or proprietary databases, exist in incompatible formats, or are of dubious quality and provenance [20]. Present methods of collecting public health data primarily depend on manual processes for reporting instances of certain communicable diseases and outbreaks of new diseases. Using the data accessible in electronic health record systems is not often done, even when possible. This disconnect impedes the effective use of available data. Unsurprisingly, secondary use of clinical data for public health purposes is usually insufficient [21]. Furthermore, many researchers were and still are unable to swiftly integrate with existing public health data infrastructures and “find the right antidote” for research demands when facing unforeseen public health crises [22].

The long-standing underinvestment in the maturity and agility of US public health data infrastructures has been frequently emphasized at the federal as well as state and local levels [23]. This gap became particularly problematic during the COVID-19 pandemic. From a broader perspective, neither early legislative efforts, such as the HITECH Act of 2009, nor more recent programs dedicated to COVID-19, such as the CDC’s Data Modernization Initiative, have successfully addressed the country’s fragilities in constructing critical infrastructures to meet the ongoing public health surveillance needs [21]. Admittedly, modern data-driven technologies have been widely implemented in both research and clinical contexts. Nevertheless, the aforementioned deficiencies hindered not only comprehensive real-time data analyses at the technical level [24] but also large-scale coordination between data holders and requesters at the social and organizational levels [25]. For example, in response to COVID-19, a plethora of data aggregation initiatives emerged, involving key actors, such as academic medical center networks as data holders and public health departments as data consumers. Subsequently, data requests from these federal- or state-level public health departments turned out to impose a substantial burden on academic medical centers’ reporting mechanisms [25].

While measures taken in response to the COVID-19 pandemic brought some data sharing challenges into sharp relief, the challenges arising at that time have a long history. A succession of diverse crises exposed the challenges that responders encountered when aiming to achieve rapid data sharing. By mapping the nature of different types of crises alongside their documented challenges, we can enhance our understanding of the most important data sharing approaches to study and maintain during such crises. In Table 1 [25-97], based on the extant research literature, we summarize 5 categories and 20 subcategories of common data sharing challenges during COVID-19, other public health crises, as well as other natural disasters and humanitarian crises.

After conducting a comprehensive search of scholarly literature in various fields such as public health, biomedicine, crisis informatics, and broader information science, we developed a corpus of papers relevant to data sharing challenges. Starting from this corpus, the aforementioned categories and subcategories were developed through an iterative qualitative analysis process to describe, conceptually order, and classify the bodies of text [98] as follows. First, we discerned challenges documented in the literature. Identifiable patterns then emerged, leading to the inductive grouping of specific challenges into common broad categories. These categories were further refined into their respective subcategories, taking into account the nuances of specific challenges and whether they were experienced during the COVID-19 pandemic, other public health crises, or additional types of crises (ie, natural disasters and humanitarian crises). Given the vast volume of publications discussing data sharing challenges, the categories and subcategories we developed are not meant to be exhaustive. Rather, they function as a starting framework for an interpretive critique of existing research gaps.

In all, the 5 categories of data sharing challenges we have identified are (1) data availability and quality, (2) data management and sharing, (3) information systems and data interoperability, (4) resource limitations, and (5) multiparty collaboration and coordination. In Table 1, we use omnibus terms, such as “stakeholders” and “data sharing entities,” to encompass a diverse range of individuals (eg, researchers, clinicians, and first responders) and organizations (eg, government agencies, health care providers, and humanitarian groups) actively involved in data exchange during crises. In the next section, we describe government-academic data collaborations—an often-underexplored type of data interaction during public health crises—and how to mitigate the challenges outlined below in these collaborations.
Table 1. Common data sharing challenges among stakeholders and example references by type of crises.

<table>
<thead>
<tr>
<th>Categories and subcategories</th>
<th>Definitions</th>
<th>Example references</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>COVID-19</td>
</tr>
<tr>
<td><strong>Category 1: data availability and quality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data collection for dynamic needs</td>
<td>Data needs can vary across different crisis phases and communities. Collecting data within a limited time frame and managing duplicated data requests from multiple parties could overly burden a data sharing entity. Solely gathering minimal data may also render them unsuitable for comprehensive use.</td>
<td>[25-28]</td>
</tr>
<tr>
<td>Data location uncertainties and awareness gaps</td>
<td>Stakeholders are generally unaware of the availability or whereabouts of potentially useful data and associated technologies.</td>
<td>[34-36]</td>
</tr>
<tr>
<td>Information overload</td>
<td>An overwhelming amount of information or data is accessible, yet organizations in need lack the sufficient capacity to effectively absorb and identify the most pertinent data points that they require.</td>
<td>[12,41-43]</td>
</tr>
<tr>
<td>Questionable data quality</td>
<td>Due to the time constraints for data collection, data quality may suffer, leading to doubts from data users or the general public regarding the credibility of data sharing entities. These doubts, in turn, raise concerns about the potential for misguided decisions based on the data.</td>
<td>[12,43,48,49]</td>
</tr>
<tr>
<td><strong>Category 2: data management and sharing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor or inconsistent data management practices</td>
<td>Stakeholders exhibit inadequate or divergent data management practices (eg, little documentation of provenance, version control, data dictionaries, or workflows and a lack of standardized procedures). These issues are further intensified by the urgency of crisis response efforts. Ensuring the long-term management of shared data is another notable challenge.</td>
<td>[26,34,54]</td>
</tr>
<tr>
<td>Insufficient data sharing frameworks</td>
<td>Insufficient data sharing policies or agreements are in place either within or between organizations to facilitate both the exchange of data and the willingness of stakeholders to participate in such activities.</td>
<td>[61-63]</td>
</tr>
<tr>
<td>Perceived risks for sharing data</td>
<td>Data holders may recognize the drawbacks that data sharing can pose to their self-interest (eg, the economic losses or governance issues that a nation may experience after reporting pandemic information internationally) and become hesitant to engage in such activities.</td>
<td>[41,67,68]</td>
</tr>
<tr>
<td>Tensions between openness and privacy</td>
<td>Concerns arise among stakeholders and the general public regarding the imperative to uphold data subject privacy and strike a balance between privacy protection and the timely sharing of data.</td>
<td>[71-73]</td>
</tr>
<tr>
<td><strong>Category 3: information systems and data interoperability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragmented data landscape</td>
<td>The data of interest are dispersed across numerous unconnected or proprietary databases, often in incompatible formats.</td>
<td>[24,34,71]</td>
</tr>
<tr>
<td>Antiquated or aging information systems and unsuccessful adoption of technologies</td>
<td>The existing IT systems holding the data may be outdated, resulting in manual operations for data sharing and related activities. In certain cases, under-resourced crisis sites lack well-implemented technological solutions to enable timely data sharing.</td>
<td>[72,78,79]</td>
</tr>
<tr>
<td>Data exchange and processing at scale</td>
<td>Data pipelines and infrastructures are largely absent or unprepared to exchange, manage, and analyze the growing volumes of data for crisis response.</td>
<td>[24,27,48]</td>
</tr>
<tr>
<td>Data standardization barriers</td>
<td>The absence of interoperable transactional- and data-level standards hinders the use of existing data tools and systems between organizations.</td>
<td>[19,26,84,85]</td>
</tr>
</tbody>
</table>
Category 4: resource limitations

<table>
<thead>
<tr>
<th>Definition</th>
<th>Example references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deficiency of data-related workforce</td>
<td>[5,62,80], [30,44,87], [52,88-90]</td>
</tr>
<tr>
<td>Absence of incentive structures</td>
<td>[19,36,64], [9,38,86,91], [6,88,92]</td>
</tr>
<tr>
<td>Inadequate legal support</td>
<td>[5,61,79], [51,74,93], [82,89,92]</td>
</tr>
<tr>
<td>Difficulties in reallocating resources for crisis response</td>
<td>[35,71,78,94], [10,31,55,87], [32,40,58]</td>
</tr>
</tbody>
</table>

Category 5: multiparty collaboration and coordination

<table>
<thead>
<tr>
<th>Definition</th>
<th>Example references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extensive geographic coverage and intricate geopolitics</td>
<td>[62,63,84], [29,30,93], [70,90]</td>
</tr>
<tr>
<td>Missing coordination mechanisms</td>
<td>[25,35,61,85], [10,29,54], [7,95,96]</td>
</tr>
<tr>
<td>Misaligned goals or competitive priorities</td>
<td>[27,62], [29,39,93], [45,82,92]</td>
</tr>
<tr>
<td>Organizational politics, bureaucracy, and power dynamics</td>
<td>[67,84,94], [37,69,97], [8,75,95]</td>
</tr>
</tbody>
</table>

Navigating Challenges: Government-Academic Collaboration as Part of the Social Infrastructure for Public Health Crisis Response

Data collected, aggregated, or provided by government entities have been observed to play a key part in managing the COVID-19 crisis. As an example that involves government data for internal use, in the United States, the city government of Boston used its preexisting data warehouse, aggregating data from 31 departments, to rapidly develop a public dashboard at the outset of the pandemic [99].

Simultaneously, there were instances globally where academia has engaged in using government data to understand the pandemic, with various forms of such cross-sector initiatives. For example, the Israeli government orchestrated a “datathon” competition—an event uniting participants from diverse sectors, including academic scientists, to devise practical, data-driven models and insights [100]. Another approach to effectively leveraging government data is by directly making them available to citizens, including academic researchers. The term “open government data” (OGD) refers to government-held data made accessible to the public to enhance transparency regarding government operations [101]. One common method for obtaining OGD is through government agencies’ open data portals [102]. This method has enabled various stakeholders to engage in data analysis before and during the COVID-19 period [103,104].

Nonetheless, there are limitations to the aforementioned data interactions between government entities and academia. Specifically, short-term data analysis competitions may not adequately support the relatively medium to long-term needs for pandemic response policy planning by government entities. Participants from various sectors are unlikely to sustainably remain within the government’s data collaboration network after the competition ends. On the other hand, in relation to OGD, there is a critical aspect of their availability—when accessing OGD, users download data independently without direct interactions with government agencies as data providers. Such a relatively unilateral data access approach may give rise to issues related to data quality and usability (eg, data integrity, granularity, and timeliness [105]), potentially impeding users’ understanding and appropriate interpretation of OGD and in turn undermining the data’s overall effectiveness to be used accurately, appropriately, and efficiently.

Consequently, there is a growing demand for enhanced “direct” collaboration between government entities and researchers on
Despite this limited information, these additional cases make applications to track the pandemic within the state [114]. Of data-driven tools (eg, symptom and vaccination monitoring), the Michigan program, attempting to connect essential workers with available health and Human Services Agency and the University of Michigan, partnered with the Michigan Department of Health and Human Services Agency and the University of Michigan, to develop a series of data-driven tools (eg, symptom and vaccination monitoring applications) to track the pandemic within the state [114]. Despite this limited information, these additional cases make it clear that such collaborations exist more broadly, and that they potentially hold value in responding to public health crises.

Our call for research into government-academic collaborations is vital given the history of previous efforts. For example, between 2005 and 2011, the CDC launched the Centers of Excellence in Public Health Informatics program. This initiative shared similarities with the form of government-academic collaboration we are advocating as it aimed to bridge the gap between public health research and practice through collaborations among academia, local or state public health departments, and other health informatics professionals [115]. The program financially supported academic institutions, such as Harvard University and Indiana University [116], to establish research centers that would translate research outcomes into public health practice. Data sharing and information exchange were integral components as well, though primarily from clinical sources to public health information systems [117]. However, after the funding concluded, the infrastructures of the research centers necessitated institutional or external backing for sustainability. This indicates the difficulty in maintaining the long-term viability of such short- or medium-term programs.

Up to this point, we have outlined several forms of government-academic data interactions and their limitations in terms of effectiveness, efficiency, and sustainability. In the following subsection, we elaborate on an alternative collaborative form that may function better during public health crises than the other aforementioned government-academic data interactions.

**Government-Academic Collaborations as a Promising Solution to Data Sharing Challenges**

**Overview**

A public health data infrastructure can be defined as “an ecosystem composed of the people, processes, procedures, tools, facilities, and technologies, which supports the capture, storage, management, exchange, and creation of data and information to support individual patient care and population health” [118]. Nonetheless, it is notable that existing discussions around these kinds of infrastructures mostly center on technological aspects, such as health information systems and their data standards for interoperability. At the same time, other key components involving people, processes, and norms have received less attention [119]. In fact, for researchers, we contend that data collaborations with government entities should be considered part of the social infrastructure that supports their research efforts toward public health crisis response. Specifically, as we argue below, many of the data sharing challenges displayed in Table 1 may be effectively mitigated through well-planned government-academic collaborations. In the upcoming sections, we explain how the 5 types of challenges (data availability and quality, data management and sharing, information systems and data interoperability, resource limitation, and multiparty collaboration and coordination) can be navigated by developing partnerships between government entities and academia, particularly during public health emergencies.
Data Availability and Quality

Government-academic data collaborations are goal oriented and usually provide researchers with prepared data sources, alleviating concerns about data location and the identification of relevant information from the vast pool of available data “outside,” which are often collected by researchers themselves. In the exemplar case described above, the IDM researchers highlighted that during their collaboration with the Washington State DOH, the COVID-19 data shared by the government was of high quality, which significantly expedited their work, enabling them to deliver outputs swiftly [111]. Furthermore, this collaborative model often implies prioritized communication channels, facilitating prompt resolution of data quality issues by both parties.

Data Management and Sharing

Intentionally built collaborations often smooth data exchange activities by facilitating the creation of a preestablished data sharing framework, which comprises agreements on data management and use, as well as clearly defined obligations and codes of conduct for both parties involved [120]. For example, in response to the external demands for Zika-related research projects, Brazil’s Secretariat of Health took a proactive approach by initiating collaboration protocols with researchers [74]. These data-related agreements clearly regulated the conditions for accessing data and thus laid the foundations for project execution at maximum speed [74]. Besides, such data sharing frameworks function as a common space where collaborators can fine-tune their collective data activities based on project performance.

Information Systems and Data Interoperability

As government entities and researchers work together around data resources, potential issues, such as inadequate technical infrastructures and inconsistent data standards, may come to light during the early stages [121]. Nonetheless, such a collaboration also presents an opportunity for both parties to acknowledge these problems and proactively work toward resolving them. To illustrate, although not identical to the collaborative model we advocate, the well-known National COVID Cohort Collaborative initiative aimed to overcome interoperability barriers. Ultimately, it managed to build a scalable data analytics infrastructure by uniting US federal agencies, health care providers, and research leaders to harmonize pandemic data across different organizations [122]. Notably, there remains much to investigate regarding the implementation of data standardization within collaborative efforts on a smaller scale, specifically between government entities and researchers.

Resource Limitations

Government-academic collaborations make dedicated investments, including workforce and funding, in their data projects. This commitment enables the efficient integration of complementary resources from both sectors, facilitating a synergistic approach to data-driven initiatives. In particular, public health and biomedical informatics experts recently stressed the need to build “a public health workforce that is skilled in informatics and data science...to meet 21st century health threats” [21]. Nonetheless, they simultaneously pointed out the challenges in recruiting incoming talent as well as in training this workforce, an ongoing problem that state and local public health departments have historically faced [21]. This further highlights the value of government-academic collaborations, in which public health authorities can borrow well-established expertise from academics in “rapid response data science” [107].

Multiparty Collaboration and Coordination

Previous ongoing collaborative relationships help to build trust in advance, eliminating the need for a cumbersome initiation period characterized by misaligned organizational priorities and conflicting power dynamics that impede the rapid circulation of data assets. In the case of the Washington State DOH and IDM described above, the IDM researchers achieved favorable outcomes in their collaboration with DOH employees by recognizing the significance of building trust [111]. Based on their accounts, the crucial element that contributed to the success of the collaboration was the ability to align needs and tasks at an early stage of the partnership [111]. In addition, such early coordination efforts may also help data holders preempt the challenges posed by the previously discussed scenario of receiving a sudden surge in requests from data consumers [25].

To summarize, we contend that placing sole emphasis on the exchange of data through technological infrastructures falls short when confronted with the challenges of a public health crisis. Government-academic data collaborations, as essential social infrastructures, encompass not only people, processes, and norms but also rely on trusting relationships within the larger legal and political context. These elements are all integral and indispensable components for the success of the data sharing enterprise. Ultimately, sharing data is not just a technical process—it should be a collaborative endeavor that transcends boundaries. Thus, we should study and implement such collaborations now, before the next public health crisis is on us. Doing so may help to establish greater readiness and more rapid responses in the future. We now outline recommendations that, if implemented, may assist in developing this crucial sociotechnical infrastructure.

Conclusion and Recommendations

In this viewpoint paper, we investigate the challenges associated with sharing data in public health crises, many of which stem from the long-standing inadequacy in the US public health data infrastructures. In particular, we have witnessed repeated appeals for increased data sharing endeavors spanning various sectors and extending in multiple directions, such as data scientists in the health care industry stressing that “sharing data should not just be a one-way street from the clinician to the researcher” [123]. However, the factors for successful collaborative data sharing across sectors in public health crises—in which government entities share data with academic researchers for effective use—need further attention. Therefore, resulting from a synthesis of extant research and our arguments, we call for more effort to be invested in building data sharing infrastructures capable of bridging and leveraging the respective strengths of government entities and academic researchers. Such infrastructures need to be established within an ecosystem that
incorporates not only technologies but also policies, processes, and personnel. This holistic framework is ideally designed to facilitate researchers in seamlessly accessing and employing data aggregated and managed by government entities for their mutual benefit.

The COVID-19 pandemic has taught us a valuable lesson, which surpasses those gained from any previous public health emergencies: that the aforementioned infrastructure for rapid and effective data sharing should be established well in advance of a crisis. Particularly, we argue that government-academic data interactions should not be thought about as only 1-time data sharing. Instead, we recommend that emphasis should be placed on the construction of robust and enduring collaborative infrastructure that not only outlasts a specific public health crisis but also is in place to respond to the next one. Ideally, these data collaborations should not be confined to emergencies or a small number of high-priority threats [51]. After all, data sharing practices both during and between crises affect crisis response efforts, albeit potentially in distinct manners. To be specific, routine data sharing practices in scientific research and the availability of preexisting baseline data before a crisis can play a crucial role in facilitating prompt planning for health relief activities [32,55]. In addition, even before implementing crisis response measures, persistent data partnerships may hold the potential to enhance the detection and early characterization of issues arising during a crisis, facilitated by the accelerated exchange of information between government entities and academics.

As of May 2023, the WHO and the US government declassified COVID-19 as a public health emergency. While most individuals have moved on, for those who have compromised immune systems or are otherwise at greater risk for negative outcomes from the virus, exchanging data to facilitate accurate disease-level reporting remains crucial for evaluating their safety. However, the termination of certain data sharing mandates and data-collection initiatives could hinder government bodies and research institutions from maintaining uninterrupted access to vital disease-related metrics [124]. With the resurgence of COVID-19 hospital admissions since July 2023 in the United States [125] and the possibility of “a new norm of summer surges” [126], it is worth considering whether we want to revert to a just-in-time approach to data sharing practices or if we should be proactive and build just-in-case resilient, long-term data infrastructures for forthcoming public health scenarios. We strongly assert that our choice should be the latter.

**Future Research Agenda**

As mentioned, it is critical to begin now to establish more effective government-academic collaborative infrastructures for public health crisis response. To do so, we must develop more systematic research on the facilitating and impeding factors for such data collaborations. In this viewpoint paper, we reviewed existing research literature and summarized data sharing challenges during different crisis scenarios (Table 1). Significantly, we conclude from the literature review that there is a conspicuous scarcity of scholarship addressing the practices and experiences related to disseminating data from government sources to researchers throughout extended or ongoing crisis response situations, including instances of global health crises. In particular, in terms of data-exchange partnerships during public health crises, the public health and biomedical informatics literature often enumerates a wide range of stakeholders [21,127] but generally lacks a specialized focus on the connections between government entities and researchers. On the other hand, literature within the realm of crisis informatics more often addresses the circulation of information and data among the public, frontline responders, and governmental bodies in natural disaster scenarios (eg, earthquakes [88], hurricanes [128], and wildfires [129]).

While government-academic collaborations that allow data exchange did exist at different administrative levels during COVID-19, there is a notable dearth of research studying these relationships. To initiate further discussions, we draw on the data sharing challenges outlined earlier and propose 3 key research questions, to foster more substantive dialogues and shape the future research agenda: (1) What types of government-academic collaborative infrastructures should we be developing? How can these infrastructures be best sustained? (2) Considering the unique characteristics of public health crises, what are the best practices for implementing data sharing and data collaborations? and (3) From the respective views of government entities and researchers, what are the incentives and disincentives that influence their willingness and capacity to engage in developing and sustaining collaborative data infrastructures?

In conclusion, the COVID-19 pandemic has emphasized the imperative for robust and durable government-academic partnerships in public health crises. As we transition beyond the pandemic, it is crucial to develop systematic research on the factors influencing these collaborations. Before the next public health crisis arises, we invite decision makers, researchers, and practitioners across government entities, academia, and various sectors to leverage the collective knowledge and expertise of diverse stakeholders, strengthening existing and building new government-academic data collaborative infrastructures. The time to act is now, and the path to a more resilient future begins with our commitment to addressing these critical challenges.

**Acknowledgments**

The authors would like to thank the Academic Data Science Alliance and Dr Jing Liu at the Michigan Institute for Data Science for their inspiration in the development of this manuscript. TCV’s effort on this manuscript was funded by an unrestricted gift from Google Health as part of its COVID-19 relief initiatives.
Authors' Contributions

JSL led the content development of the manuscript, prepared the initial draft, and coordinated contributions and subsequent revisions. ARBT took the lead in drafting an earlier version of the manuscript and contributed to subsequent revisions of the current version. TCV and EY provided critical revisions to this manuscript, enhancing its intellectual content. All authors contributed content and feedback that informed the drafting process and agreed to be responsible for all aspects of the work, ensuring its integrity and accuracy.

Conflicts of Interest

None declared.

References

1. Mahase E. COVID-19: government's failure to share data and face scrutiny have undermined response, say MPs. BMJ 2021;372:n717 [FREE Full text] [doi: 10.1136/bmj.n717] [Medline: 33722799]


42. Ma F. How can information and data management be used to address global health crisis. Data Inf Manag 2020;4(3):127-129 [FREE Full text] [doi: 10.2478/dim-2020-0018] [Medline: 35382100]


117. Dixon BE, Grannis SJ. Why "what data are necessary for this project?" and other basic questions are important to address in public health informatics practice and research. Online J Public Health Inform 2011;3(3):ojphi.v3i3.3792 [FREE Full text] [doi: 10.5210/ojphi.v3i3.3792] [Medline: 23569618]


Abbreviations

CDC: Centers for Disease Control and Prevention
DOH: department of health
IDM: Institute for Disease Modeling
OGD: open government data
WHO: World Health Organization
Abstract

Background: Beyond the direct effect of COVID-19 infection on young people, the wider impact of the pandemic on other infectious diseases remains unknown.

Objective: This study aims to assess changes in the incidence and mortality of 42 notifiable infectious diseases during the pandemic among children and adolescents in China, compared with prepandemic levels.

Methods: The Notifiable Infectious Disease Surveillance System of China was used to detect new cases and fatalities among individuals aged 5-22 years across 42 notifiable infectious diseases spanning from 2018 to 2021. These infectious diseases were categorized into 5 groups: respiratory, gastrointestinal and enterovirus, sexually transmitted and blood-borne, zoonotic, and vector-borne diseases. Each year (2018-2021) was segmented into 4 phases: phase 1 (January 1-22), phase 2 (January 23-April 7), phase 3 (April 8-August 31), and phase 4 (September 1-December 31) according to the varying intensities of pandemic restrictive measures in 2020. Generalized linear models were applied to assess the change in the incidence and mortality within each disease category, using 2018 and 2019 as the reference.

Results: A total of 4,898,260 incident cases and 3701 deaths were included. The overall incidence of notifiable infectious diseases decreased sharply during the first year of the COVID-19 pandemic (2020) compared with prepandemic levels (2018 and 2019), and then rebounded in 2021, particularly in South China. Across the past 4 years, the number of deaths steadily decreased. The incidence of diseases rebounded differentially by the pandemic phase. For instance, although seasonal influenza dominated respiratory diseases in 2019, it showed a substantial decline during the pandemic (percent change in phase 2 2020: 0.21, 95% CI 0.09-0.50), which persisted until 2021 (percent change in phase 4 2021: 1.02, 95% CI 0.74-1.41). The incidence of gastrointestinal
and enterovirus diseases decreased by 33.6% during 2020 but rebounded by 56.9% in 2021, mainly driven by hand, foot, and mouth disease (percent change in phase 3 2021: 1.28, 95% CI 1.17-1.41) and infectious diarrhea (percent change in phase 3 2020: 1.22, 95% CI 1.17-1.28). Sexually transmitted and blood-borne diseases were restrained during the first year of 2021 but rebounded quickly in 2021, mainly driven by syphilis (percent change in phase 3 2020: 1.31, 95% CI 1.23-1.40) and gonorrhea (percent change in phase 3 2020: 1.10, 95% CI 1.05-1.16). Zoonotic diseases were not dampened by the pandemic but continued to increase across the study period, mainly due to brucellosis (percent change in phase 2 2020: 0.94, 95% CI 0.75-1.16). Vector-borne diseases showed a continuous decline during 2020, dominated by hemorrhagic fever (percent change in phase 2 2020: 0.68, 95% CI 0.53-0.87), but rebounded in 2021.

Conclusions: The COVID-19 pandemic was associated with a marked decline in notifiable infectious diseases in Chinese children and adolescents. These effects were not sustained, with evidence of a rebound to prepandemic levels by late 2021. To effectively address the postpandemic resurgence of infectious diseases in children and adolescents, it will be essential to maintain disease surveillance and strengthen the implementation of various initiatives. These include extending immunization programs, prioritizing the management of sexually transmitted infections, continuing feasible nonpharmaceutical intervention projects, and effectively managing imported infections.

(JMIR Public Health Surveill 2024;10:e47626) doi:10.2196/47626

KEYWORDS
children and adolescents; COVID-19; notifiable infectious diseases

Introduction

Globally, the nearly 1.9 billion children and adolescents younger than 20 years, who account for approximately 26% of the world’s population, are particularly vulnerable to infectious diseases [12]. Yet, previous research has suggested that compared with adults, children and adolescents are more likely to be asymptomatic or have milder symptoms of COVID-19, experience a shorter course of illness, and have a lower risk of developing severe disease [3,4]. They are also less likely to experience multisystem inflammatory syndrome [5]. In China, there has been relatively limited epidemiological research on COVID-19 infections among children and adolescents and the bulk of studies have focused on the original virus strains from the early stages of the outbreak. For instance, the largest published Chinese population study included only 2135 children and adolescents infected with COVID-19 from January to February 2020 [6]. With the easing of COVID-19–related restrictions worldwide, there has been a sharp increase in the number of children and adolescents infected with the virus [3,4].

In the early response to the COVID-19 pandemic, countries variably introduced a series of social and behavioral measures (eg, border closures, quarantine, lockdowns, school closures, travel restrictions, and face masks) to reduce viral transmission. These approaches are relatively effective [5,7-11]. For example, short-term prospective observational research has indicated that school closures and restricted social distancing were associated with a 38% reduction in the incidence of COVID-19 [12]. Another study conducted in 11 countries demonstrated a 15% reduction in the incidence of COVID-19 associated with multiple social restrictions; additionally, it found that earlier implementation of lockdown measures was associated with a larger reduction in the incidence of COVID-19 [13]. Social restrictions aimed at limiting the transmission of COVID-19 have also had an indirect effect on the incidence of a wide range of other infectious diseases, likely reflecting the interruption of the transmission chain [14]. Respiratory diseases have been particularly affected [14-17]. Chinese surveillance indicates that during the early stages of the COVID-19 pandemic in 2020, the rates of positive tests for all respiratory viruses declined. Reductions varied from 17.2% for respiratory syncytial virus to 87.6% for influenza virus [16]. These patterns were also evident in countries with a low prevalence of COVID-19, such as Australia, which experienced a 36% reduction in antibiotic dispensing for respiratory tract infections in the first year of the pandemic [18]. In England, in the first 12 months after the onset of the COVID-19 pandemic, large and sustained reductions were found in the rates of hospital admissions for respiratory infections as well as for a spectrum of other severe and vaccine-preventable childhood infections, such as meningitis [16].

In any country, it is apparent that a balance of factors has affected the transmission and severity of COVID-19, including the timing and intensity of lockdowns and their subsequent lightening, evolving virus strains, and access to vaccines. What is less clear is how these dynamic changes have impacted other infectious diseases. For instance, in China, hand, foot, and mouth disease (HFMD) surged following the gradual lightening of COVID-19 restrictions, distinct from scarlet fever or seasonal influenza, both of which remained at lower levels [19]. Although several studies have investigated the impacts of COVID-19–related restrictions early in the pandemic, there has been less attention to children and adolescents. To date, there has been a lack of systematic exploration in children and adolescents regarding the impacts of the COVID-19 pandemic and the associated social restrictions on the spectrum of infectious diseases. In addition, little attention has been paid to the effects following the lifting of these restrictions [15-17,19]. This study analyzed data for 42 notifiable infectious diseases in children and adolescents in mainland China from 2018 to 2021 to investigate the indirect effect of COVID-19 restrictions on a wide spectrum of infectious diseases. Specifically, we aimed to determine the variability and persistence of this indirect effect across age, sex, time, region, and disease category, with the expectation that this information can inform the nature of
protective strategies for children and adolescents within subsequent pandemics.

**Methods**

**Data Collection**

Using the China Information System for Disease Control and Prevention (CISDCP), we identified data on daily new cases and deaths for 42 notifiable infectious diseases reported in children and adolescents aged 5-22 years between 2018 and 2021. For each case, we recorded the date and location of disease onset and death, diagnosed disease, age, and sex. The CISDCP covers 55,077 national health facilities in 397 cities across all 31 provinces of mainland China and maintains a detailed surveillance protocol [20]. The CISDCP had a national average coverage rate that remained stable and exceeded 95% for all health facilities at or above the county level during the study period, and in 2017, web-based reporting was available for 87.2% of national health facilities [20]. In this analysis, we only included children and adolescents with a confirmed diagnosis of any of the 42 notifiable infectious diseases. Cases from Hong Kong and Macao were excluded. Demographic information at the city level was provided by the National Bureau of Statistics of China [21].

**Ethical Considerations**

Ethical approval was not required for this study because it exclusively used data that were deidentified and publicly available, making it exempt from review by an institutional review board or ethics committee.

**Classification**

In this study, we focused on 42 notifiable infectious diseases that are captured within the CISDCP. To classify these diseases, we revised our previous approach to categorization [20], classifying vaccine-preventable infectious diseases and bacterial infections based on their respective modes of transmission. These 42 infectious diseases were grouped into 5 categories: (1) respiratory diseases, (2) gastrointestinal and enterovirus diseases, (3) sexually transmitted and blood-borne diseases, (4) zoonotic diseases, and (5) vector-borne diseases. Respiratory diseases included 10 diseases: seasonal influenza, mumps, tuberculosis, scarlet fever, rubella, pertussis, measles, meningococcal meningitis, leprosy, and diphtheria. Gastrointestinal and enterovirus diseases included 8 diseases: HFMD, infectious diarrhea, dysentery (amoebic dysentery and bacterial dysentery), acute hemorrhagic conjunctivitis, typhoid and paratyphoid, hepatitis A, cholera, and poliomyelitis. Sexually transmitted and blood-borne diseases included 6 diseases: syphilis, gonorrhea, HIV/AIDS, hepatitis B, hepatitis C, and hepatitis D. Vector-borne diseases include 9 diseases: hemorrhagic fever, dengue, Japanese encephalitis, typhus, malaria, kala-azar, schistosomiasis, filariasis, and plague. Zoonotic diseases included 9 diseases: brucellosis, hepatitis E, hydatid disease, rabies, anthrax, leptospirosis, H5N1, H7N9, and severe acute respiratory syndrome.

**Epidemic Stages and Phases**

As shown in Figure 1, we divided the 4-year study period into 3 stages of the COVID-19 pandemic based on the timing of the major restrictions that were implemented in response to the pandemic [19]: 2018 and 2019 were considered pre–COVID-19 pandemic years, 2020 was categorized as COVID-19 year 1, and 2021 as COVID-19 year 2.

Because of the different intensities of COVID-19–related restrictive measures taken to prevent and control COVID-19, we further divided 2020 into 4 phases: phase 1 (January 1-22), phase 2 (January 23-April 7), phase 3 (April 8-August 31), and phase 4 (September 1-December 31). In phase 1, COVID-19 had just broken out and no specific interventions or COVID-19–related restrictions were implemented in mainland China. In phase 2, the most intense COVID-19 restrictions were implemented, including school closures and travel restrictions, routine temperature monitoring, mask wearing and social distancing, and isolation of high-risk groups and those with COVID-19. In phase 3, marked by the lifting of the lockdown in Wuhan (April 8, 2020), all provinces of mainland China downgraded their response to the public health emergency, and there was a return to regular education, work, and public transit in cities without major COVID-19 outbreaks. During this phase, a variety of COVID-19–related restrictions remained in practice, including social distancing, mask wearing, routine temperature monitoring, school closures, and capping the number of meetings. Phase 4 began on September 1, signaling the return of businesses, recreational activities, and school reopenings across the nation. However, routine temperature monitoring and mask wearing were still widely practiced. To mitigate the influence of seasonal characteristics on infectious diseases, we applied these same 4 phases, identified in 2020, to each of the study years, which resulted in each year, from 2018 to 2021, being divided into 4 phases. During all epidemic phases in 2021, routine temperature monitoring and mask wearing were still practiced, as described in Figure 1. In China, widespread vaccination against COVID-19 began in 2021.
Figure 1. Categorization of the 3 stages (pre–COVID-19 (2018 and 2019), year 1 (2020), and year 2 (2021)) of the pandemic and 4 phases of the first year of the pandemic in China (2020) by intensity of COVID-19 restrictions.

Statistical Analysis
Incidence and mortality were used to present trends by year from 2018 to 2021. Incidence (per 100,000) was calculated by the number of incident cases divided by the number of the population aged 5-22 years. Mortality (per 100,000) was calculated by the number of deaths divided by the number of the population aged 5-22 years (per 100,000). To assess changes in the 4 epidemic phases, the percent change (PC) for “during versus pre” was calculated using the following formula: \[
\frac{\text{incidence (2020)} - \text{incidence (2018-2019)}}{\text{incidence (2018-2019)}} \times 100\%
\]
and the PC for “COVID-19 year 2 versus COVID-19 year 1” was assessed using the following formula:
\[
\frac{\text{incidence (2021)} - \text{incidence (2020)}}{\text{incidence (2020)}} \times 100\%.
\]
Heat maps were used to show the characteristics of each disease for incidence, number of cases, and number of deaths. Stacked plots were used to visualize the proportion of each infectious disease. To illustrate the trends in infectious diseases from 2018 to 2021, the spiral visualization was used to represent the daily number of incident cases.

To assess the changes in the epidemic phases, the PC in the incidence of each category was calculated at the city level as follows:
\[
\frac{\text{incidence (phase x in 2020-2021)} - \text{incidence (phase x in 2018-2019)}}{\text{incidence (phase x in 2018-2019)}} \times 100\%,
\]
where incidence (phase x in 2020-2021) indicated the average incidence in the corresponding phase in 2021 or 2020 and incidence (phase x in 2018-2019) indicated the average incidence in the specific phase during 2018-2019. The PC in specific phases was calculated by disease, disease category, year, and prefecture-level city. To eliminate the impact of season and quantify the impact of different COVID-19–related restrictions in the epidemic phases on the incidence of infectious diseases, multivariable generalized linear models were applied and used to compare changes in regional variation and specific categories of infectious disease changes by epidemic phases in pandemic years compared with the pre–COVID-19 years of the pandemic in 2018 and 2019. A quasi-Poisson model was fitted for prediction using the indicators as follows: phase category for 2020 and 2021, with number of person-days (population size times the number of days in the month) as an offset. All models were fitted for the daily number of cases. Incidence rate ratios (IRRs) associated with the phase indicators estimated reflect the effects of COVID-19 restrictions on the incidence of each notifiable disease in the pandemic years. IRRs associated with the disease categories were also estimated to assess the impact of COVID-19–related restrictions on various disease categories. All statistical analyses were performed using the R program (version 4.1.1; R Foundation).

Role of the Funding Source
The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The authors from Peking University had full access to the data in the study.

Results
Changes in the Incidence and Mortality of Notifiable Infectious Diseases, 2018-2021
Between 2018 and 2021, nearly 5 million Chinese children and adolescents between the ages of 5 and 22 years (2,818,718 males and 2,079,542 females) were diagnosed with 1 of 42 notifiable infectious diseases. A total of 3701 deaths were reported during this period (2530 in males and 1171 in females). As depicted in Figure 2, the number of reported cases of notifiable infectious diseases fluctuated throughout the study period. In 2018, 911,522 cases were reported, which rose to 2,268,809 in 2019 and declined to 813,635 in 2020, before increasing again to 904,294 in 2021. From 2018 to 2021, the overall incidence of notifiable infectious diseases was 248.848, 738.338, 266.051, and 281.664 per 100,000, respectively. During the first year of the COVID-19 pandemic in 2020, the incidence of notifiable infectious diseases dropped by 46.1% but then rebounded by 5.9% in 2021. Nevertheless, the total mortality and number of deaths decreased steadily over the 4-year period, with rates of 0.304 (n=1112 cases), 0.340 (n=1045 cases), 0.286 (n=869 cases), and 0.210 (n=675 cases) per 100,000 over each year, respectively (Multimedia Appendices 1 and 2).
Trends of Notifiable Infectious Diseases, by Category, 2018-2021

Table 1 and Figure 3 present the overall and temporal trends across the 5 categories of infectious diseases, before and during the COVID-19 pandemic. Among the 5 categories, respiratory, sexually transmitted and blood-borne, and gastrointestinal and enterovirus diseases were the most commonly reported notifiable infectious diseases in those aged 5-22 years (Multimedia Appendix 3). Although the overall incidence of notifiable infectious diseases in children and adolescents declined during the first year of the COVID-19 pandemic and then increased during COVID-19 year 2, the different categories of infectious diseases exhibited diverse temporal changes. Generally, the peak of infectious diseases was reached from December 2019 to January 2020, attributed to a seasonal influenza outbreak. The incidence of notifiable infectious diseases in the first year of the COVID-19 pandemic (2020) was higher than that in the prepandemic years, but then dropped dramatically from February 2020 to March 2021, before rising at the end of the reporting period in the last quarter of 2021 (Figure 4).

The seasonality of respiratory diseases is generally pronounced in winter in China. This was not experienced in the first year of the COVID-19 pandemic, with the fourth quarter of 2020 and the first quarter of 2021 presenting a low incidence of respiratory diseases. This appeared to rebound the following year, with much higher levels in the fourth quarter of 2021. Seasonal influenza dominated these trends; before the pandemic, the incidence of influenza in 2019 was 8 times more common than in 2018 (incidence rate of 480.541 vs 52.858, respectively), and remained elevated compared with 2018 throughout 2020 and 2021. Excluding seasonal influenza, the incidence of all other respiratory diseases continued to decline during the COVID-19 pandemic (50.768 and 43.241 per 100,000 in 2020 and 2021, respectively), compared with pre–COVID-19 levels (average 106.735 per 100,000), especially for mumps, tuberculosis, scarlet fever, and rubella. The 1 exception was pertussis. The leading causes of respiratory deaths were tuberculosis and seasonal influenza. The number of deaths decreased steadily during these 4 years (Multimedia Appendix 4).

Generally in China, gastrointestinal and enterovirus diseases peak in incidence from April to July. However, this peak was not experienced during the COVID-19 pandemic in 2020. Compared with the 2 pre–COVID-19 years, the incidence of gastrointestinal and enterovirus diseases decreased by 33.6% in 2020, but rebounded by 56.9% in 2021. HFMD and infectious diarrhea were the main contributors to this rebound of
gastrointestinal and enterovirus diseases. Infectious diarrhea surpassed HFMD as the most common notifiable infectious disease among the gastrointestinal and enterovirus diseases category in 2020.

Different trends were apparent in the sexually transmitted and blood-borne diseases category. In the initial year of the COVID-19 pandemic, the previous upward trend in incidence observed in the prepandemic years was halted (32.133 and 39.336 per 100,000 in 2018 and 2019, respectively, vs 35.562 per 100,000 in 2020). However, this was succeeded by a noticeable rebound in the second year of the pandemic (38.753 per 100,000 in 2021), representing a 9.0% increase from 2020. The incidence of hepatitis B, C, and D and HIV/AIDS remained stable or declined slightly during the COVID-19 pandemic; however, the incidence of syphilis and gonorrhea increased in both years of the pandemic. In 2018, the disease with the highest incidence within the sexually transmitted and blood-borne diseases category was hepatitis B. Hepatitis B was overtaken by syphilis from 2019 to 2021, particularly in adolescents aged 15-19 years (Figure 5). Across the 4 years, HIV/AIDS remained the major cause of death from any infectious disease in children and adolescents, although the mortality rate decreased by 26.6% over this period, from 0.267 per 100,000 in 2018 to 0.196 per 100,000 in 2021.

In China, zoonotic diseases typically show a seasonal pattern with the highest incidence in summer. Zoonotic diseases seemed less affected by COVID-19, with the incidence continuing to increase from 0.660 per 100,000 in 2018 to 0.982 per 100,000 in 2021, an increase of 18.2%. While this was mainly attributed to brucellosis, which has the highest incidence among zoonotic diseases, the previous growth trends of other zoonotic diseases such as hepatitis E and hydatid disease were also constrained. The incidence and mortality of rabies continued to decline during the study period, although it remained the third leading cause of death from any notifiable infectious disease in children and adolescents in China (followed by HIV/AIDS and tuberculosis).

There was a substantial prepandemic increase in the incidence of vector-borne disease from 2018 to 2019 (from 0.542 to 1.175 per 100,000 in 2018 and 2019, respectively), which remained low during 2020 and 2021. The observed trend was attributed to dengue, Japanese encephalitis, typhus, malaria, and kala-azar, which exhibited significant declines in 2020 (0.335 per 100,000) and remained at low levels in 2021 (0.344 per 100,000). Conversely, hemorrhagic fever seemed to experience a resurgence during the COVID-19 pandemic, increasing by 26.8%.
Table 1. Incidence (per 100,000) and mortality (per 100,000) for 42 notifiable infectious diseases among 5-22-year olds, by year.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>248,847</td>
<td>738.338</td>
<td>266.051</td>
<td>281.664</td>
<td>0.210</td>
<td>-46.1</td>
</tr>
<tr>
<td>Respiratory diseases</td>
<td>135.956</td>
<td>600.910</td>
<td>170.913</td>
<td>150.056</td>
<td>0.008</td>
<td>-53.6</td>
</tr>
<tr>
<td>(seasonal influenza</td>
<td>83.100</td>
<td>120.369</td>
<td>50.768</td>
<td>43.241</td>
<td>0.007</td>
<td>-50.1</td>
</tr>
<tr>
<td>(seasonal influenza</td>
<td>52.858</td>
<td>480.541</td>
<td>120.145</td>
<td>106.815</td>
<td>0.001</td>
<td>-55.0</td>
</tr>
<tr>
<td>included)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mumps</td>
<td>47.578</td>
<td>68.429</td>
<td>21.947</td>
<td>18.071</td>
<td>&lt;0.001</td>
<td>-62.2</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>23.769</td>
<td>29.085</td>
<td>25.979</td>
<td>19.969</td>
<td>0.007</td>
<td>-3.1</td>
</tr>
<tr>
<td>Scarlet fever</td>
<td>10.586</td>
<td>13.606</td>
<td>2.572</td>
<td>4.266</td>
<td></td>
<td>-78.7</td>
</tr>
<tr>
<td>Rubella</td>
<td>0.633</td>
<td>8.296</td>
<td>0.458</td>
<td>0.093</td>
<td></td>
<td>-89.7</td>
</tr>
<tr>
<td>Pertussis</td>
<td>0.371</td>
<td>0.697</td>
<td>0.137</td>
<td>0.804</td>
<td></td>
<td>-74.3</td>
</tr>
<tr>
<td>Measles</td>
<td>0.142</td>
<td>0.227</td>
<td>0.039</td>
<td>0.018</td>
<td></td>
<td>-78.9</td>
</tr>
<tr>
<td>Meningococcal</td>
<td>0.011</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>-50.0</td>
</tr>
<tr>
<td>meningitis</td>
<td>0.010</td>
<td>0.012</td>
<td>0.011</td>
<td>0.009</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>Leprosy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diphtheria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal and</td>
<td>79.556</td>
<td>&lt;0.001</td>
<td>96.056</td>
<td>58.341</td>
<td>0.001</td>
<td>-33.6</td>
</tr>
<tr>
<td>enterovirus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand, foot, and mouth</td>
<td>42.638</td>
<td>47.122</td>
<td>13.038</td>
<td>36.875</td>
<td></td>
<td>-70.9</td>
</tr>
<tr>
<td>disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infectious diarrhea</td>
<td>30.801</td>
<td>41.484</td>
<td>40.858</td>
<td>50.615</td>
<td></td>
<td>13.0</td>
</tr>
<tr>
<td>Dysentery</td>
<td>3.217</td>
<td>&lt;0.001</td>
<td>2.613</td>
<td>2.438</td>
<td></td>
<td>-25.6</td>
</tr>
<tr>
<td>Acute hemorrhagic</td>
<td>2.010</td>
<td>2.694</td>
<td>1.229</td>
<td>1.110</td>
<td></td>
<td>-47.7</td>
</tr>
<tr>
<td>conjunctivitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typhoid and paratyphoid</td>
<td>0.551</td>
<td>0.605</td>
<td>0.415</td>
<td>&lt;0.001</td>
<td>0.367</td>
<td>-28.2</td>
</tr>
<tr>
<td>Hepatitis A</td>
<td>0.338</td>
<td>0.344</td>
<td>0.187</td>
<td>0.124</td>
<td></td>
<td>-45.2</td>
</tr>
<tr>
<td>Cholera</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>&lt;0.001</td>
<td></td>
<td>-98.5</td>
</tr>
<tr>
<td>Poliomyelitis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexually transmitted</td>
<td>32.133</td>
<td>39.336</td>
<td>35.562</td>
<td>38.753</td>
<td>0.197</td>
<td>-0.5</td>
</tr>
<tr>
<td>and blood-borne</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatitis B</td>
<td>11.900</td>
<td>12.626</td>
<td>9.800</td>
<td>9.200</td>
<td>0.001</td>
<td>-21.9</td>
</tr>
<tr>
<td>Syphilis</td>
<td>9.133</td>
<td>12.659</td>
<td>13.335</td>
<td>15.916</td>
<td>0.001</td>
<td>22.4</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>3.099</td>
<td>3.492</td>
<td>2.964</td>
<td>2.942</td>
<td>0.196</td>
<td>-10.1</td>
</tr>
<tr>
<td>Hepatitis C</td>
<td>0.647</td>
<td>&lt;0.001</td>
<td>0.525</td>
<td>&lt;0.001</td>
<td>0.459</td>
<td>-20.8</td>
</tr>
<tr>
<td>Hepatitis D</td>
<td>0.002</td>
<td>0.004</td>
<td>0.001</td>
<td>0.002</td>
<td></td>
<td>-66.7</td>
</tr>
<tr>
<td>Zoonotic</td>
<td>0.660</td>
<td>0.861</td>
<td>0.899</td>
<td>0.982</td>
<td>0.004</td>
<td>18.2</td>
</tr>
<tr>
<td>Brucellosis</td>
<td>0.364</td>
<td>0.497</td>
<td>0.63</td>
<td>0.721</td>
<td>&lt;0.001</td>
<td>46.3</td>
</tr>
<tr>
<td>Hepatitis E</td>
<td>0.144</td>
<td>0.186</td>
<td>0.117</td>
<td>0.131</td>
<td></td>
<td>-29.1</td>
</tr>
<tr>
<td>Hydatid disease</td>
<td>0.132</td>
<td>0.158</td>
<td>0.138</td>
<td>0.117</td>
<td></td>
<td>-4.8</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Rabies</td>
<td>0.011</td>
<td>0.011</td>
<td>0.008</td>
<td>0.007</td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td>Anthrax</td>
<td>0.007</td>
<td>&lt;0.001</td>
<td>0.008</td>
<td>—</td>
<td>0.003</td>
<td>—</td>
</tr>
<tr>
<td>Leptospirosis</td>
<td>0.002</td>
<td>—</td>
<td>0.003</td>
<td>—</td>
<td>0.004</td>
<td>—</td>
</tr>
<tr>
<td>H5N1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>H7N9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Severe acute respiratory syndrome</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Vector-borne</td>
<td>0.542</td>
<td>0.005</td>
<td>1.175</td>
<td>0.001</td>
<td>0.335</td>
<td>0.002</td>
</tr>
<tr>
<td>Hemorrhagic fever</td>
<td>0.205</td>
<td>0.001</td>
<td>0.169</td>
<td>—</td>
<td>0.179</td>
<td>0.001</td>
</tr>
<tr>
<td>Dengue</td>
<td>0.167</td>
<td>—</td>
<td>0.854</td>
<td>0.039</td>
<td>—</td>
<td>0.001</td>
</tr>
<tr>
<td>Japanese encephalitis</td>
<td>0.104</td>
<td>0.004</td>
<td>0.052</td>
<td>0.001</td>
<td>0.035</td>
<td>0.001</td>
</tr>
<tr>
<td>Typhus</td>
<td>0.039</td>
<td>—</td>
<td>0.071</td>
<td>—</td>
<td>0.067</td>
<td>—</td>
</tr>
<tr>
<td>Malaria</td>
<td>0.022</td>
<td>—</td>
<td>0.023</td>
<td>—</td>
<td>0.0100</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Kala-azar</td>
<td>0.005</td>
<td>—</td>
<td>0.005</td>
<td>—</td>
<td>0.005</td>
<td>—</td>
</tr>
<tr>
<td>Schistosomiasis</td>
<td>&lt;0.001</td>
<td>—</td>
<td>0.001</td>
<td>—</td>
<td>0.001</td>
<td>—</td>
</tr>
<tr>
<td>Filarisis</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Plague</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*a*No cases.

**Figure 3.** The trends in (A) incidence, (B) number of cases, and (C) number of deaths for 42 notifiable infectious diseases by disease category, from 2018 to 2021. AHC: acute hemorrhagic conjunctivitis; HD: hydatid disease; HF: hemorrhagic fever; HFMD: hand, foot, and mouth disease; ID: infectious diarrhea; JE: Japanese encephalitis; MM: meningococcal meningitis; SF: scarlet fever; SI: seasonal influenza; SM: schistosomiasis; T/P: typhoid and paratyphoid; TB: tuberculosis.
Figure 4. Temporal changes in the incidence of infectious diseases by disease category and day, from 2018 and 2019 (pre–COVID-19), year 1 (2020), and year 2 (2021) of the COVID-19 pandemic. (A) Overall infectious diseases; (B) Gastrointestinal and enterovirus; (C) Respiratory diseases; (D) Sexually transmitted and bloodborne; (E) Vector borne; (F) Zoonotic.

Figure 5. The leading infectious diseases by incidence by disease category and across age, from 2018 to 2021. HF: hemorrhagic fever; HFMD: hand, foot, and mouth disease; ID: infectious diarrhea; JE: Japanese encephalitis; NA: not applicable; SI: seasonal influenza; TB: tuberculosis.

Regional Variations in Infectious Disease Changes by Epidemic Phase

In the pre-pandemic years, there was evidence of a geographic North-South demarcation (the “Qinling-Huaihe River”), which separates China into 2 regions with varying incidences of notifiable infectious diseases. Compared with North China, South China has a more serious burden of infectious diseases, a pattern that remained evident in 2020 and 2021 (Figure 6 and Multimedia Appendix 5).

We then proceeded to compare changes in regional variation using the categories from 2018, 2019, and 2021, which were delineated based on the 4 phases of the pandemic in 2020.
comparison with the corresponding time phases in 2018 and 2019, the overall incidence of infectious diseases among children and adolescents was greater in each region during phase 1 of 2020. In phase 2, the overall incidence of infectious diseases among children and adolescents markedly declined across the country, and the incidence of notifiable infectious diseases was also lower in almost all regions of China than in the same phase in 2018 and 2019 (Multimedia Appendix 6). This pattern continued to be evident in phase 3 and phase 4 in 2020. However, in 2021, the incidence of infectious diseases among children and adolescents began to markedly recover, with even higher levels than in the prepandemic years in some regions. Considering the magnitude of the seasonal influenza outbreak from 2019 to 2021, we conducted sensitivity analyses by excluding seasonal influenza, and the results remained consistent (Multimedia Appendix 7).

Figure 6. (A) Regional variation in infectious diseases incidence by year and (B) percent change of incidence in each epidemic phase during COVID-year 1 in 2020 and COVID-year 2 in 2021 compared with the average pre–COVID-19 levels for 2018-2019 at the city level.

Persistence and Variability of the COVID-19 Pandemic by Notifiable Disease Category

Overview
As shown in Figure 7 and Multimedia Appendix 8, an indirect effect of the COVID-19 pandemic was evident on infectious diseases, characterized by a significant decrease in overall incidence during the initial year of the pandemic (2020), succeeded by a resurgence toward pre–COVID-19 pandemic levels for certain diseases. However, the degree of this phenomenon varied across the phases depending on the disease category.
**Respiratory Diseases**

The indirect impact of COVID-19 on respiratory diseases among Chinese children and adolescents appeared to be relatively consistent, as evidenced by a reduced incidence of each respiratory disease during the 2 pandemic years compared with the 2 prepandemic years. A rebound in the incidence rate of respiratory diseases was observed only during phase 4 in 2021, coinciding with a return to prepandemic rates. COVID-19–related restrictions appeared to impact the incidence of seasonal influenza, mumps, tuberculosis, scarlet fever, rubella, pertussis, and measles, as their incident IRRs were always lower during the 2 pandemic years than in the corresponding phases before the COVID-19 pandemic. The overall rebound in phase 4 of 2021 was driven mainly by seasonal influenza, pertussis, and meningococcal meningitis with larger IRRs of 1.02 (95% CI 0.74-1.41), 2.83 (95% CI 2.51-3.18), and 1.21 (95% CI 1.02-1.43) in phase 4 of 2021, respectively.

**Gastrointestinal and Enterovirus Diseases**

The incidence rebound started from phase 3 of 2020 and reached a similar level to prepandemic years in phase 4 of 2020, which was maintained during 2021. The rebound of gastrointestinal and enterovirus diseases was primarily led by infectious diarrhea, with incident IRRs in phases 2, 3, and 4 of 2020 being 0.45 (95% CI 0.41-0.50), 1.22 (95% CI 1.17-1.28), and 1.48 (95% CI 1.41-1.55), respectively. A similar pattern was observed in other gastrointestinal and enterovirus diseases, which exhibited varying degrees of rebound from phase 3 of 2020 onward, although they did not fully return to the levels seen before the COVID-19 pandemic in 2018 and 2019.

**Sexually Transmitted and Blood-Borne Diseases**

The indirect impact of COVID-19 on reducing sexually transmitted and blood-borne diseases was not sustained, as there was a rapid rebound starting from phase 3 of 2020 when China relaxed its most stringent prevention measures. Subsequently, the incidence of these diseases returned to pre–COVID-19 levels.
levels. This rapid rebound, particularly notable in South China, was largely driven by increases in the incident IRRs for syphilis and gonorrhea during phase 3 of 2020, reaching 1.31 (95% CI 1.23-1.40) for syphilis and 1.10 (95% CI 1.05-1.16) for gonorrhea (see Multimedia Appendices 9 and 10).

**Zoonotic Diseases**

The indirect impact of COVID-19 on zoonotic diseases appeared to be weak, as there was a rapid rebound observed after phase 3 of 2020, with the incidence levels remaining similar to each phase of the pre–COVID-19 years. Brucellosis stood out as the most pronounced rebound point, showing an incident IRR of 2.08 (95% CI 1.86-2.33) in phase 3 of 2020, especially notable in North China.

**Vector-Borne Diseases**

The indirect impact of COVID-19 on vector-borne diseases appeared to be relatively enduring, as the incidence during the COVID-19 pandemic years consistently remained lower than the levels observed before the pandemic. Despite rebounds in incidence for diseases such as hemorrhagic fever and typhus, these were insufficient to alter the overall trend for this group of diseases.

**Discussion**

**Principal Findings**

This examination of the incidence and mortality of 42 notifiable infectious diseases among Chinese children and adolescents aged 5-22 years in the period immediately preceding and the 2 years following the COVID-19 pandemic reveals that initially, the pandemic had a significant indirect effect, leading to notable decreases in all infectious diseases. However, this effect was not sustained, and the incidence of infectious diseases rebounded to previous levels during the second year of the pandemic, especially in southern China, where COVID-19 restrictions were largely lifted by 2021. The indirect effect of the pandemic varied by disease category, region, and time, with sexually transmitted and blood-borne diseases rebounding first, followed by gastrointestinal and enterovirus, and zoonotic diseases. By contrast, the indirect effect of the COVID-19 pandemic seemed to have a more continuous impact on vector-borne and respiratory diseases. Vector-borne diseases remained below prepandemic levels throughout the study period, while respiratory diseases only rebounded toward the end of the study period.

**Comparison With Other Studies**

Numerous studies have demonstrated a decrease in the incidence of infectious diseases during the initial phases of the COVID-19 pandemic, coinciding with the implementation of strict COVID-19–related restrictions [16,22-27]. For instance, hospital admission data from England revealed reductions in various infectious diseases during the pandemic, with the most significant decline observed in the incidence of seasonal influenza [16]. Other data from China indicated reductions in all infectious diseases, with respiratory diseases experiencing the most substantial declines [19,22,24]. However, a research study in Australia suggested that while various vaccine-preventable infectious diseases declined, the incidence of sexually transmitted and blood-borne diseases increased during the pandemic [26]. We demonstrated that while the incidence of hepatitis B, C, and D, as well as HIV/AIDS, remained stable or experienced slight declines during the COVID-19 pandemic, cases of syphilis and gonorrhea increased. Our study provides new evidence indicating that the implementation of national restrictions in response to the COVID-19 pandemic did not consistently or uniformly reduce the incidence of infectious diseases. This emphasizes the significance of public health professionals collaborating with governments to formulate comprehensive prevention and control policies addressing a broader spectrum of infectious diseases, beyond solely focusing on pandemic measures. This effort needs to encompass various aspects, ranging from health education to immunization, and from individuals to institutions, with a focus on both children and adolescents as well as their parents. A key finding from this study was the significant geographic disparities in infectious disease distribution observed both before and during the COVID-19 pandemic years. For instance, zoonotic diseases primarily affected western regions, whereas vector-borne diseases were mostly concentrated in coastal zones. While reflecting prepandemic patterns, these disparities stem from various factors including economic, geographic, climatic, and social differences. Urbanization, for example, aggregates populations and potentially increases the risk of disease transmission [28-35]. The complexity of these regional distribution patterns highlights the importance of conducting in-depth research and disease surveillance to unravel the nuanced relationships between environmental factors and socioeconomic determinants of health.

The resurgence in respiratory diseases observed in the last quarter of 2021 was primarily fueled by upticks in seasonal influenza, pertussis, and meningococcal meningitis. Interestingly, these conditions also demonstrated the most pronounced sustained decline in incidence. This indicates that the most significant indirect effect of COVID-19 restrictions was observed in other respiratory infections, likely reflecting the impact of sociobehavioral restrictions on the transmission of respiratory diseases [36,37]. The persistent behavioral restrictions aimed at reducing COVID-19 transmission, which remained in place even after school reopenings, might have contributed to the relative delay observed in returning to pre–COVID-19 levels of seasonal influenza, mumps, scarlet fever, rubella, and measles. The rebound in the incidence of seasonal influenza, pertussis, and meningococcal meningitis coincided with the peak season for these infectious diseases, which also aligned with the lifting of restrictions. During the pandemic, the decreased incidence of respiratory infectious diseases resulted in lower viral exposure and reduced immunization rates, leading to an “immunity debt.” This phenomenon increases the proportion of individuals susceptible to infection while gradually reducing herd immunity in the population [38]. In addition to immunization against influenza, pertussis, meningococcal meningitis, and COVID-19, wearing masks during the peak season for infectious respiratory diseases may represent an effective response to mitigate the scale of rebound in respiratory infections, as evidenced by ongoing efforts in Europe [38].
During the COVID-19 pandemic, there was a notable decrease in the incidence of gastrointestinal and enterovirus diseases among Chinese children and adolescents, which gradually increased as COVID-19 restrictions loosened, eventually returning to average levels. This pattern aligns with the impact of COVID-19 restrictions on social contact. In line with previous research, the most significant decline in gastrointestinal and enterovirus diseases was observed for HFMD, a seasonal virus that frequently affects children in school settings [19,22,25]. We demonstrated that infectious diarrhea and dysentery rebounded initially, followed by HFMD. However, acute hemorrhagic conjunctivitis, typhoid and paratyphoid, and hepatitis A exhibited persistently lower levels during the monitoring period. Understanding these nuances is crucial. For instance, 2 rounds of rebound have been described for HFMD: the first coinciding with the reopening of schools and the second corresponding to the seasonal increases experienced in spring and early summer [39].

During the initial phase of the COVID-19 pandemic, the incidence of sexually transmitted and blood-borne diseases decreased but rebounded to prerestriction levels when COVID-19–related restrictions were partially lifted. Gonorrhea and HIV/AIDS rebounded to average levels observed across 2018 and 2019, while syphilis rebounded to a higher incidence during the second year of the pandemic. Research conducted in the United States [40] and Germany [41] has indicated a positive correlation between sexually transmitted diseases and population mobility, implying that the COVID-19 restrictions effectively curbed population movement. While restrictions on population mobility might have reduced access to sexual partners and decreased high-risk sexual behavior and injecting drug use, the inability to attend hospitals for screening might have also contributed to the reduction of these infections during the pandemic [42]. As restrictions on population mobility were lifted before the reopening of schools, the rise in the incidence of gonorrhea and HIV/AIDS to levels exceeding historical norms is potentially driven by out-of-school adolescents. Previous research has also indicated that restrictions implemented to reduce COVID-19 transmission were ineffective in stemming the spread of syphilis [43]. In the United States, syphilis is more prevalent among individuals engaging in unprotected sex or having multiple partners, those who are HIV positive, and those who engage in sexual activity with peers, with recent significant increases observed among women [44]. One possible factor during the pandemic was the reduced access to condoms, which may have led to an increase in unprotected intercourse [45]. These findings underscore the importance that, during any pandemic, efforts to ensure the preservation of programs and interventions aimed at identifying and treating sexually transmitted infections are critical, alongside specific responsive measures.

Vector-borne diseases, along with respiratory infectious diseases, were the conditions that experienced the most pronounced benefits from COVID-19 pandemic measures, with incidence levels remaining lower for an extended period, despite some rebound observed in hemorrhagic fever and typhus. The restrictions on population mobility and the stringent control of students’ mobility during the pandemic seem to have substantially reduced opportunities for children and adolescents to come into contact with vectors and animal reservoirs of vector-borne and zoonotic diseases. This has led to a decreased incidence of both types of infectious diseases, consistent with evidence observed in adults [19]. While the highly cautious approach by the Chinese government to international travel may have contributed to this phenomenon, a resurgence of vector-borne diseases is anticipated once international travel resumes [46], as evidenced in Italy [47]. International travel has minimal impact on zoonoses, diseases that appear to be more prevalent in rural areas where children and adolescents have increased interaction with animals. School closures in these areas may elevate the risk of exposure to diseases such as brucellosis for students residing in rural areas [48].

Implications

Human civilization has encountered numerous pandemics throughout history. Interventions implemented during the COVID-19 pandemic have proven effective in “buying time” for vaccine development and reducing human mortality. However, there has been considerably less focus on measures to control the rebound in the incidence of common infectious diseases that often resurge during the later phases of any epidemic. In formulating strategies to address future pandemics, public health policy makers and governments are urged to consider approaches aimed at limiting the anticipated rebound in the incidence of other infectious diseases, particularly for disease categories where the rebound is projected to surpass prepandemic levels. Priority infectious diseases should be selected based on regional monitoring data, and comprehensive strategies should be developed for all categories of infectious diseases. Implementing interventions in schools may prove effective in preventing the rebound of gastrointestinal and enterovirus diseases while expanding vaccination programs can help address the immunization deficit resulting from limited exposure to respiratory infectious viruses or vaccine shortages during the pandemic. Maintaining testing and treatment services for sexually transmitted diseases is crucial across any pandemic while developing response strategies for imported vector-borne illnesses is imperative in any postpandemic phase.

Strengths and Limitations

This study possesses several notable strengths. First, we used data from the CISDCP, a long-term, systematic surveillance system covering over 85% of health facilities in China [20,49]. In China, health monitoring is largely conducted within schools. However, the utilization of the CISDCP allowed us to access data for out-of-school children and adolescents as well, thereby further increasing the representativeness of our study. While the CISDCP encompasses suspected cases, carriers of pathogens, and confirmed cases, we specifically included only patients with diagnoses supported by both uniform clinical standards and laboratory tests. This approach undoubtedly enhanced the accuracy of our results. Several potential limitations should be noted as well. The incidence of infectious diseases is influenced by multiple factors during the COVID-19 pandemic, and this study solely analyzed the impact of COVID-19–related restrictions, without considering specific meteorological, travel, and human mobility factors. Furthermore, the collective nature...
of COVID-19–related restrictions poses challenges in assessing the effects of individual restrictions. In addition, the variability in COVID-19 policy responses across Chinese provinces in 2021, such as fluctuating lockdowns and reopening, as well as variations in policy enforcement, complicates the distinction between affected and unaffected areas within these analyses. Indeed, the variability in policy implementation could potentially result in an underestimation of the impact of 2021 policies on the incidence rates of infectious diseases.

Conclusions
This study has unveiled the indirect impact of COVID-19–related restrictions on the incidence of infectious diseases among 5-22-year-old children and adolescents in China. Throughout the COVID-19 pandemic, there was a notable reduction in the incidence of the majority of infectious diseases, particularly respiratory and vector-borne diseases. However, the lifting of national COVID-19–related restrictions led to a rapid rebound in gastrointestinal and enterovirus diseases, sexually transmitted and blood-borne diseases, and zoonotic diseases, particularly in southern China. The overall resurgence of infectious diseases was primarily driven by respiratory diseases such as seasonal influenza, pertussis, and meningococcal meningitis, as well as gastrointestinal and enterovirus diseases such as HFMD and infectious diarrhea. Some sexually transmitted and blood-borne diseases, such as syphilis and gonorrhea, did not exhibit any reductions and instead showed persistently rising levels over the pandemic years. Planning for future pandemics must acknowledge that while mitigating strategies for the specific infectious agent are crucial, investment in broader efforts must also continue to protect children and adolescents. Beyond health education and access to routine immunizations, strategies should encompass precise approaches for different infectious diseases; strengthening disease surveillance; and ensuring access to prevention, diagnosis, and treatment services for sexually transmitted infections.

Acknowledgments
This study was supported by the National Natural Science Foundation (grant 82103865 to YD and grant 82273654 to YS) and Beijing Natural Science Foundation (grant 7222244 to YD and grant 7222247 to YS). We acknowledge Professor David Burgner (Murdoch Children’s Research Institute, Royal Children’s Hospital) for helpful preliminary discussions and suggestions.

Data Availability
All of the data (deidentified) collected in the surveillance system can be shared with investigators whose proposed use of the data has been approved by an independent review committee identified for this purpose by contacting the corresponding author. The study protocol and statistical analysis plan must be approved by the committee. Proposals should be directed to dongyanhui@bjmu.edu.cn, songyi@bjmu.edu.cn, or majunt@bjmu.edu.cn.

Authors’ Contributions
YD and YS have equal contributions to this study. LC, LW, and YX conceptualized and designed the study, completed the statistical analyses, drafted the initial manuscript, and reviewed and revised the manuscript. YD, JEM, and YS contributed to the conceptualization and design of the study; supervised the data collection, statistical analyses, and initial drafting of the manuscript; and reviewed and revised the manuscript. SS and JM contributed to the interpretation of the data, and critically reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Conflicts of Interest
None declared.

Multimedia Appendix 1
The trends in number of cases, incidence, number of deaths, and mortality rate for 42 notifiable infectious diseases by year and quarter.
[DOCX File , 18 KB - publichealth_v10i1e47626_app1.docx ]

Multimedia Appendix 2
The changes in incidence (per100,000) and mortality (per 100,000) for 42 notifiable infectious diseases in China, from 2018 to 2021.
[DOCX File , 33 KB - publichealth_v10i1e47626_app2.docx ]

Multimedia Appendix 3
The trends in incidence and proportion for 42 notifiable infectious diseases by disease category, from 2018 to 2021.
Multimedia Appendix 4
Ranking of incidence of each of the 42 notifiable infectious disease by year, from 2018 to 2021.

Multimedia Appendix 5
The incidence of 5 categories for 42 notifiable infectious diseases at the city level, from 2018 to 2020.

Multimedia Appendix 6
The percent changes of cumulative incidence of 5 categories for 42 notifiable infectious diseases by epidemic phase at the city level.

Multimedia Appendix 7
Incidence of overall infectious diseases and percent changes of cumulative incidence in each epidemic phase between 2020 and 2021 and the average of 2018-2019 at the city level (excluding seasonal influenza).

Multimedia Appendix 8
IRRs for incidence for 42 notifiable infectious diseases in China, from 2018 to 2021. IRR: incidence rate ratio.

Multimedia Appendix 9
IRRs for incidence for 42 notifiable infectious diseases in North China, from 2018 to 2021. IRR: incidence rate ratio.

Multimedia Appendix 10
IRRs for incidence for 42 notifiable infectious diseases in South China, from 2018 to 2021. IRR: incidence rate ratio.

References


**Abbreviations**

CISDCP: China Information System for Disease Control and Prevention

HFMD: hand, foot, and mouth disease

IRR: incidence rate ratio
PC: percent change
Exploring the Association Between Structural Racism and Mental Health: Geospatial and Machine Learning Analysis

Fahimeh Mohebbi, MS; Amir Masoud Forati, PhD; Lucas Torres, PhD; Terri A deRoon-Cassini, MS, PhD; Jennifer Harris, MEd; Carissa W Tomas, PhD; John R Mantsch, PhD; Rina Ghose, PhD

1. College of Engineering and Applied Science, University of Wisconsin-Milwaukee, Milwaukee, WI, United States
2. Department of Medicine, University of Wisconsin-Madison, Madison, WI, United States
3. Department of Psychology, Marquette University, Milwaukee, WI, United States
4. Division of Trauma & Acute Care Surgery, Department of Surgery, Medical College of Wisconsin, Milwaukee, WI, United States
5. Community Relations-Social Development Commission, Milwaukee, WI, United States
6. Division of Epidemiology, Institute for Health and Equity, Medical College of Wisconsin, Milwaukee, WI, United States
7. Department of Pharmacology & Toxicology, Medical College of Wisconsin, Milwaukee, WI, United States

Corresponding Author:
John R Mantsch, PhD
Department of Pharmacology & Toxicology
Medical College of Wisconsin
8701 Watertown Plank Rd
Milwaukee, WI, 53226
United States
Phone: 1 4149558861
Email: jomantsch@mcw.edu

Abstract

Background: Structural racism produces mental health disparities. While studies have examined the impact of individual factors such as poverty and education, the collective contribution of these elements, as manifestations of structural racism, has been less explored. Milwaukee County, Wisconsin, with its racial and socioeconomic diversity, provides a unique context for this multifactorial investigation.

Objective: This research aimed to delineate the association between structural racism and mental health disparities in Milwaukee County, using a combination of geospatial and deep learning techniques. We used secondary data sets where all data were aggregated and anonymized before being released by federal agencies.

Methods: We compiled 217 georeferenced explanatory variables across domains, initially deliberately excluding race-based factors to focus on nonracial determinants. This approach was designed to reveal the underlying patterns of risk factors contributing to poor mental health, subsequently reintegrating race to assess the effects of racism quantitatively. The variable selection combined tree-based methods (random forest) and conventional techniques, supported by variance inflation factor and Pearson correlation analysis for multicollinearity mitigation. The geographically weighted random forest model was used to investigate spatial heterogeneity and dependence. Self-organizing maps, combined with K-means clustering, were used to analyze data from Milwaukee communities, focusing on quantifying the impact of structural racism on the prevalence of poor mental health.

Results: While 12 influential factors collectively accounted for 95.11% of the variability in mental health across communities, the top 6 factors—smoking, poverty, insufficient sleep, lack of health insurance, employment, and age—were particularly impactful. Predominantly, African American neighborhoods were disproportionately affected, which is 2.23 times more likely to encounter high-risk clusters for poor mental health.

Conclusions: The findings demonstrate that structural racism shapes mental health disparities, with Black community members disproportionately impacted. The multifaceted methodological approach underscores the value of integrating geospatial analysis and deep learning to understand complex social determinants of mental health. These insights highlight the need for targeted interventions, addressing both individual and systemic factors to mitigate mental health disparities rooted in structural racism.

(JMIR Public Health Surveill 2024;10:e52691) doi:10.2196/52691

https://publichealth.jmir.org/2024/1/e52691

JMIR Public Health Surveill 2024 | vol. 10 | e52691 | p.2380
(page number not for citation purposes)
KEYWORDS
machine learning; geospatial; racial disparities; social determinant of health; structural racism; mental health; health disparities; deep learning

Introduction

Structural Racism and Discrimination (SRD) is a fundamental determinant of health disparities and poor health outcomes among historically marginalized communities in the United States [1]. SRD refers to systemic or institutional racism and societal norms that constrain the chances, resources, influence, and welfare of people and communities due to their racial and ethnic background and other characteristics. Racial segregation in urban centers is an outcome of SRD-driven policy inequalities such as discriminatory mortgage lending practices (redlining), confining Black Americans to central city neighborhoods, which became sites of concentrated poverty and heightened inequities. Residents of segregated neighborhoods experience disproportionate exposure, susceptibility, and vulnerability to economic and social inequality, environmental pollution, toxic substances, and unsafe conditions, thereby affecting individual health conditions, health practices, and access to health care services [2]. Further, neighborhood-level racial and ethnic segregation determines and limits access to educational, employment, and health-related resources [1]. Studies have emphasized the significance of neighborhood segregation on health inequity [3].

Past research has focused on the relationship between interpersonal discrimination and health; however, SRD is likely to have broad downstream effects on psychological, biological, physiological, and behavioral processes [4]. Mental health is of particular importance, especially since the start of the COVID-19 pandemic, where estimates of pooled prevalence of depression are 7 times higher than expected [5], with minority Americans experiencing more severe and chronic symptoms across time [6]. As individuals are embedded within larger systems of influence, it is essential to understand the relationship between SRD and mental health at a community level [5]. The socioecological model of health provides a framework to examine how individual health and behavior are impacted by interpersonal, neighborhood, and societal factors [7]. The National Institute on Minority Health and Health Disparities (NIMHD) has encouraged a place-based approach, as “relationships between SRD and physical/mental health are influenced by numerous place-based factors… individual-level factors (i.e., health-related behaviors, ways of coping) are understood best within the context of the lived environment and structural policies that perpetuate inequities” [1].

To examine the contribution of SRD to inequalities, various measures have been used including racial residential segregation [8]. However, no domain of structural racism operates in isolation, multiple index measures of SRD have been created and applied. Dougherty et al [9] developed a structural racism index measure by combining 7 measures of SRD: housing dissimilarity index, school dissimilarity index, high school graduation ratio, incarceration ratio, poverty ratio, primary care ratio, and ambulatory care ratio. Still, there are no consistent, agreed-upon relevant content domains of structural racism [10].

Instead of selecting 1 approach or developing yet another index, we examined mental health disparities without reference to race to determine whether communities experiencing SRD can be identified using an unsupervised retrospective approach.

This study used the NIMHD framework to explore the relationship between SRD and mental health. By using machine learning algorithms with geodata science, we conducted spatial modeling and geovisualization to investigate how location-based factors, indicative of SRD, impact mental and physical health. This approach allows for a statistical analysis of these dynamics over time and space, offering a comprehensive analysis of the effects of place on health outcomes [11]. Data were analyzed at multiple geographic scales—county, city, and census tracts. Understanding mental health at the community level is important, given the complex intersectionality of factors that promote or hinder health in the United States [12].

Our study site is Milwaukee County (population of 918,661), the most populous county in Wisconsin, with a racially diverse population of Black (n=241,608, 26.3%), Hispanic or Latino (n=143,311, 15.6%), and White origin (n=541,091, 58.9%) [13]. The county includes the hypersegregated city of Milwaukee (population 636,305), where redlining confined the African American population to its central city neighborhoods. Devastated by deindustrialization and disinvestments, these neighborhoods exhibit concentrated urban poverty, heightened spatial inequalities, and dramatic health disparities [14], where Black residents experience a poverty rate 5 times higher than that of White residents and White residents outlive Black residents by almost 14 years [13]. Acknowledging the damaging legacy of SRD on health, Milwaukee County was among the first jurisdictions in the United States to declare racism as a public health crisis in 2019, with 170 jurisdictions following its suit.

Methods

Overview

To identify geospatial determinants of health across behavioral indicators, built environment, sociocultural environment, and health care (based on the NIMHD framework), georeferenced data sets were acquired from the United States Census Bureau. The United States Census Bureau anonymizes and deidentifies data before releasing them to the public. Detailed demographic data can be obtained only at the Census Tract level (each tract comprises 4000 residents) and not at an individual level. In accordance with census policies, we analyzed all data at the census tract level, and no individual-level data have been obtained or used. Selected variables (eg, age, gender, population, race, ethnicity, marital status, educational attainment, educational enrollment, employment, neighborhood stability, and poverty) were compiled into a data set of 217 explanatory variables. Notably, race-based factors were deliberately withheld during the initial stages of our unsupervised machine learning analysis. This approach allowed the model to operate without
the assumption that racial disparities significantly influence mental health outcomes. It was only in the final stage of the analysis that race and ethnicity variables were incorporated, providing an opportunity to observe whether the emerging patterns of poor mental health prevalence correlated with racial factors. All variables were joined to the administrative boundary shapefile of Milwaukee County census tracts collected from the TIGER/Line database11 (using ArcGIS Desktop 10.7; Environmental Systems Research Institute).

**Variable Selection Procedure**

**Overview**

Variable selection minimizes the number of predictors in quantitative models to improve efficiency and reduce complexity. Public health research commonly uses conventional techniques such as subject matter expert selection and regression-based stepwise selection. However, tree-based methods such as random forest can handle nonlinear, nonparametric relationships and provide more robust results [15].

Extreme multicollinearity can cause parameter estimate instability, unintuitive parameter signs, high $R^2$ diagnostics despite few or no significant parameters, and inflated standard errors of the parameter estimates [16]. To avoid this, variance inflation factor (VIF) and Pearson correlation analysis are used to detect multicollinearity and select the most uncorrelated variables for the random forest variable selection model. The variable selection method [17] implemented in the Variable Selection Using Random Forest (VSURF) R package (R Foundation for Statistical Computing) yields the best results compared with other variable selection and baseline methods [18]. The method ranks variables by their importance, eliminating the least important ones and constructing a sequence of random forest models. Eventually, variables of the most optimized model are selected.

**Geographically Weighted Random Forest**

Spatial modeling enables the examination of how variables behave across geographical space to identify spatial heterogeneity and dependence. Local spatial modeling offers a more informed approach to understanding complex phenomena compared with conventional global approaches [19]. The geographically weighted random forest (GWRF), a localized model, is ideal for conducting public health research. However, GWRF has not been used to examine the variability in relationships between place-based risk factors and the prevalence of poor mental health. This study used GWRF to investigate nonstationarity and localized associations between risk factors and the prevalence of poor mental health.

Local feature importance in GWRF measures how much each feature contributes to the accuracy of the model within a specific region, calculated by the increase in mean squared error or the decrease in node impurities averaged over all trees in the model. These measures are derived from the out-of-bag (OOB) error, which is a measure of the model’s performance on data not used during training [20].

**Self-Organizing Maps**

Self-organizing maps (SOMs) are unsupervised artificial “neural” networks that create a 2D space topographic map of a data set. “Neurons,” or relational clusters, are organized to preserve their context or neighborhood, and SOMs use closeness or neighborhood function to display input space properties. We implemented a SOM model to explore a data set containing 296 census tracts in Milwaukee County×12 determinant factors of poor mental health. The SOM was trained on this data set using a 5×5 hexagonal grid, selected for its ability to create a more comprehensive and representative map of the data. After training the SOM, we delved deeper into the clusters formed on the map by combining the SOM with k-means clustering, an algorithm known for its efficiency in partitioning data into distinct groups or clusters. We determined the optimal number of clusters for our data set by evaluating the “within-cluster sums of squares” and the “average silhouette” statistics [21].

**Ethical Considerations**

This study uses anonymized and deidentified publicly available data sets that were released by federal or state agencies for public consumption. All publicly available government data are aggregated at the census tract level, where each tract is composed of 4000 individuals. We received full approval from the Medical College of Wisconsin institutional review board with exemption from oversight by Code of Federal Regulations 46.104 (d) 4, as we conducted secondary analyses of publicly available anonymized and deidentified data sets that do not contain any personally identifiable information and do not pose any risk to ethical violations. Given that the project did not involve direct contact with subjects, an informed consent process was not required and no compensation was provided. Moreover, the project did not require a waiver of HIPAA (Health Insurance Portability and Accountability Act) authorization as outlined in 45 Code of Federal Regulations 164.514(e) because it used a Limited Data Set.

**Results**

**Input Factor Selection**

We compiled a database of 217 geospatial determinants of health across behavioral indicators, built environment, sociocultural environment, and health care to analyze the relationship between place-based factors and mental health. Next, we excluded all race-based factors and used the remaining 184 variables to identify highly colinear variables, VIF, and Pearson correlation analysis was applied (VIF threshold=7.5; Pearson correlation threshold=0.75). As a result, 105 uncorrelated factors were selected to be used as the input for the variable selection (VSURF) procedure.

**User Statistics**

The VSURF procedure was applied to the 105 place-based factors to find a sufficiently parsimonious set of significant indicators of poor community mental health (ie, the percentage of adults who stated that their mental health was not good 14 or more days in the past month) and measures of SRD. The VSURF selected 98 variables at the thresholding step, then 24 variables at the interpretation step, and 12 variables at the
prediction step. The most important place-based factors of mental health were adults who smoke, insufficient sleep, adults without health insurance, adults who are obese, adults who are sedentary, marriage rate, people living below the poverty level, childhood opportunity index score, median age, homeownership, full-time employment, and educational attainment (all variables are percentages except for Age and Childhood Opportunity Index score). These variables account for 95.11% of the variability in poor mental health in communities, with a mean of squared residuals of 0.74.

Partial dependence plots depicting the relationship of each of the 12 predictive variables for poor mental health were constructed. Figure 1 illustrates the correlation between the selected 12 variables and poor mental health. The results of the partial dependence plot approach showed that the relationship among the prevalence of smoking, poverty, insufficient sleep, lack of health insurance, single households, sedentariness, and poor mental health was positive. In contrast, the childhood opportunity index, median age, homeownership, and employment rate exhibit negative relationships with poor mental health. Notably, obesity and educational attainment did not follow linear trends in relation to poor mental health, illustrating the complexity of these factors. The largest difference makers were the prevalence of smoking, lack of insurance, poverty, insufficient sleep, employment, and age.

Figure 1. Partial dependence plots illustrate the marginal effect of each of the 12 selected predictor variables on the prevalence of poor mental health. The Y-axis in each plot indicates the prevalence of poor mental health outcomes, quantified as a percentage, and the X-axes represent the value of each predictive variable. The curve in each partial dependence plot shows the relationship between the predictor variable and the prevalence of poor mental health outcomes. An ascending curve implies that an increase in the predictor variable corresponds with an increase in poor mental health outcomes, while a descending curve implies the opposite.

GWRF Implementation
Next, we used GWRF to investigate and visualize local associations between poor mental health prevalence and the 12 identified place-based risk factors at multiple scales—the county, the city, and census tract levels. Six place-based risk factors (prevalence of smoking, lack of insurance, poverty, insufficient sleep, employment, and age) were identified as having a spatially heterogeneous impact. We trained the GWRF with 45 nearest neighbors (census tracts) with bootstrapped 5000 “ntrees” and 4 “mtry” in each tree. It yielded a local model with an $R^2$ square (OOB) of 94.91%, mean squared error (OOB) of 0.051, and Akaike Information Criterion (OOB) of –856.447. Spatial variation of the local contribution of the 6 risk factors to the prediction of poor mental health is shown in Figure 2.
Figure 2. Geographic variation in place-based risk factors impacting mental health. Maps represent a visualization of the localized influences of 6 place-based risk factors—prevalence of smoking, lack of health insurance, poverty level, insufficient sleep, employment status, and median age—on predictions of poor mental health. Visualization is based on the findings of the GWRF model. Each risk factor is represented in a separate map, with color gradients signifying areas of varying impact on mental health. Darker shades indicate regions where a risk factor significantly correlates with poor mental health, while lighter shades correspond to weaker correlations. These visual representations underscore the spatial heterogeneity of mental health determinants and the significance of targeted interventions. GWRF: geographically weighted random forest. For a higher-resolution version of this figure, see Multimedia Appendix 1.

SOMs Implementation

A SOM approach was used to determine impactful clusters of the 12 most significant place-based factors related to mental health. Census tracts were grouped based on similarities in these mental health–related factors. The SOM was integrated with k-means clustering to further delineate and group census tracts based on their similarity. The original 25 SOM grid codes were condensed into 3 clusters, depicted in Figure 3. Each cluster represents a collection of census tracts displaying similar patterns or characteristics across the 12 mental health–related factors. Clusters 1, 2, and 3 differ markedly in their composition, each corresponding to varying levels of mental health risk. Although a more detailed understanding of each cluster necessitates additional data, particularly in terms of specific demographic, socioeconomic, and environmental variables that may influence these risk levels, these clusters can be generally categorized as representing high (cluster 3), moderate (cluster 2), and low (cluster 1) risk for poor mental health. This categorization is based on the aggregated mental health factors within each cluster. Deeper analysis with more comprehensive data could reveal more nuanced distinctions and cluster-specific contributing factors.

Figure 4 represents the geospatial distribution of the 3 mental health risk clusters across Milwaukee County. Classification into low-, moderate-, and high-risk labels refers to the collective intensity of the 12 key place-based mental health factors within each cluster. For instance, the “high-risk” label indicates the prevalence of higher levels of risk factors (ie, smoking prevalence, lack of health insurance, and poverty) known to contribute to poor mental health but does not necessarily mean that all individuals in this cluster have poor mental health.
To examine racial disparities within community clusters, we calculated a disproportionality index, the ratio of the percentage of a specific racial demographic within each cluster compared with that within a base population. Within the low-risk group (cluster 1), the disproportionality index is 1.5 for the White population, indicating an overrepresentation; the Black and Hispanic populations are underrepresented with indices of 0.21 and 0.6, respectively.

In the moderate risk group (cluster 2), the White population is underrepresented, with an index of 0.6; the Black and Hispanic populations are overrepresented, with indices of 1.52 and 1.51, respectively. The most significant disparity was found in the high-risk group (cluster 3), where the Black population’s disproportionality index was 2.23, and the White and Hispanic populations were underrepresented with indices of 0.34 and 0.51, respectively.

The findings of our unsupervised approach demonstrate significant mental health disparities that align with the racial demographics in Milwaukee County, despite our exclusion of race-related factors. The overrepresentation of Black populations in the high-risk cluster combined with the overrepresentation of the White population in the low-risk cluster indicates that disparities observed across clusters are not isolated occurrences but rather are indicative of underlying socioeconomic disparities likely caused by structural racism and discrimination. Notably, while Hispanic community members were underrepresented in the low-risk cluster 1 and overrepresented in the moderate-risk cluster 2, they were underrepresented in the high-risk cluster 3. This highlights that there are differences across racial minority demographics that influence mental health. For example, differences in upward financial mobility have been identified between Black and Hispanic communities [22].

The spatial dependency of risk areas was investigated to discover the locationality of poor mental health areas. To examine the degree of similarity, the Global Moran I index was calculated to measure the degree of similarity between an area and its neighboring areas. The calculated Moran Index is 0.53 with a z score of 21.63, indicating that there is <1% likelihood that the clustering of poor mental health risk was the result of random chance. The findings, displayed in Figure 4, show that clusters 2 and 3 are primarily located in central city neighborhoods of Milwaukee County, aligning with predominantly Black communities in the central and north side of the city and with the Hispanic community in Milwaukee’s south side.

Although generated without incorporating location as a factor, Figure 4 demonstrates spatial autocorrelation, signifying the presence of spatial dependency in mental health outcomes. This suggests that mental health issues are not randomly distributed but rather display geographically linked patterns. Spatial dependency confirms the influence of location or spatial context on mental health. Our results align with previous findings, further demonstrating that mental health outcomes are partly a product of spatially located phenomena such as socioeconomic conditions, access to health care, environmental factors, and SRD. The observations reinforce the assertion that conventional global models are not ideal for examining complex phenomena such as community mental health or SRD [19,23]. Local spatial modeling offers more precise outcomes when processes are affected by both their geographical location and the fluctuating conditions of underlying variables across different times and locations.

Figure 3. Integrated self-organizing map and k-means showing the 3 clusters of census tracts in Milwaukee County based on similarities among the 12 mental health risk factors. The differences between clusters are maximized and within clusters are minimized. The size or area of each wedge reflects its proportional “influence” on mental health in the corresponding census track grouping.
Discussion

Principal Findings

To address the complex issue of SRD and its impact on mental health disparities, our study used a 5-step process. First, we compiled a comprehensive database of place-based explanatory factors, also known as “geospatial determinants of health,” at the census tract scale [24]. Second, we deliberately eliminated all race-based factors, an unprecedented approach to assess whether SRD effects on mental health disparities could be examined without directly referencing race. Third, we conducted a variable selection process, identifying the most defining place-based characteristics that shape the mental health conditions of an affected community. Fourth, using SOMs, we clustered census tracts based on mental health risk factor similarity. Finally, we investigated the racial composition, demographic characteristics, and spatial dependency of risk areas to shed light on racial disparities, confirmed SRD, and illuminated the locationality of mental health and SRD (Table 1).

Our study introduces a novel approach to analyzing mental health disparities and their structural underpinnings, diverging from traditional models. By integrating advanced geospatial and machine learning techniques, it provides a more granular, community-specific insight into these disparities. This method contrasts with conventional approaches that may not fully capture the intricate, locality-specific interactions of factors impacting mental health [8]. In line with the study by Groos et al [8] based on the exploration of structural racism quantification methods, this study exemplifies the shift toward more sophisticated, nuanced methodologies in understanding the complex effects of SRD on health outcomes.
poor mental health in both the central city neighborhoods and the affluent shoreline areas of Milwaukee County.

Age is a determinant of mental health. Adult mental health diagnoses often begin in adolescence, with approximately half of all adult mental health disorders emerging in the teenage years [33]. The median onset age ranges from 8 to 35 years, and increased age is associated with better mental health [34]. Chen et al [35] found that the relationship between census tracts’ median age and prevalence of poor mental health follows a negative curvilinear trend, consistent with a large-scale meta-analysis of 192 epidemiological studies. Our results demonstrate that poor mental health in the central city neighborhoods of Milwaukee County is more significantly associated with age than in other parts of the county, indicating that the influence of age on mental health varies locally.

Although high-risk clusters showed high spatial alignment with socioeconomically disadvantaged communities in Milwaukee, several individual factors were strong predictors of mental health in more affluent and socioeconomically secure areas (eg, in Milwaukee’s lakefront communities). These include unemployment, lack of insurance, sleep deprivation, and smoking. These observations suggest that while these communities are not locations of high-risk clusters, mental health associations are evident in subgroups of community members.

### Geographical Disparities in Mental Health Challenges

Insights gained from the disproportionality indices reveal the need for targeted interventions and policy adjustments in Milwaukee communities. These indices do not only demonstrate disparities but are also critical for guiding approaches that can address racial imbalances in mental health, steering toward a more equitable health care system. Our findings align with research demonstrating that mental health outcomes are the products of spatially located phenomena, including socioeconomic conditions, access to health care, environmental factors, and deeply rooted SRD [36]. This confluence of factors reinforces the assertion that conventional global “one-size-fits-all” approaches fall short of understanding complex phenomena such as community mental health or the intricacies of structural racism and discrimination [19,23]. When

<table>
<thead>
<tr>
<th>Cluster (n)</th>
<th>White, n (%)</th>
<th>Black, n (%)</th>
<th>Hispanic, n (%)</th>
<th>Homeownership, n (%)</th>
<th>Poverty, n (%)</th>
<th>Median household income (US $)</th>
<th>Employment, n (%)</th>
<th>Educational attainment (bachelor), n (%)</th>
<th>Poor mental health, n (%)</th>
<th>Poor physical health, n (%)</th>
<th>Incarceration, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (488,282)</td>
<td>408,839 (83.73)</td>
<td>30,185 (6.53)</td>
<td>42,822 (8.77)</td>
<td>282,715 (57.9)</td>
<td>24,463 (5.01)</td>
<td>71,913</td>
<td>285,547 (58.48)</td>
<td>152,832 (31.30)</td>
<td>93,701 (19.19)</td>
<td>106,543 (21.82)</td>
<td>11,182 (2.29)</td>
</tr>
<tr>
<td>2 (409,305)</td>
<td>144,853 (35.39)</td>
<td>119,145 (29.43)</td>
<td>90,334 (22.07)</td>
<td>138,550 (33.85)</td>
<td>90,538 (22.12)</td>
<td>39,801</td>
<td>179,808 (43.93)</td>
<td>41,012 (10.02)</td>
<td>97,374 (23.79)</td>
<td>107,115 (26.17)</td>
<td>21,161 (5.17)</td>
</tr>
<tr>
<td>3 (56,622)</td>
<td>10,753 (18.99)</td>
<td>38,718 (68.38)</td>
<td>4286 (7.57)</td>
<td>7474 (13.2)</td>
<td>27,320 (48.25)</td>
<td>20,889</td>
<td>15,616 (27.58)</td>
<td>3397 (6)</td>
<td>16,664 (29.43)</td>
<td>17,190 (30.36)</td>
<td>3907 (6.9)</td>
</tr>
<tr>
<td>County (954,209)</td>
<td>531,208 (55.67)</td>
<td>292,370 (30.64)</td>
<td>199,601 (14.63)</td>
<td>407,638 (42.72)</td>
<td>161,548 (16.93)</td>
<td>52,485</td>
<td>466,704 (48.91)</td>
<td>183,781 (19.26)</td>
<td>212,407 (22.26)</td>
<td>234,831 (24.61)</td>
<td>38,455 (4.03)</td>
</tr>
</tbody>
</table>

*With the exception of median household income, all data represent a percentage of the population in the census tracts that comprise each cluster. Data sets were sourced from the Centers for Disease Control and Prevention’s 500 Cities Project and the United States Census Bureau.*
processes are governed by their location and by conditions that fluctuate over time, the use of localized spatial modeling approaches offers a more realistic alignment, yielding more accurate and nuanced results. This perspective challenges traditional frameworks and calls for a tailored approach that considers the unique characteristics of each community, reflecting a more comprehensive understanding of mental health disparities (Table 1) [37-39].

While our approach introduces innovative methods for analyzing mental health disparities, it is crucial to acknowledge the inherent limitations of observational studies. Such limitations highlight the need for caution in interpreting causality from our findings. Our study’s results, therefore, should be considered indicative of associations rather than definitive causal relationships. For example, the relationships of several factors with mental health (eg, smoking, sleep deprivation, and unemployment) are likely bidirectional. This acknowledgment is vital in guiding future research and formulating public health policies that are based on a comprehensive understanding of mental health determinants.

Finally, while our study demonstrates that Black residents are overrepresented in neighborhoods where high-risk factors for poor mental health are localized, our approach does not address risk for Black community members who live in non-Black majority communities. Moreover, our classification of communities as Black, Hispanic, and White does not recognize diversity within these racial demographics (eg, Puerto Rican vs Mexican heritage) [40].

Conclusions
Understanding the complex interplay between SRD and mental health is vital to informing public health policies and interventions. For example, acknowledging that SRD places additional burdens on an individual’s ability to cope with life’s demands calls for targeted support in areas such as mental health services, education, employment, and community infrastructure. Our study, deploying advanced Geographic Information System and unsupervised machine learning analyses, unravels complex spatiotemporal relationships predicting poor mental health while excluding explicit race-related variables. The findings highlight that the risk for poor mental health is intertwined with structural and spatially localized factors that correspond with disproportional racial representation within communities. These insights illustrate a need for reinvestment strategies that recognize, protect, and promote mental health, with a focus on communities disproportionately affected by SRD. Such strategies must be implemented in a manner that considers the multifaceted risks and includes protections against further exacerbating disparities. In an era where mental health disparities persist, our research emphasizes the importance of a targeted and localized approach, prioritizing communities with historical burdens of discrimination, ensuring equitable access to resources, and ultimately fostering a more resilient and inclusive mental health care system.

Acknowledgments
This work was supported by a grant from the Advancing a Healthier Wisconsin Foundation.

Data Availability
The data sets used in this study are publicly available and were sourced from the Centers for Disease Control and Prevention’s 500 Cities Project and the United States Census Bureau.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Geographic variation in place-based risk factors impacting mental health. Maps represent a visualization of the localized influences of 6 place-based risk factors—prevalence of smoking, lack of health insurance, poverty level, insufficient sleep, employment status, and median age—on predictions of poor mental health. Visualization is based on the findings of the GWRF model. Each risk factor is represented in a separate map, with color gradients signifying areas of varying impact on mental health. Darker shades indicate regions where a risk factor significantly correlates with poor mental health, while lighter shades correspond to weaker correlations. These visual representations underscore the spatial heterogeneity of mental health determinants and the significance of targeted interventions. GWRF: geographically weighted random forest.

References


Abbreviations

GWRF: geographically weighted random forest
HIPAA: Health Insurance Portability and Accountability Act
NIMHD: National Institute on Minority Health and Health Disparities
OOB: out-of-bag
SOM: self-organizing map
SRD: structural racism and discrimination
VIF: variance inflation factor
VSURF: Variable Selection Using Random Forest
This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
An Accessible Web-Based Survey to Monitor the Mental Health of People With Mild Intellectual Disability or Low Literacy Skills During the COVID-19 Pandemic: Comparative Data Analysis

Monique CJ Koks-Leensen1,2*, PhD; Anouk Menko1,3*, MSc; Fieke Raaijmakers1,3,4, MA, MSc; Gerdine AJ Fransen-Kuppens1,3,5, Prof Dr; Kirsten E Bevelander1,2,3, PhD

1Department of Primary and Community Care, Radboud university medical center, Nijmegen, Netherlands
2Academic Collaborative Intellectual Disability and Health - Sterker op Eigen Benen (SOEB), Nijmegen, Netherlands
3Academic Collaborative AMPHI - Integrated Health Policy, Nijmegen, Netherlands
4Safety and Health Region Gelderland-Midden, Arnhem, Netherlands
5Municipal Health Service Gelderland Zuid, Nijmegen, Netherlands
*these authors contributed equally

Corresponding Author:
Monique CJ Koks-Leensen, PhD
Department of Primary and Community Care
Radboud university medical center
Geert Grooteplein 21
Nijmegen, 6525 EZ
Netherlands
Phone: 31 243618181
Email: monique.koks-leensen@radboudumc.nl

Abstract

Background: The COVID-19 pandemic and related control measures affected the mental health of all populations. Particular subgroups are underrepresented in mainstream surveys because they are hard to reach, and study measurements are not adapted to their skills. These subgroups include people with lower cognitive and literacy skills, such as people with mild intellectual disability (MID), who were considered vulnerable during the COVID-19 pandemic given their low socioeconomic status, small social networks, increased risks of health problems, and difficulties understanding health-related information.

Objective: This study examines the impact of the COVID-19 pandemic on mental health among people with MID or low literacy skills compared with those predominantly represented in national surveys.

Methods: A repeated cross-sectional study of people with MID or low literacy skills and a general population sample was conducted in the Netherlands. An easy-read web-based survey was co-designed with, and tested among, people with MID or low literacy skills and conducted in 3 rounds within 1 year of the COVID-19 pandemic (T1: November to December 2020, T2: March to April 2021, and T3: September to October 2021). The survey contained questions about demographics and 6 aspects of mental health: feeling happy, feeling energized, feeling stressed, worry, feeling lonely, and sleeping problems.

Results: Our adapted survey and recruitment procedure enabled 1059 persons with MID or low literacy skills to participate (T1: n=412, 38.9%; T2: n=351, 33.1%; and T3: n=296, 28%). They were significantly younger, had a lower level of education, and more often than not were born outside the Netherlands compared to the general population sample (P<.001). Approximately half of them (604/1059, 57.03%) received professional care. They displayed poorer mental health scores than the general population sample. The percentages of people with MID or low literacy skills who reported more negative feelings in T1 ranged from 20.6% (85/412) reporting feeling lonely often or almost always to 57.8% (238/412) reporting feeling happy almost never or sometimes. The general population sample’s percentages were 5.4% (160/2930) and 32.2% (941/2918), respectively. Although scores improved over time in both populations, the disproportional effects remained.

Conclusions: General COVID-19–related restrictions for the entire Dutch population affected people with MID or low literacy skills more negatively than the general population. Our study underscores the relevance of including these subpopulations in public health research because they are often overlooked in regular health data. An accessible web-based survey particularly targeted at this population enabled us to do so, and we reached a group of respondents significantly different from regular survey
participants. This survey’s results provided insights into the health of people with MID or low literacy skills and gained knowledge to be used by care organizations and policy makers to reduce health disparities during a pandemic and in general.

*(JMIR Public Health Surveill 2024;10:e44827) doi:10.2196/44827*

**KEYWORDS**

monitoring; mental health; intellectual disabilities; low literacy; COVID-19; web-based survey

**Introduction**

Studying COVID-19–Related Impact

The COVID-19 pandemic and related disease control measures affected the entire world. People were advised to adhere to strict hygiene measures and to work from home (if possible), and public places and nonessential shops were closed. In addition, social distancing and visiting restrictions were in place during lockdowns. In general, these restrictions had a major impact on daily routines, social contacts, and mental health [1-3], affecting some individuals more than others [4]. In the Netherlands, the *Rijksinstituut voor Volksgezondheid en Milieu* (RIVM; National Institute for Public Health and the Environment) started conducting a national longitudinal survey to monitor the impact of COVID-19 and related measures on Dutch citizens [5,6]. The monitor is informative regarding disease control strategies and policy making [7]; however, there is an external validity bias because participant characteristics show that a majority of the participants have a high educational level and are middle-aged or older and women [6,8,9]. Indeed, studies have demonstrated that particular subgroups are often underrepresented and excluded from mainstream surveys because general recruitment strategies are unsuitable for reaching them, and study measurements are not adapted to their cognitive level or literacy skills [10-12]. These groups often comprise people with lower socioeconomic status and higher risks of health problems, and, in the case of the COVID-19 pandemic, more difficulties complying with preventive measures [13-15], given their housing or work situation.

This study examines the impact of the COVID-19 pandemic and related restrictions on mental health and well-being among people with lower cognitive and literacy skills in addition to those who are predominantly represented in the national survey. An accessible survey based on the national survey was developed, and alternative recruitment techniques were used to specifically include these underrepresented subgroups.

**Subpopulations at Risk for Greater COVID-19–Related Impact**

In the Netherlands, approximately 19% of the adult population (ie, 2.5 million adults) have limited reading, writing, or numeracy skills [16]. These limitations have various causes, such as a low level of educational attainment, migrant background, parents’ level of education and literacy, or low information-processing skills [17]. The last item plays an important role in people with mild intellectual disability (MID), who experience considerable limitations in both intellectual functioning and adaptive behavior and often need support in their daily life [18]. It is estimated that 4% to 8% of the Dutch population have an MID [19]. People with low literacy skills or MID often have limited work and income, poor health, and small social networks [20-24]. In general, studies have shown that people with low education and health literacy as well as those without social support, a stable income, a daily routine, and access to services are more at risk of mental health problems such as anxiety, general distress, and loneliness arising from the COVID-19 pandemic [1,3,5,14,25,26]. Therefore, it is likely that the COVID-19 pandemic had a higher impact on the mental health of people with MID or low literacy skills compared with the general population. However, during the first months of the pandemic, very limited knowledge was available about the impact on this subpopulation, and our study was set up to provide both these essential insights and practical recommendations for policy makers and care providers.

During the COVID-19 pandemic, studies on mental health specifically aimed at people with mild or more severe intellectual disability (ID) showed negative impacts as a result of social isolation or a lack of social support, the rapid changes in COVID-19–related measures and difficulty understanding these measures, difficulty accessing services, and disruption of daily routines [27-29]. Two European surveys among people with ID found that more than half reported stress or anxiety [30,31] or felt more anxious than usual because of the pandemic and subsequent lockdown [31]. A US survey found that 41% of the participants with ID had experienced more mental health problems or symptoms since the pandemic began; worry and stress were most often mentioned [32]. Similarly, people with low health literacy experienced more anxiety disorders, bouts of depression, and sleeping disorders during the COVID-19 pandemic than those showing sufficient health literacy [14]. Altogether, these studies—primarily conducted during the first lockdown periods—showed a great impact on the mental health of people with MID or low literacy skills, which contributed to an increase in preexisting inequalities in health and well-being [24,33].

The current underrepresentation in national surveillance and surveys of people with MID or low literacy skills, as well as the consequent lack of information about them, adds to existing health disparities. To better understand the impact of the COVID-19 pandemic on people with MID or low literacy skills, the specific factors driving this impact, and their specific needs, more information is urgently needed. Knowledge acquired through monitoring population health in its local context can provide a basis for government and health organizations to develop appropriate strategies to reduce this impact accordingly. In addition, the course of the pandemic and the ever-changing COVID-19 disease control strategies over time are important aspects regarding the context in which people were affected. Due to the rapidly changing situation and regulations during the pandemic (eg, when vaccinations were offered or restrictions

https://publichealth.jmir.org/2024/1/e44827
were lifted), a dynamic impact on mental health was expected, and more insight is needed into how people responded to this unpredictable course.

**Objectives**

This study examines the impact of the COVID-19 pandemic on people with MID or low literacy skills in direct comparison with the general population and over the course of the pandemic at 3 different time points. A unique survey study was set up that complemented the RIVM national survey. This study developed an accessible version of the web-based survey together with representatives of the target population and used suitable techniques to reach people with MID or low literacy skills.

**Methods**

**Study Design**

A repeated cross-sectional study of people with MID or low literacy skills and a general population sample was conducted during the COVID-19 pandemic in the Netherlands. The inclusion criteria were people with reading and writing difficulties, aged ≥16 years, living in the Netherlands, and completion of the survey. No exclusion criteria were used. A control question to assess participants on literacy skills or intellectual ability was not included because this was expected to be too sensitive for the participants. For reasons of comparison, the same survey was presented to 2 panels used to represent the general Dutch population.

The survey was administered 3 times in a 1-year period between November 2020 and November 2021. The first survey (T1) was distributed during a nationwide second lockdown (November to December 2020), the second survey (T2) was administered after the second lockdown and when the Dutch vaccination program had started (April to May 2021), and the third survey (T3) was distributed after the summer when most COVID-19–related restrictions had been lifted (September to October 2021). Figure 1 shows the timeline and the severity of the COVID-19 pandemic in the Netherlands by means of excess mortality rates. Data were derived from Statistics Netherlands [34].

**Study Population and Recruitment**

The surveys were disseminated via organizations working with people with MID or low literacy skills, such as advocacy organizations, care facilities for people with MID, language education organizations and libraries supporting and educating people with no or low literacy skills, social workplaces, the Dutch center of expertise on health disparities Pharos, Special Olympics, and a website offering accessible web-based information and programs for people with limited digital or literacy skills [35]. The surveys were open for between 4 and 6 weeks, giving the organizations time to distribute them within their network. Support was available to allow people who were anxious or unable to complete the survey independently to participate in this study. At the end of each survey, participants were asked to participate in future research, resulting in a panel of participants who could be contacted directly for the subsequent survey rounds.

Our easy-read survey was also distributed within the same time period to 2 municipal health service (MHS) panels: GGD Gelderland-Zuid (n=approximately 2500) and GGD Gelderland-Midden (n=approximately 7000). Each MHS panel consisted of residents in its service area who are regularly asked to complete health surveys. There is a known bias to these panels, in that they generally consist of older and more highly
educated residents, with an overrepresentation of women [36]. We particularly used these panels in our study to obtain comparative data from a Dutch population sample and because these panels were also invited to participate in the national RIVM survey.

All participants received the same survey. Participation was voluntary, and participants could stop completing the survey at any time. Completion of a prior survey was not mandatory for participation in the next round. Data were obtained anonymously; therefore, matching between surveys and paired within-group analysis over time were not possible.

Web-Based Survey Development

The RIVM survey on Dutch citizens’ perception of the COVID-19-related measures, their impact on well-being, and whether people were complying formed the basis of this monitor but was adapted to provide an easy-read version for this study [6]. The national survey consisted of >100 questions (eg, about participants’ well-being, trust in the government, adherence to COVID-19–related measures, the risk of COVID-19 infection, and the understanding of COVID-19 information). Over time, new topics such as willingness to receive COVID-19 vaccinations and vaccine hesitancy were added to the survey. To create an easy-read version, we adapted the national survey in three steps by (1) shortening the survey (ie, we selected only a limited number of relevant topics), (2) reducing the number of response categories, and (3) adjusting the language level. This procedure was based on literature insights [37] and carried out in collaboration with professionals (researchers: n=2, care providers: n=3, and policy makers: n=2 working with people with MID or low literacy skills regarding health-related issues) as well as 2 experiential experts with MID or low literacy skills trained to advise, and experienced in advising, research projects. First, to shorten the survey, we discussed, prioritized, and selected the most relevant topics in light of our research objective to measure mental health, relevance to the target population, and the target population’s cognitive capacity to answer the questions. Second, we reduced the number of response categories, in terms of Likert-scale options [38,39], by verifying the distinctiveness between the response categories and checking the relevance of the categories [40]. During the third and final step, abstract concepts, time references, and the language level were adjusted [41].

The easy-read survey consisted of 40 to 60 questions, with the length depending on the answers given to previous questions. The survey was pilot-tested with people with MID or low literacy skills (n=6) using the think-aloud methodology in cognitive interviews [42]. We used this method to verify the intended constructions of the questions and to assess the language level and the fit of the questions and response categories. Next, a web-based version of the survey was created on a web-based platform (called I Coresearch) designed with and for people with MID [43]. The platform has a clear layout, the possibility to enlarge font size, icons that can be added to response categories, and a speech-to-text and text-to-speech function. Additional pilot tests were carried out in which we observed participants (n=4–7 in each wave) while they were completing the survey to evaluate and improve the usability of the platform. The observation sessions were followed by retrospective interviews. The tests resulted in minor adjustments to the questions, the response categories, and the web-based platform. This procedure resulted in a final short easy-read web-based survey.

These same steps were followed to modify and revise the survey for the second and third rounds. After each survey round, the findings were discussed in 4 to 5 focus groups with either people with MID or low literacy skills or care and support professionals and policy makers concerned with these subgroups. This not only led to a quick dissemination of our findings accompanied by solutions or practical tips to put into practice, but the group discussions also provided input for the subsequent survey rounds in which questions that became less relevant over time (eg, adherence to specific measures and difficulty coping with changes in specific daily activities) were replaced by new questions (eg, about vaccination).

Measures

The easy-read survey consisted of various topics, and we report the measures used for this specific study only (for details, refer to Multimedia Appendix 1).

Demographics and Contextual Factors

Similar to the national survey, an extensive section on demographics was included, such as age, gender, educational level, country of birth, and living situation. Furthermore, contextual factors known to have a potential influence on mental well-being were selected from the national survey and included, such as health status (eg, a rating of physical health and whether the participant had experienced COVID-19 infection), having social contacts, and socioeconomic status (eg, work status and the cessation of main activities because of COVID-19–related restrictions) [24,25]. To fit our target group’s everyday experience, work status included paid work, volunteer work, school, and day care. In addition to the national survey questions, we included receiving professional care because this is an important characteristic describing the support needs of our target population, as well as survey completion methods (alone or with help), about which participants with MID or low literacy skills were asked (refer to Multimedia Appendix 1 for all questions listed in the easy-read questionnaire).

Mental Health

To gain a better understanding of the impact of the COVID-19 pandemic on mental health, a final set of 6 outcome measures regarding mental health were defined. The RIVM survey incorporated a mix of newly developed and existing validated scales or items in its well-being module, including psychological well-being (the 5-item Mental Health Inventory [44]), loneliness (the 6-item De Jong Gierveld Loneliness Scale [45]), life satisfaction, resilience, positive and negative effects experienced due to the COVID-19 pandemic, and emotional response (ie, the extent of worry, stress, or fear people experience) to monitor various aspects of mental health during the COVID-19 pandemic [6]. The most relevant items for our target group and research aim were selected from this set in the first developmental step of item generation. Subsequently, the response format was evaluated. For reasons of uniformity, the response categories...
in this series of questions were all adapted to a 4-point Likert scale. All questions had to be revised in accordance with the language level and understanding of the target group, and we ensured that overlapping concepts were avoided. These steps resulted in the following questions: (1) “Did you feel happy in the last couple of days?” (2) “Did you feel full of energy in the last couple of days?” (3) “Did you worry in the last couple of days?” (4) “Did you feel stressed in the last couple of days?” (5) “Did you feel lonely in the last couple of days?” and (6) “Did you have problems falling asleep in the last couple of days?” The outcomes were measured on a 4-point Likert scale: 1=yes, almost always; 2=yes, often; 3=yes, sometimes; and 4=no, almost never (Multimedia Appendix 1).

Statistical Analyses

Mental health was measured at 3 time points over a 1-year period among people with MID or low literacy skills (referred to as the target panel) as well as among members of the MHS panels. First, we calculated the frequencies and medians of the descriptive and contextual measures at each time point for the target panel and the MHS panels. To assess differences between the panels, in each round, Pearson chi-square tests were conducted for nominal or ordinal variables, and nonparametric t tests (2-tailed) were performed for age. Second, we calculated the differences in frequencies of mental health scores using Pearson chi-square tests between survey rounds within each panel and between panels for each survey round. Third, we analyzed the impact of the group differences on mental health using linear regression analyses, while controlling for gender- and age-related differences. Given the large number of participants in each round and the multitude of comparisons made in analysis, differences and associations were considered statistically significant if P values were <.01 [46]. Statistical analyses were conducted in SPSS (version 25.0; IBM Corp).

Ethical Considerations

The study was reviewed by the medical research ethics committee of Radboud University Medical Center, which ruled that this study did not fall under the Medical Research Involving Human Subjects Act and was therefore exempt from formal ethical review (2020-7033). We conducted the study in accordance with the General Data Protection Regulation and standard operating procedures of our research center.

All participants received the survey after they had been fully informed, in plain language, about the purpose of this study. All participants provided web-based written informed consent regarding participation and the use of their data for this study and for future purposes before filling out each questionnaire. For each survey, 20 vouchers worth €50 (US $53.9) each were raffled among people with MID or low literacy skills as motivation for participation.

Contact information used for the purpose of this raffle or future research was obtained and saved in a separate environment so that survey data could be obtained anonymously. Therefore, matching between surveys and paired within-group analysis over time was not possible.

Results

Participant Characteristics

Our web-based survey and adapted recruitment procedure enabled 1059 persons with MID or low literacy skills to participate (T1: n=412, 38.9%; T2: n=351, 33.1%; and T3: n=296, 28%). Background and contextual characteristics per survey round for the target panel and the MHS panels are presented in Table 1. Over the 3 time periods, 46.6% (138/296) to 53.2% (219/412) of the participants with MID or low literacy skills were women, with median ages ranging from 42 (IQR at T1: 27-57; IQR at T3: 28-54) to 45 (IQR 30-57 at T2) years and >70% (T1: 292/412, 70.9%; T2: 253/351, 72.1% and T3: 221/296, 74.7%) reporting no education or a low educational level. The majority (299/412, 72.6% at T1; 295/351, 84% at T2 and to 253/296, 85.5% at T3) were born in the Netherlands, 49.5% (204/412) to 62.2% (184/296) received professional care, and 22.6% (93/412) to 31.8% (94/296) reported living in a residential setting. Approximately half of the respondents (217/412, 52.7% at T1; 164/351, 46.7% at T2 and 170/296, 57.4% at T3) in each round reported very good or good physical health.

A total of 9305 MHS panel members completed our survey (T1: n=2930, 31.49%; T2: n=3213, 34.53%; and T3: n=3162, 33.98%). On the MHS panels over the 3 survey rounds, 55.85% (1766/3162) to 62.53% (1832/2930) of the participants were women, with median ages ranging from 52 to 62 years, and 12.73% (373/2930 at T1) to 17.08% (540/3162 at T3) had no education or a low educational level (ie, >70% had an intermediate or advanced educational level). The majority (2790/2930, 95.22% at T1; 3073/3213, 95.64% at T2 and 3023/3162, 95.6% at T3) were born in the Netherlands, only 3.28% (96/2930 at T1; 92/3213, 2.86% at T2 and 105/3162, 3.32% at T3) received professional care, and <1% (6/2930, 0.2% at T1; 6/3213, 0.19% at T2 and 3/3162, 0.1% at T3) lived in a residential setting. In addition, 77.12% (2478/3213 at T2 and 2435/3162, 77% at T3) to 79.39% (2326/2930 at T1) reported having very good or good physical health.

Altogether, this suggests that we successfully included a sample that represented our target population (ie, people with MID or low literacy skills). In addition, the characteristics of the MHS panels resemble those of the national sample, which is often used to represent the general Dutch population [9,36].

---

https://publichealth.jmir.org/2024/1/e44827
### Table 1. Distribution of demographic and contextual characteristics (health and socioeconomic status, social contacts, and target group–specific characteristics) per survey round by target panel and municipal health service (MHS) panels.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>T1&lt;sup&gt;a&lt;/sup&gt; target panel (n=412)</th>
<th>T1 MHS panels (n=2930)</th>
<th>T2&lt;sup&gt;a&lt;/sup&gt; target panel (n=351)</th>
<th>T2 MHS panels (n=3113)</th>
<th>T3&lt;sup&gt;a&lt;/sup&gt; target panel (n=296)</th>
<th>T3 MHS panels (n=3162)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (y), median (IQR)</strong></td>
<td>62 (27-77)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>52 (41-64)</td>
<td>45 (30-57)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>60 (45-71)</td>
<td>42 (28-54)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>62 (46-72)</td>
</tr>
<tr>
<td><strong>Gender&lt;sup&gt;f&lt;/sup&gt; (woman), n (%)</strong></td>
<td>219 (53.2)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1832 (62.5)</td>
<td>185 (52.7)&lt;sup&gt;h&lt;/sup&gt;</td>
<td>1865 (58)</td>
<td>138 (46.6)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1766 (55.9)</td>
</tr>
<tr>
<td><strong>Educational level&lt;sup&gt;e&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No to primary education</td>
<td>91 (22.1)</td>
<td>12 (0.4)</td>
<td>63 (17.9)</td>
<td>17 (0.5)</td>
<td>82 (27.7)</td>
<td>25 (0.8)</td>
</tr>
<tr>
<td>Low</td>
<td>201 (48.8)</td>
<td>361 (12.3)</td>
<td>190 (54.1)</td>
<td>477 (14.8)</td>
<td>139 (47)</td>
<td>515 (16.3)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>53 (12.9)</td>
<td>769 (26.2)</td>
<td>64 (18.2)</td>
<td>811 (25.2)</td>
<td>45 (15.2)</td>
<td>790 (25)</td>
</tr>
<tr>
<td>Advanced</td>
<td>27 (6.6)</td>
<td>1730 (59)</td>
<td>8 (2.3)</td>
<td>1848 (57.5)</td>
<td>12 (4.1)</td>
<td>1785 (56.5)</td>
</tr>
<tr>
<td>Other&lt;sup&gt;i&lt;/sup&gt;</td>
<td>39 (9.5)</td>
<td>56 (1.9)</td>
<td>25 (7.1)</td>
<td>59 (1.8)</td>
<td>17 (5.7)</td>
<td>46 (1.5)</td>
</tr>
<tr>
<td><strong>Born in the Netherlands, n (%)</strong></td>
<td>299 (72.6)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>2790 (95.2)</td>
<td>295 (84)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>3073 (95.6)</td>
<td>253 (85.5)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>3023 (95.6)</td>
</tr>
<tr>
<td><strong>Living situation&lt;sup&gt;e&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>101 (24.5)</td>
<td>388 (13.2)</td>
<td>123 (35)</td>
<td>495 (15.4)</td>
<td>88 (29.7)</td>
<td>516 (16.3)</td>
</tr>
<tr>
<td>With family</td>
<td>209 (50.7)</td>
<td>2525 (86.2)</td>
<td>149 (42.5)</td>
<td>2708 (84.4)</td>
<td>112 (37.8)</td>
<td>2636 (83.4)</td>
</tr>
<tr>
<td>In residential setting</td>
<td>93 (22.6)</td>
<td>6 (0.2)</td>
<td>79 (22.5)</td>
<td>6 (0.2)</td>
<td>94 (31.8)</td>
<td>3 (0.1)</td>
</tr>
<tr>
<td><strong>Very good or good physical health, n (%)</strong></td>
<td>364 (14.7)</td>
<td>45 (15.2)</td>
<td>461 (14.3)</td>
<td>61 (17.4)&lt;sup&gt;k&lt;/sup&gt;</td>
<td>461 (14.3)</td>
<td>61 (17.4)&lt;sup&gt;k&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>COVID-19 infection, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Daily activities&lt;sup&gt;m&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paid work</td>
<td>145 (35.2)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1920 (65.5)</td>
<td>133 (37.9)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1628 (50.7)</td>
<td>117 (39.5)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1528 (48.3)</td>
</tr>
<tr>
<td>Volunteer work</td>
<td>92 (22.3)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>748 (25.5)</td>
<td>74 (21.1)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>652 (20.3)</td>
<td>62 (20.9)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>738 (23.3)</td>
</tr>
<tr>
<td>School</td>
<td>100 (24.3)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>154 (5.3)</td>
<td>42 (12)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>87 (2.7)</td>
<td>47 (15.9)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>73 (2.3)</td>
</tr>
<tr>
<td>Day care</td>
<td>121 (29.4)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>43 (1.5)</td>
<td>103 (29.3)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>32 (1)</td>
<td>103 (34.8)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>30 (0.9)</td>
</tr>
<tr>
<td>Other (sport, hobby, care, other)</td>
<td>197 (47.8)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1558 (53.2)</td>
<td>179 (51)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1760 (54.8)</td>
<td>176 (59.5)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1668 (52.8)</td>
</tr>
<tr>
<td>None</td>
<td>32 (7.8)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>351 (12)</td>
<td>19 (5.4)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>379 (11.8)</td>
<td>9 (3)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>373 (11.8)</td>
</tr>
<tr>
<td><strong>Change in daily activity&lt;sup&gt;v&lt;/sup&gt;, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced or stopped</td>
<td>125 (30.3)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>385 (13.1)</td>
<td>98 (27.9)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>538 (16.7)</td>
<td>67 (22.6)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>229 (7.2)</td>
</tr>
<tr>
<td>Nothing to do (bored)&lt;sup&gt;j&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>18 (5.1)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>54 (1.7)</td>
<td>19 (6.4)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>49 (1.5)</td>
</tr>
<tr>
<td><strong>Social contacts, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No one I can talk to</td>
<td>42 (10.2)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>188 (6.4)</td>
<td>45 (12.8)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>134 (4.2)</td>
<td>20 (6.8)&lt;sup&gt;w&lt;/sup&gt;</td>
<td>135 (4.3)</td>
</tr>
<tr>
<td>No one who can help me</td>
<td>59 (14.3)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>103 (3.5)</td>
<td>34 (9.7)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>88 (2.7)</td>
<td>24 (8.1)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>79 (2.5)</td>
</tr>
<tr>
<td>Receives professional care, n (%)</td>
<td>204 (49.5)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>96 (3.3)</td>
<td>216 (61.5)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>92 (2.9)</td>
<td>184 (62.2)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>105 (3.3)</td>
</tr>
<tr>
<td><strong>Survey completion, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>187 (45.4)</td>
<td>—</td>
<td>202 (57.5)</td>
<td>—</td>
<td>169 (57.1)</td>
<td>—</td>
</tr>
<tr>
<td>With help</td>
<td>223 (54.2)</td>
<td>—</td>
<td>147 (41.9)</td>
<td>—</td>
<td>127 (42.9)</td>
<td>—</td>
</tr>
</tbody>
</table>

---

*aCategory totals do not always add up to 100% because some categories (I don’t know and I don’t want to answer) and item nonresponse are not shown. Percentages are based on presented variable totals per category.*

*bT1: first survey (November to December 2020).*

*cT2: second survey (April to May 2021).*

*dT3: third survey (September to October 2021).*

*eValue for the target panel is significantly different from that for the regional panel (*P*<.001). Italicized values emphasize significance.*

https://publichealth.jmir.org/2024/1/e44827
f.<1% indicated their gender as "other."

Value for the target panel is significantly different from that for the regional panel ($P=.001$). Italicized values emphasize significance.

Value for the target panel is not significantly different from that for the regional panel ($P=.15$).

The answer category “other” is chosen by respondents when it does not fit any of the provided options. This may be because they do not know their educational level, do not recognize their education category from the option list, or they were educated in a country other than the Netherlands.

This question was added to the survey from T2.

Value for the target panel is not significantly different from that for the regional panel ($P=.12$).

Value for the target panel is not significantly different from that for the regional panel ($P=.72$).

Respondents could provide multiple answers; the category total can therefore add up to more than 100%.

Value for the target panel is significantly different from that for the regional panel ($P=.005$). Italicized values emphasize significance.

Value for the target panel is not significantly different from that for the regional panel ($P=.16$).

Value for the target panel is not significantly different from that for the regional panel ($P=.73$).

Value for the target panel is not significantly different from that for the regional panel ($P=.38$).

Value for the target panel is not significantly different from that for the regional panel ($P=.04$).

Value for the target panel is not significantly different from that for the regional panel ($P=.18$).

Value for the target panel is not significantly different from that of the regional panel ($P=.03$).

Value for the target panel is not significantly different from that for the regional panel ($P=.01$).

This variable is constructed concerning the daily activities of paid work, volunteer work, and day care.

Value for the target panel is not significantly different from that for the regional panel ($P=.02$).

Mental Health

The analyses of the distributions of the frequencies of mental health scores within the target panel show no differences between T1 and T2. There are significant differences between T1 and T3 regarding feeling happy, feeling energized, feeling stressed, worry, feeling lonely, and sleeping problems among people with MID or low literacy skills (Figure 2) and between T2 and T3 for these aspects, except for worry and feeling lonely; the percentage of people reporting positive feelings often or almost always increased, and the percentage of people reporting negative feelings often or almost always decreased over time. There were no differences observed regarding sleeping problems.

Regarding the MHS panels, there were significant differences between each survey round for feeling happy, feeling energized, and feeling stressed. For worry and feeling lonely, significant differences were observed only between T3 and the 2 previous rounds. The direction of the differences is similar to that observed in the target panels. Similar to the target panels, the MHS panels did not report differences regarding sleeping problems.

The analyses between the different panels within the survey rounds show that the percentage of participants in the target panel reporting negative feelings on mental health outcomes was significantly higher compared with the members of the MHS panels, especially within T1 and T2 (Figure 2); for example, looking at the 6 outcome measures within T1, the percentages of people who reported more negative feelings range from 20.8% (85/408) feeling lonely often or almost always to 58.3% (238/408) feeling happy almost never or only sometimes. The MHS panels show a different and more positive distribution on all outcome measures. The percentages of people on the MHS panels who reported more negative feelings range from 5.5% (160/2904) feeling lonely often or almost always to 32.2% (941/2918) feeling happy almost never or only sometimes. Figure 2 presents more details on the distribution of all mental health outcomes for the target panel and the MHS panels.
Figure 2. Distribution of mental health outcomes for participants in the target panels and members of the municipal health service (MHS) panels at 3 different time periods. T1: first survey; T2: second survey; T3: third survey. a) Difference between groups within each round ($P<.001$), b) Difference between rounds within target group (T1 vs T3: $P<.001$), c) Difference between rounds within panel (T1 vs T2, T2 vs T3, and T1 vs T3: $P<.001$), d) Difference between rounds within target group (T2 vs T3, and T1 vs T3: $P<.001$), e) Difference between rounds within target group (T1 vs T3: $P=.005$), f) Difference between rounds within panel (T1 vs T2, and T1 vs T3: $P<.001$ and T2 vs T3 $P=.01$).

Impact of Literacy Skills on Mental Health

Regression analyses adjusted for age and gender by using them as covariates show that the differences found between the 2 panels exist for almost all mental health outcomes in each survey round, except feeling stressed in T3 ($P=.07$). In addition, the differences between the panels during T1 (age- and gender-adjusted $\beta$ ranging from $-0.376$ to $0.525$) are larger than...
those observed in T3 (age- and gender-adjusted β ranging from −0.257 to 0.509), except for sleeping problems. Table 2 presents details for all time periods and outcome measures.

Table 2. Results for each survey round with panel as independent variable, mental health measures as dependent variables, and gender and age as covariates.

<table>
<thead>
<tr>
<th></th>
<th>T1a</th>
<th>T2b</th>
<th>T3c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feeling happy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel</td>
<td>−0.386b (−0.464 to −0.287)</td>
<td>−0.385 (−0.481 to −0.288)</td>
<td>−0.257 (−0.362 to −0.152)</td>
</tr>
<tr>
<td>Gender</td>
<td>−0.037 (−0.096 to 0.022)</td>
<td>−0.092 (−0.151 to −0.034)</td>
<td>−0.065 (−0.122 to −0.07)</td>
</tr>
<tr>
<td>Age</td>
<td>0.000 (−0.002 to 0.002)</td>
<td>0.000 (−0.002 to 0.002)</td>
<td>0.000 (−0.002 to 0.002)</td>
</tr>
<tr>
<td>Feeling energized</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel</td>
<td>−0.156 (−0.249 to −0.063)</td>
<td>−0.243 (−0.340 to −0.145)</td>
<td>−0.143 (−0.249 to −0.037)</td>
</tr>
<tr>
<td>Gender</td>
<td>−0.039 (−0.102 to −0.023)</td>
<td>−0.080 (−0.139 to −0.021)</td>
<td>−0.058 (−0.117 to 0.000)</td>
</tr>
<tr>
<td>Age</td>
<td>0.004 (0.002 to 0.006)</td>
<td>0.003 (0.001 to 0.005)</td>
<td>0.004 (0.002 to 0.006)</td>
</tr>
<tr>
<td>Worry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel</td>
<td>0.328 (0.246 to 0.409)</td>
<td>0.324 (0.255 to 0.429)</td>
<td>0.227 (0.135 to 0.319)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.158 (0.104 to 0.213)</td>
<td>0.217 (0.165 to 0.270)</td>
<td>0.181 (0.131 to 0.232)</td>
</tr>
<tr>
<td>Age</td>
<td>0.000 (−0.001 to 0.002)</td>
<td>−0.002 (−0.004 to 0.000)</td>
<td>−0.002 (−0.003 to 0.000)</td>
</tr>
<tr>
<td>Feeling stressed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel</td>
<td>0.274 (0.195 to 0.353)</td>
<td>0.199 (0.117 to 0.281)</td>
<td>0.081 (−0.168 to 0.006)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.158 (0.105 to 0.211)</td>
<td>0.189 (0.140 to 0.239)</td>
<td>0.195 (0.147 to 0.243)</td>
</tr>
<tr>
<td>Age</td>
<td>−0.010 (−0.012 to −0.008)</td>
<td>−0.012 (−0.014 to −0.011)</td>
<td>−0.013 (−0.014 to −0.011)</td>
</tr>
<tr>
<td>Sleeping problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel</td>
<td>0.525 (0.439 to 0.611)</td>
<td>0.518 (0.427 to 0.609)</td>
<td>0.509 (0.411 to 0.606)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.220 (0.163 to 0.278)</td>
<td>0.296 (0.241 to 0.351)</td>
<td>0.240 (0.187 to 0.294)</td>
</tr>
<tr>
<td>Age</td>
<td>0.003 (0.001 to 0.005)</td>
<td>0.000 (−0.002 to 0.002)</td>
<td>0.002 (0.000 to 0.003)</td>
</tr>
<tr>
<td>Feeling lonely</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel</td>
<td>0.479 (0.409 to 0.549)</td>
<td>0.461 (0.387 to 0.535)</td>
<td>0.297 (0.229 to 0.365)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.095 (0.049 to 0.142)</td>
<td>0.161 (0.116 to 0.206)</td>
<td>0.071 (0.034 to 0.109)</td>
</tr>
<tr>
<td>Age</td>
<td>−0.001 (−0.003 to 0.000)</td>
<td>−0.002 (−0.004 to −0.001)</td>
<td>−0.002 (−0.003 to 0.000)</td>
</tr>
</tbody>
</table>

aT1: first survey (November to December 2020).
bT2: second survey (April to May 2021).
cT3: third survey (September to October 2021).
dItalicized values indicate significant regression results, with P<.01 level.

Discussion

Principal Findings

This is the first study to monitor the mental health and well-being of people with MID or low literacy skills and a general population sample over the course of 1 year during the COVID-19 pandemic. With our adapted web-based survey co-designed with representatives from our target population, we were able to reach subgroups that are usually underrepresented in surveys. Our study showed that feelings of happiness, energy, worry, stress, and loneliness improved in both populations over the course of the pandemic. However, the COVID-19 pandemic and related restrictions had a much bigger impact on the mental health of people with MID or low literacy skills than on that of the general population.

In general, our findings show that, during the second lockdown in the Netherlands (ie, at the time of the first survey round), people with MID or low literacy skills as well as the general population sample reported poorer mental well-being than 1 year later when all restrictions were lifted, and the COVID-19 infections became less severe (ie, at the time of the third survey round). These findings are in line with research on people in vulnerable positions [1] as well as the general population who experienced fewer negative feelings over the course of the COVID-19 pandemic [5]. Previous literature has shown that the impact on mental health and well-being is correlated to the...
stringency of disease control measures [25, 47]; for example, the closure of social care services, workplaces, and day care activities negatively influenced daily structure and social interactions [29-31, 48, 49], thereby increasing stress and anxiety [50], and quarantines and social isolation were found to have an effect on loneliness, fear, and boredom [27, 28, 51]. When these measures were relaxed, there was a partial improvement in mental health [47]. Although our study was not designed to prove any causation between the stringency of disease control measures and mental health impact, our findings show a similar pattern of decreasing worries, stress, and loneliness whereas feelings of happiness and energy increased over time. Notably, this was also the case in our general population sample [1, 3, 5, 14, 25, 26]; however, the pandemic disproportionately impacted people with MID or low literacy skills, who reported more negative mental health outcomes in all survey rounds.

Qualitative studies among people with MID show that long-term social restrictions in particular had an extensive impact on their daily life by limiting social connections and work activities [27, 29]. Our target population reported these limitations in daytime activities to a greater extent than the general population sample. Interviews by Voermans et al [29] provide more in-depth assessment of the consequences of these limitations for people with MID, showing a major impact in terms of social isolation, difficulties coping with negative thoughts, struggles with autonomy in society, stigmatization, a lack of routine and purpose, boredom, and lower self-worth. As awareness is raised about the significant value of meaningful social contacts and daytime activities, professionals and policy makers should provide tailored policies that consider both health risks and the risks of social isolation. Societal participation initiatives should be organized and sustained for people with MID or low literacy skills, both during and outside of a pandemic.

Besides the disruptive impact of disease control measures on the target population’s daily routines and social contacts, the high levels of confusion and uncertainty that resulted from the rapidly changing measures as well as fear and loss of control may have played a role in their reduced mental health [32] in periods of both stringent measures and relaxation of control measures [14]. In addition, people with lower health literacy skills are known to have less resilience, which affects their feelings of anxiety, stress, or worry [52, 53], thereby putting them at greater risk of mental health problems. Ongoing support should be provided to enhance resources of resilience and coping strategies in people with MID or low literacy skills through either formal or informal caregivers.

Our findings highlight the need to prioritize the mental health consequences of the pandemic and the disease control measures for people with MID or low literacy skills [1, 54, 55]. The majority of our sample received support from formal and informal caregivers, who are an important source of support. Studies have shown detrimental effects on the mental health of these caregivers as well [28, 56]. Therefore, we suggest tailoring generic disease control measures to the specific situations of groups considered vulnerable and their support system, instead of widely implementing measures such as social distancing, visiting restrictions, and closure of schools or day care facilities (e.g., by developing strategies to maintain social inclusion during pandemic challenges through a combination of supportive carers, assisted digital communication technologies, and safe social activities) [57, 58]. Hence, engaging groups considered vulnerable and their support system in policy making and decision-making is essential in the tailoring process [59].

This study underscores the relevance of including people with MID or low literacy skills in health research and therefore endorses current calls to action in practice and science [60]. Health information systems are crucial for providing data for policy making and decision-making, but the underrepresentation of people with MID or low literacy skills in health data may lead to biased policy decisions, with adverse and detrimental effects on existing health disparities [13, 33]. Previous research has suggested that, to reduce disparities and guide policy, researchers should evaluate how health outcomes are distributed among specific demographic groups and compare these distributions with those of the overall population [61], as was done in our study. Collecting information about people with MID or low literacy skills should become routine in demographic and public health data collection. We have shown that, by co-designing an adapted survey and using an accessible web-based platform and specific recruitment procedures, it is possible to collect information among people with MID or low literacy skills, even during lockdown periods; for example, the sample characteristics showed that participants with MID or low literacy skills differed from the general population in educational level, country of birth, and daily activities. With our adjusted approach to data collection, we were quickly able to obtain relevant information about people with MID or low literacy skills and disseminate our findings and recommendations, thereby facilitating policy makers to guide disease control measures and health promotion activities that address the immediate as well as longer-term health needs of people with MID or low literacy skills or other populations considered vulnerable.

**Limitations of the Study**

Executing a repeated cross-sectional survey among people with MID or low literacy skills during a pandemic is fraught with challenges. Therefore, our study has some limitations. First, we did not collect longitudinal data because we wanted to lower the threshold for participation by choosing an anonymous design. In addition, people were not obliged to complete all 3 surveys. As a result, it was impossible to track individual participants over time. Second, inevitably, the validated questions had to be revised to include people with MID or low literacy skills in our survey. However, we tested the questions in cognitive interviews, and the project team worked as inclusively as possible together with the target population to create a valid survey to obtain reliable data. More than half of the participants over all survey rounds (558/1059, 52.69%) were able to complete the web-based survey themselves, and support was arranged for the remaining group of respondents (497/1059, 46.93%). This should encourage future researchers to consider easy-read web-based surveys among people with MID or low literacy skills as long as target group representatives are closely involved in designing and testing these surveys. Third, our study started after the onset of the COVID-19 pandemic and lacks a baseline measurement of mental health before the pandemic.
Therefore, it remains inconclusive as to whether people with MID or low literacy skills experienced greater mental health problems during the COVID-19 pandemic than before the pandemic. Fourth, we relied mostly on organizations in our network (eg, health care organizations and public libraries with special literacy programs) to contact and recruit people with MID or low literacy skills. Besides the possible sampling bias that this may have caused, we could not track how many people were approached to take part in the survey. Therefore, we were unable to report information about response rates. Fifth, because the survey was conducted on the web, those without access to the internet or sufficient digital literacy skills may have been excluded. Sixth and last, there was also a bias in our general population sample. However, although the sample was not representative of the Dutch general population in terms of age, gender, and educational level, it allowed us to contextualize our findings and gain a deeper understanding of the challenges faced by people with MID or low literacy skills compared with the general population sample, while controlling for differences in gender and age. We were able to do so because we used the same easy-read questionnaire in both groups to prevent survey bias. An open-ended question about the experience of members of the MHS panels with this type of questionnaire at T3 revealed that the majority of participants (1829/2337, 78.26%) appreciated this approach, given that a broader population was enabled to participate. Although this may indicate that easy-read questionnaires can be used for broader purposes and other populations, rather than being aimed specifically at people with low literacy skills alone, this single question does not provide sufficient information regarding a broad survey approach, and more research is required.

Conclusions

In conclusion, our study enabled insight into the impact of the COVID-19 pandemic and related control measures on the mental health of people with MID or low literacy skills. General disease control measures for the entire Dutch population had a more negative impact on people with MID or low literacy skills than on the general population. Although mental health improved over the course of the pandemic in both populations as the COVID-19–related restrictions were gradually lifted, and the severity of the disease decreased over time, the disproportional effect remained. Professionals should be aware of this and pay attention to the needs of people with MID or low literacy skills in research, practice, and policy by tailoring measures that consider physical, social, and mental health effects and providing support to overcome such effects.

This study underscores the relevance of including people with MID or low literacy skills in public health research because they are often overlooked in regular health data. An accessible and structural web-based monitor for people with MID or low literacy skills enabled us to do so and provides better knowledge for care providers and policy makers to react to unexpected events such as a pandemic. To prevent existing health disparities from increasing, greater account should be taken of the impact of control measures on people who are relatively more vulnerable.

Acknowledgments

The authors offer grateful thanks to their coresearchers Anneke van der Cruijsen and Paméla Melkert who were involved during all steps of the study as well as to all experts by experience who participated in the development, testing, or evaluation of the web-based survey and supported survey completion or the discussion of the survey results. The authors thank all organizations that participated in the recruitment of participants and, in particular, MEE Gelderse Poort, GGD Gelderland-Zuid, GGD Gelderland-Midden, and Pharos for their valuable contributions and cooperation during the execution of the study. This study was supported by a grant from the Netherlands Organization for Health Research and Development (ZonMw; grant 1043002201000). The funder had no role in study design, data collection, data analysis, data interpretation, the writing of the manuscript, and the decision to publish.

Data Availability

The data sets generated and analyzed during this study are available in the Radboud Data Repository.

Authors’ Contributions

MCJKL, FR, and GAJFK participated in the study conception and design. AM and MCJKL conducted the literature search and coordinated survey development and participant recruitment. KEB, AM, and MCJKL performed survey development, AM was involved in survey programming, and KEB facilitated the implementation of the survey on the web-based platform. AM, FR, and MCJKL contributed to the acquisition of data. AM accessed and verified the data; and KEB, AM, and MCJKL analyzed the data and plotted the tables and figures. AM and MCJKL drafted the manuscript. All authors participated in interpreting the data and study findings, critically reviewing and contributing to the revision of the manuscript, and approving the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
References


35. Steffie. URL: https://www.steffie.nl/ [accessed 2024-04-29]


43. Ik onderzoek mee. URL: https://www.ikonderzoekmee.nl/ [accessed 2024-04-29]


54. Ik onderzoek mee. URL: https://www.ikonderzoekmee.nl/ [accessed 2024-04-29]


Predicting Efficacies of Fractional Doses of Vaccines by Using Neutralizing Antibody Levels: Systematic Review and Meta-Analysis

Zhanwei Du¹²⁺, PhD; Caifen Liu¹²⁺, MSc; Yuan Bai¹², PhD; Lin Wang³, PhD; Wey Wen Lim¹, PhD; Eric H Y Lau¹², PhD; Benjamin J Cowling¹², PhD

¹ these authors contributed equally

Corresponding Author:
Zhanwei Du, PhD

Abstract

Background: With the emergence of SARS-CoV-2 variants that have eluded immunity from vaccines and prior infections, vaccine shortages and vaccine effectiveness pose unprecedented challenges for governments in expanding booster vaccination programs. The fractionation of vaccine doses might be an effective strategy for helping society to face these challenges, as fractional doses may have efficacies comparable with those of the standard doses.

Objective: This study aims to investigate the relationship between vaccine immunogenicity and protection and to project efficacies of fractional doses of vaccines for COVID-19 by using neutralizing antibody levels.

Methods: In this study, we analyzed the relationship between in vitro neutralization levels and the observed efficacies against both asymptomatic infection and symptomatic infection, using data from 13 studies of 10 COVID-19 vaccines and from convalescent cohorts. We further projected efficacies for fractional doses, using neutralization as an intermediate variable, based on immunogenicity data from 51 studies included in our systematic review.

Results: In comparisons with the convalescent level, vaccine efficacy against asymptomatic infection and symptomatic infection increased from 8.8% (95% CI 1.4%-16.1%) to 71.8% (95% CI 63%-80.7%) and from 33.6% (95% CI 23.6%-43.6%) to 98.6% (95% CI 97.6%-99.7%), respectively, as the mean neutralization level increased from 0.1 to 10 folds of the convalescent level. Additionally, mRNA vaccines provided the strongest protection, which decreased slowly for fractional dosing with dosages between 50% and 100% of the standard dose. We also observed that although vaccine efficacy increased with the mean neutralization level, the rate of this increase was slower for vaccine efficacy against asymptomatic infection than for vaccine efficacy against symptomatic infection.

Conclusions: Our results are consistent with studies on immune protection from SARS-CoV-2 infection. Based on our study, we expect that fractional-dose vaccination could provide partial immunity against SARS-CoV-2 and its variants. Our findings provide a theoretical basis for the efficacy of fractional-dose vaccines, serving as reference evidence for implementing fractional dosing vaccine policies in areas facing vaccine shortages and thereby mitigating disease burden. Fractional-dose vaccination could be a viable vaccination strategy comparable to full-dose vaccination and deserves further exploration.

(JMIR Public Health Surveill 2024;10:e49812) doi:10.2196/49812

KEYWORDS
COVID-19; SARS-CoV-2; dose fractionation; neutralizing antibody level; vaccination; review; vaccine

Introduction

COVID-19 continues to threaten fragile health care and socioeconomic systems, exacting a devastating human and economic toll around the world. The primary means of COVID-19 control are the widespread implementation of vaccination and the preservation of public health and social measures. Worldwide, 123 vaccine candidates have been tested in humans as of August 31, 2022, with 52 in the final phases of clinical trials [1]. Although 68% of the world population has received at least one dose of the COVID-19 vaccine as of October 1, 2022 [2], global vaccine shortages and inequities persist [3], with only 18.55% and 1.27% of people in low-income countries being fully vaccinated or receiving booster doses, respectively [2].
As of October 2022, the Omicron subvariant BA.5 has displaced BA.2 as the predominant strain of SARS-CoV-2 in countries around the world. The timely achievement of administering the optimal threshold of doses to the population was critical to preventing a new mutation of the virus, thereby averting accelerated transmission dynamics in society and mitigating consequential socioeconomic issues [4-6]. Although vaccine effectiveness is expected to wane quickly, a booster shot can restore the protection against infection by both Omicron subvariants to 30% - 60% and increase protection against severe disease from a high level to a very high level [7]. A highly vaccinated population is still not enough to combat Omicron’s spread due to immune escape and the waning of vaccine-derived immunity against the ancestral SARS-CoV-2 strain [8]. Countries should time the ramp-up of their booster doses to account for the risks of waning vaccine effectiveness against infection, disease, and death. Mass vaccination with fractional doses of COVID-19 vaccines to boost immunity in a vaccinated population could be a cost-effective strategy for mitigating the public health costs of resurgences caused by vaccine-evasive variants [9].

Given accelerating vaccination, expanding booster programs, and concerns about vaccine safety, the fractional dosing of vaccines could help by providing partial protection for a significant number of people [10]. In the clinical trial of the mRNA-1273 vaccine, 2 fractional doses (half of the full dose) gave geometric mean plaque reduction neutralization test (PRNT) titers that, after 2 weeks, were comparable with those of 2 standard doses [11]. Further, the model-predicted efficacy of the half dose and the measured efficacy of the standard dose were nearly 95% for symptomatic disease [12]. Fractional-dose vaccination has also successfully addressed vaccine shortages in outbreak events [13]. For example, Angola and the Democratic Republic of the Congo adopted fractional doses (one-fifth of the standard dose) of the 17DD yellow fever vaccine to accelerate vaccine rollout during their 2016 yellow fever outbreaks and finally won the war against yellow fever [14,15]. During the mpox (monkeypox) outbreaks in 2022, the US Food and Drug Administration (FDA) issued an emergency use authorization on August 9, 2022, for the JYNNEOS vaccine to prevent mpox infections by administering intradermal injections with a lower volume (one-fifth) to adults aged ≥18 years [16]. During the COVID-19 pandemic period, fractional dosing for young people was approved in some countries, considering that young adults have a robust immune response to vaccines, and lower doses may also be linked with fewer side effects. For example, the FDA approved the use of one-third of the standard doses of the Pfizer-BioNTech COVID-19 vaccine (10 μg) for 5- to 11-year-old children, and one-tenth doses (3 μg) are used for children aged <5 years [17-19].

Fractional-dose vaccination could be an effective strategy for mitigating the epidemic risks of COVID-19 (eg, it allows for the deployment of more vaccines to reach more individuals in settings with limited health care budgets [12,20]) and reducing its disease burden, and this strategy’s effectiveness may be comparable to that of the full-dose strategy [21,22]. However, vaccine efficacy has mainly been tested for standard doses (ie, via clinical trials) [23] rather than for fractional doses, of which the efficacies are measured via phase 1 and 2 clinical trials. These trials involve various vaccines, with immune response being measured based on neutralizing antibody titers. Więcek et al [12] derived the efficacy of fractional doses against symptomatic infection by using its relationship with neutralizing antibody titers in standard doses [23]. However, other clinical outcomes of COVID-19 (eg, asymptomatic infection) are related to measured immunity and how vaccine efficacies can be predicted by using dose fractions based on corresponding immunity levels remain unclear. In this study, we first investigated the relationship between neutralizing antibody titers and vaccine efficacies against SARS-CoV-2 infection. We then predicted vaccine efficacies for fractional doses by using real-world data obtained through a systematic review of neutralizing antibody levels induced by fractional doses of vaccines.

**Methods**

**Sample and Data**

To collect the neutralization data of COVID-19 vaccines that were reported in phase 1 and 2 clinical trial studies, we performed a systematic review of peer-reviewed studies in PubMed on March 15, 2022. We searched for studies in PubMed by using a combination of the following search terms, with no restriction on publication language: (1) “COVID-19” OR “SARS-CoV-2” OR “2019-nCoV” OR “coronavirus”; (2) “vaccin*”; (3) “fractiona*” OR “dose”; and (4) “efficacy” OR “effectiveness” OR “neutralizing antibody” OR “neutralising antibody” OR “neutralization titer” OR “neutralization level” OR “antibody titer” OR “immune response” OR “immune protection” OR “immunogenicity” OR “reactogenicity” OR “safety” OR “adverse event” OR “adverse reaction” OR “adverse effect”; the final search term included search terms 1 to 4, which were combined by using the Boolean operator AND. The searched studies were published between January 1, 2020, and March 15, 2022. The inclusion and exclusion criteria are summarized in **Table 1**. Studies were excluded if they were duplicate publications, preliminary animal studies, preprints, reviews, or commentaries. We reported studies in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.
We performed a systematic review of peer-reviewed studies that reported on the neutralization of COVID-19 vaccines in phase 1 and 2 clinical trials and were published in PubMed between January 1, 2020, and March 15, 2022. We excluded studies, based on the full-text assessment, if they (1) lacked original neutralizing antibody data; (2) reported vaccines that were no longer being processed; or (3) reported vaccines for which standard dosages were not determined.

### Measures of Variables
We calculated the dose fraction for each sample as the fraction of the dose tested in the group divided by the corresponding standard dose. The mean neutralization level was defined as the ratio of the vaccine-induced neutralizing antibody level to the number of neutralizing antibodies in convalescent sera (ie, those measured in the same study). By doing so, we standardized the neutralizing antibody level and enabled the comparison of neutralization titers across studies that used different assays. The $\log_{10}$ transformation of the mean neutralization level was assumed to be normally distributed [23]. This $\log_{10}$-transformed mean neutralization level was the key variable in the models and data analyses reported in the Results section.

### Results
We identified 2811 studies, through the electronic search on PubMed, that were published between January 1, 2020, and March 15, 2022. A total of 2777 studies were left after excluding duplicates. After 2694 studies were excluded based on the title and abstract screening, we retrieved 83 studies eligible for the full-text screening. After we excluded 32 studies based on the full-text screening, 51 studies met the inclusion criteria and were included in the systematic review (Figure 1, Multimedia Appendix 1, and Figure S1 in Multimedia Appendix 3).
Alongside the published studies included in the systematic review, we collected 13 records of the efficacy of 10 vaccines (eg, BNT162b2, ChAdOx1 nCoV-19, and CoronaVac) against asymptomatic and symptomatic infections for 11 COVID-19 variants (eg, Alpha, Beta, and Delta) from PubMed (Multimedia Appendix 2). Informed by the information on standard-dose vaccines, the nonlinear relationship between the standardized mean neutralization level and vaccine efficacies (which were stratified into 4 dosage groups) was fitted by using a logistic model (Figure S3 in Multimedia Appendix 3); the parameter estimates and root mean square deviations of the asymptomatic and symptomatic infection models are shown in Table 2. The slope parameter \( k \) for the asymptomatic infection model was estimated to be 1.84 and was smaller than that for the symptomatic infection model (3.10), reflecting that although vaccine efficacy increased with the mean neutralization level, the rate of this increase was slower for vaccine efficacy against asymptomatic infection than for vaccine efficacy against symptomatic infection (Figure 2). We estimated the mean neutralization level for 50% protection (10n50) against asymptomatic and symptomatic infections to be 262% (95% CI 190% - 361%) of the mean convalescent level and 20% (95% CI 14% - 28%) of the mean convalescent level, respectively. We predicted the vaccine efficacy of fractional doses against infection outcomes based on the established model (Figure 2). As the mean neutralization level increased from 0.1 to 10 folds
of the convalescent level, the predicted vaccine efficacy against asymptomatic and symptomatic infections increased from 8.8% (95% CI 1.4%-16.1%) to 71.8% (95% CI 63%-80.7%) and from 33.6% (95% CI 23.6%-43.6%) to 98.6% (95% CI 97.6%-99.7%), respectively (Figure 2). mRNA vaccines provided the strongest protection, which decreased slowly for fractional dosing with dosages between 50% and 100% of the standard dose.

Table. Model parameter estimates for the logistic function of vaccine efficacy against asymptomatic and symptomatic infections. We fit the nonlinear relationship between the standardized mean neutralization level and vaccine efficacies (which were stratified into 4 dosage groups) by using a logistic model that was informed by the information on standard-dose COVID-19 vaccines (Figure S3 in Multimedia Appendix 3). The parameter estimates and root mean square deviations (RMSDs) of the asymptomatic and symptomatic infection models are shown in this table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate (95% CI)</th>
<th>RMSD, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptomatic infection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope (k)</td>
<td>1.84 (1.15 - 2.94)</td>
<td></td>
</tr>
<tr>
<td>50% protection (10n50)</td>
<td>2.62 (1.90 - 3.61)</td>
<td>2.43</td>
</tr>
<tr>
<td>Symptomatic infection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope (k)</td>
<td>3.10 (2.19 - 4.38)</td>
<td>0.16</td>
</tr>
<tr>
<td>50% protection (10n50)</td>
<td>0.20 (0.14 - 0.28)</td>
<td></td>
</tr>
</tbody>
</table>

*aParameter estimates and CIs for the logistic function (Equation 1 in Multimedia Appendix 3) describe the protective efficacy experienced by individuals with a given log_{10}-transformed neutralization level.

*bThe 50% protective neutralization values reported here have been converted from the log_{10} scale back to the linear scale (as reported in the main text).
Figure 2. Estimating vaccine efficacy by using neutralizing antibody levels. Estimated vaccine efficacy against (A) asymptomatic infection and (B) symptomatic infection. The x-axis and y-axis denote the reported mean neutralization levels from phase 1 and 2 trials and the predicted protective efficacies for vaccines or the convalescent cohort, respectively. The gray solid line and shading indicate the best fit and the 95% CIs of the logistic model, respectively. Each dot indicates the predicted efficacy for vaccines (eg, mRNA, protein subunit, viral vector, inactivated, and DNA) or convalescent individuals.

The results of fitting the generalized additive model on dose fractions and mean neutralization levels are presented in Table S1 in Multimedia Appendix 3. The estimated coefficients of the \( \log_{10} \) of the dose fractions were 0.746 (\( P = .02 \)), 0.803 (\( P = .002 \)), and 0.543 (\( P = .02 \)) for mRNA, protein subunit, and nonreplicating viral vector vaccines, respectively. The fitted models showed that the efficacies of mRNA (\( P = .02 \)) and protein subunit (\( P = .002 \)) vaccines significantly increased as the dose fraction increased (Figure S4 in Multimedia Appendix 3). The predicted efficacies of dose fractions for different vaccine types are presented in Figures 2-4. For example, the mRNA vaccine efficacy against asymptomatic and symptomatic infections increased from 12.1% (95% CI 7.2%-19.7%) to 74.8% (95% CI 64.6%-82.9%) and from 52.9% (95% CI 29.8%-74.8%) to 99.5% (95% CI 98.9%-99.8%), respectively, for dose fractions ranging from 0.03 to 5. The protein subunit vaccines had efficacies of 5.2% (95% CI 2.1%-12.2%) to 59.6% (95% CI 46.5%-71.4%) against asymptomatic infection and 19.4% (95% CI 4.8%-53.2%) to 98.4% (95% CI 96.2%-99.3%) against symptomatic infection for dose fractions in the study range of 0.03 to 5. The nonreplicating viral vector vaccines had efficacies of 6% (95% CI 3.3%-10.9%) to 37.1% (95% CI 31.2%-43.4%).
against asymptomatic infection and 23.6% (95% CI 9.5%-47.7%) to 92.9% (95% CI 89.3%-95.3%) against symptomatic infection for dose fractions in the study range of 0.03 to 5.

**Figure 3.** Estimating vaccine efficacy against asymptomatic infection by using neutralizing antibody levels of vaccines. The x-axis and y-axis denote the fraction of dose and the estimated protective efficacy for 3 types of vaccines (e.g., mRNA, protein subunit, and NR viral vector), respectively. The solid line and shading indicate the best fit and the 95% CIs of the logistic model, respectively. NR: nonreplicating.
Figure 4. Estimating vaccine efficacy against symptomatic infection by using neutralizing antibody levels of vaccines. The x-axis and y-axis denote the fraction of dose and the estimated protective efficacy for 3 types of vaccines (eg, mRNA, protein subunit, and NR viral vector), respectively. The solid line and shading indicate the best fit and the 95% CIs of the logistic model, respectively. NR: nonreplicating.

Discussion

This review revealed the immunogenicity, efficacy, and safety of fractional doses as of December 2021, indicating that they would be safe and comparably effective to the standard doses. The included articles predominantly focused on 10 vaccines (eg, BNT162b2, ChAdOx1 nCoV-19, and CoronaVac) for 11 COVID-19 variants (eg, Alpha, Beta, and Delta). Informed by available data from published studies of immune responses for both standard doses and fractional doses, we predicted vaccine efficacies by using neutralizing antibody levels for both standard-dose vaccines and fractional-dose vaccines.

We investigated the relationship between neutralizing antibody titers and the vaccine efficacies against infection for standard doses and derived vaccine efficacies for fractional doses. This study contributes further evidence supporting the use of neutralization titers as significant predictors of vaccine efficacy against asymptomatic infections, symptomatic infections, and severe diseases, in line with previous studies [23,24]. We additionally explored the relationship between dose fractions and vaccine efficacy, using neutralization titers as intermediate variables and accounting for vaccine type.

There would be public health and economic advantages to using a dose-sparing strategy to increase vaccine supply and
vaccination coverage around the world [12,20]. According to the World Health Organization, fractional dosing strategies have the potential to save lives, but only after a thorough review of immunogenicity data [25]. Using untested fractionated vaccine doses in a low-resource setting should also be considered, that is, in cases where higher-resource settings have access to the full vaccine doses and ethics and politics have been taken into account [20]. Further, fractional-dose vaccination could help to optimize the availability of vaccine doses during vaccine shortages, and even in cases where the vaccine supply is increasing, this strategy would still have a high value for vulnerable individuals, as lower doses may also be linked with fewer side effects [17-19].

The following five vaccine types were included in the final analysis: mRNA, protein subunit, nonreplicating viral vector, inactivated, and DNA vaccines. Other vaccine types and mix-and-match vaccine regimens have been used in the rollout of booster programs. For example, previous studies showed that the mix-and-match CoronaVac/BNT162b2 vaccination regimen was superior when compared to the CoronaVac/CoronaVac regimen in terms of immunogenicity [26]. As more studies are conducted and evidence accumulates, our model can be expanded to incorporate combinations of vaccines, thereby enhancing its robustness and utility. Given that the immunity and efficacy data we used were all from the general population, our results may not apply to specific groups, such as individuals who are immunocompromised or high-risk occupational populations, since neutralization levels and efficacies in these groups can be different [27,28].

Despite the vital role of a timely vaccination plan for reducing negative pandemic impacts in society [29,30], studies have suggested that manifold factors (eg, psychological, sociodemographic, cultural, institutional, environmental, and economic factors) in society that drive infections and numbers of deaths should also be considered and emphasized in plans for facing the next pandemic crisis [31-33].

Despite the robustness of our qualitative results, we identified certain limitations. First, our study does not explicitly include age groups, across which levels of vaccine efficacy and safety may differ [34]. Second, our study does not include the waning rate of vaccine-derived immunity for fractional-dose vaccination, which may vary with respect to dose size [12]. Third, the neutralizing antibody data were extracted for ancestral strains, but data on efficacy against asymptomatic infection were collected for different variants. Ideally, we should have used the neutralizing antibody and efficacy data of the same strains, but the neutralizing antibody data of the different variants were limited, and the neutralizing antibody levels of the different variants may have differed from those of the ancestral strains [35-37]. Hence, our prediction of efficacy against COVID-19 variants was based on neutralization levels specific to ancestral strains. Caution is advised when applying the model; however, it remains valid if neutralization levels of ancestral strains are tested and used as the predictor. Our findings provide a theoretical basis for the efficacy of fractional-dose vaccines, serving as reference evidence for implementing fractional dosing vaccine policies in areas with vaccine shortages to reduce disease burden. Effective governance would facilitate the implementation of fractional dosing strategies at the early phase of the next pandemic, thereby providing time for vaccine production to increase and meet supply and demand requirements [6].

To summarize, our results are consistent with studies on predicting immune protection against symptomatic SARS-CoV-2 infection [23], and we further projected vaccine protection against asymptomatic infection. We expect that fractional-dose vaccination could provide partial immunity against, for example, SARS-CoV-2 infection. This strategy would be cost-effective in curbing the COVID-19 outbreak, as it balances the lower efficacy of smaller doses with faster vaccine coverage and lower side effects, especially if global vaccine shortage issues persist.

Acknowledgments

We acknowledge the financial support from the AIR@InnoHK Programme of the Innovation and Technology Commission of the Government of the Hong Kong Special Administrative Region; the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services (contract number: 75N93021C00015); the National Natural Science Foundation of China (grant number: 82304204); and the Shenzhen-Hong Kong-Macau Science and Technology Project (category C; project number: SGDX20230821091559022). The funders of this study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. We thank the research assistants Shuqi Wang, Zengyang Shao, and Mingda Xu for their help with reviewing and checking eligible studies.

Data Availability

All data were collected from open sources, and a detailed description is provided in the Methods section. The data used in this study can be found in Multimedia Appendices 1 and 2. The code used for data analysis is freely available upon request.

Authors’ Contributions

ZD, CL, and BJC conceived this study, designed statistical and modeling methods, conducted analyses, interpreted results, and wrote and revised the manuscript. LW, WWL, YB and EHYL interpreted results and revised the manuscript.
Conflicts of Interest
BJC reports honoraria from AstraZeneca, Fosun Pharma, GlaxoSmithKline, Moderna, Pfizer, Sanofi Pasteur, and Roche. The authors report no other potential conflicts of interest.

Multimedia Appendix 1
Information extracted from the selected studies.
[XLSX File, 145 KB - publichealth_v10i1e49812_app1.xlsx ]

Multimedia Appendix 2
Symptomatic infection and asymptomatic infection outcomes.
[XLSX File, 57 KB - publichealth_v10i1e49812_app2.xlsx ]

Multimedia Appendix 3
Data analysis procedure and further results from our models.
[DOCX File, 1478 KB - publichealth_v10i1e49812_app3.docx ]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 16 KB - publichealth_v10i1e49812_app4.docx ]

References
Abbreviations

**FDA:** US Food and Drug Administration
monkeypox

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses

---

Edited by A Mavragani; submitted 09.06.23; peer-reviewed by J Kaewkungwal, M Coccia; revised version received 04.06.24; accepted 06.06.24; published 12.07.24.

**Please cite as:**
Predicting Efficacies of Fractional Doses of Vaccines by Using Neutralizing Antibody Levels: Systematic Review and Meta-Analysis
JMIR Public Health Surveill 2024;10:e49812
URL: https://publichealth.jmir.org/2024/1/e49812
doi:10.2196/49812

© Zhanwei Du, Caifen Liu, Yuan Bai, Lin Wang, Wey Wen Lim, Eric H Y Lau, Benjamin J Cowling. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 12.7.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.

Bushra Mahmood¹*, PhD; Prince Adu²,³*, PhD; Geoffrey McKee²,⁴, MD; Aamir Bharmal²,⁴, MD; James Wilton², MPH; Naveed Zafar Janjua²,⁴,⁵, DrPH

¹Division of Endocrinology, University of British Columbia, Vancouver, BC, Canada
²British Columbia Center for Disease Control, Vancouver, BC, Canada
³Department of Social Medicine, Heritage College of Osteopathic Medicine, Ohio University, Dublin, OH, United States
⁴School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada
⁵Centre for Health Evaluation and Outcome Sciences, St Paul’s Hospital, Vancouver, BC, Canada

*these authors contributed equally

Corresponding Author:
Bushra Mahmood, PhD
Division of Endocrinology
University of British Columbia
2775 Laurel St 4116
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 7787073711
Email: bushra.mahmood@bccdc.ca

Abstract

Background: Racialized populations in the United States, Canada, and the United Kingdom have been disproportionately affected by COVID-19. Higher vaccine hesitancy has been reported among racial and ethnic minorities in some of these countries. In the United Kingdom, for example, higher vaccine hesitancy has been observed among the South Asian population and Black compared with the White population, and this has been attributed to lack of trust in government due to historical and ongoing racism and discrimination.

Objective: This study aimed to assess vaccine receipt by ethnicity and its relationship with mistrust among ethnic groups in British Columbia (BC), Canada.

Methods: We included adults ≥18 years of age who participated in the BC COVID-19 Population Mixing Patterns Survey (BC-Mix) from March 8, 2021, to August 8, 2022. The survey included questions about vaccine receipt and beliefs based on a behavioral framework. Multivariable logistic regression was used to assess the association between mistrust in vaccines and vaccine receipt among ethnic groups.

Results: The analysis included 25,640 adults. Overall, 76.7% (22,010/28,696) of respondents reported having received at least 1 dose of COVID-19 vaccines (Chinese=86.1%, South Asian=79.6%, White=75.5%, and other ethnicity=73.2%). Overall, 13.7% (3513/25,640) of respondents reported mistrust of COVID-19 vaccines (Chinese=7.1%, South Asian=8.2%, White=15.4%, and other ethnicity=15.2%). In the multivariable model (adjusting for age, sex, ethnicity, educational attainment, and household size), mistrust was associated with a 93% reduced odds of vaccine receipt (adjusted odds ratio 0.07, 95% CI 0.06-0.08). In the models stratified by ethnicity, mistrust was associated with 81%, 92%, 94%, and 95% reduced odds of vaccine receipt among South Asian, Chinese, White, and other ethnicities, respectively. Indecision, whether to trust the vaccine or not, was significantly associated with a 70% and 78% reduced odds of vaccine receipt among those who identified as White and of other ethnic groups, respectively.

Conclusions: Vaccine receipt among those who identified as South Asian and Chinese in BC was higher than that among the White population. Vaccine mistrust was associated with a lower odds of vaccine receipt in all ethnicities, but it had a lower effect on vaccine receipt among the South Asian and Chinese populations. Future research needs to focus on sources of mistrust to better understand its potential influence on vaccine receipt among visible minorities in Canada.

(JMIR Public Health Surveill 2024;10:e48466) doi:10.2196/48466
KEYWORDS
COVID-19; vaccine hesitancy; mistrust; trust; ethnic minorities; South Asian; vaccine; vaccination; hesitancy; ethnic; ethnicity; minority; cultural; racial; minorities; SARS-CoV-2; coronavirus; Asia; Asian

Introduction

Vaccines are a successful, evidence-based measure against many infectious diseases and a key response to contain the spread of SARS-CoV-2 and prevent severe outcomes such as hospitalizations and death. Soon after the approval of COVID-19 vaccines by Health Canada, nationwide vaccination began on December 14, 2020. As of March 2023, approximately 87.1% of the Canadian population (aged 5 years and older) had received at least 1 dose of COVID-19 vaccines, and 84.5% had completed the 2-dose primary series of vaccination [1]. Although these numbers are impressive, vaccine receipt in Canada has varied by subgroups with some expressing more resistance or hesitancy [2]. Vaccine hesitancy, characterized as a “delay in acceptance or refusal of vaccines despite availability of vaccination services” [3], is a major barrier to prevent severe outcomes for SARS-CoV-2 infection.

The analysis of COVID-19 cases and mortality rates shows disproportionate infections and deaths across various racial and ethnic populations in many countries, including the United Kingdom, the United States, and Canada [4-6]. In the United Kingdom, despite an attenuation of the elevated risk of COVID-19 mortality after adjusting for sociodemographic characteristics and health status, risk remained substantially higher among South Asian people from Bangladeshi and Pakistani background in both the first and the second waves of the pandemic [5]. Similarly, in the United States, the African American or Black and Hispanic populations experienced disproportionately higher rates of SARS-CoV-2 infection and COVID-19–related mortality [4]. In Canada, the South Asian population, one of the fastest growing ethnic minorities [7], has been disproportionately impacted by the pandemic with an increased risk of both infection and mortality from COVID-19 [6]. Age-standardized mortality rates among the South Asian population (31 deaths per 100,000 population) were second to Black (49 deaths per 100,000 population), followed by Chinese (22 deaths per 100,000 population) [6]. Studies from the United Kingdom have also identified the South Asian population as one of the most affected groups with a 10%-50% higher risk of death related to COVID-19 compared with the White British population [8]. However, despite this elevated risk, studies have reported significantly low levels of vaccine acceptance among this group in the United Kingdom [9,10]. Gauging current levels of vaccine acceptance across various population groups (specifically those at higher risk of infection and severe outcomes) and identifying modifiable factors to increase vaccine receipt are important considerations to control COVID-19 [11] and reduce mortality and morbidity. Numerous studies have evaluated COVID-19 vaccine hesitancy among the general population. These studies have attributed various factors to low acceptance of vaccine including concerns around safety and effectiveness of the vaccine [12], potential side effects, and short duration of immunity [13]. Other factors include a low socioeconomic status, identifying as a member of a marginalized population (such as Black, Hispanic, Indigenous, or other person of color), and dissatisfaction or mistrust of the system based on past experiences of discrimination and systematic racism [14,15].

The rapid development of COVID-19 vaccines has been possible due to substantial public funding and unprecedented levels of scientific collaboration and innovation to address global public needs. However, building trust in the vaccine to encourage its receipt has been an uphill task for many governments and public health agencies [16]. Trust is defined as one’s belief that another person or institution will act in accordance with one’s expectations of positive behavior by others [17], and institutional trust is recognized as a key measure of government performance [18]. Some early studies on vaccine hesitancy implicated a general mistrust in the benefits and safety of vaccines and concerns about their unforeseen effect [19,20]. Medical mistrust is a key factor that underlies general health inequities [21] and pertains to both the absence of trust and a sense that someone or something is acting against one’s best interests; this can include health care providers, systems, and government [22,23]. A study investigating the role of medical mistrust in receipt of vaccines among ethnically diverse communities in the United Kingdom found medical mistrust to be associated with COVID-19 vaccine hesitancy and lower intention to vaccinate. Suspicion related to the benefits and safety of vaccines and concerns about their unforeseen effect were the main contributor to this relationship [21].

In this study, we investigated COVID-19 vaccine receipt and its relationship with mistrust in vaccines by ethnicity in British Columbia (BC), Canada.

Methods

Design

We used data from the BC COVID-19 Population Mixing Patterns Survey (BC-Mix), a repeated web-based survey, designed and launched to assess population mixing patterns in BC during the COVID-19 pandemic [24]. To capture participants from a broad demographic range, the survey invitation was disseminated through Instagram, Facebook, YouTube, WhatsApp, Twitter, and Google search engine results pages. The Google Ads Audience manager and Facebook Ads manager allow for paid advertisements to be targeted at specific audiences. We used these tools to target the survey advertisement campaigns to only residents of BC who are 18 years and older. To help capture underrepresented groups, we promoted the survey to various ethnic populations. For instance, a South Asian community organization promoted the survey on their social media pages and also sent the survey to individuals on their mailing address. Although the survey was in English, it was also promoted in different languages (specifically, Korean and Farsi) to members of minority community groups in BC on their social media pages. Flyers
were also distributed at grocery stores and restaurants particularly including those frequented by minority groups.

As of August 30, 2022, over 94,000 participants had participated in the survey since its launch in September 2020. We have described the survey development, design, and domains in detail elsewhere [24,25].

**Measures**

Participants were asked about their receipt of at least 1 dose of any of the approved COVID-19 vaccines in Canada and were also assessed for their level of mistrust in the COVID-19 vaccines.

Responses were rated on a 5-point scale ranging from 1 to 5, with 1 being “strongly disagree” and 5 being “strongly agree.” For the purpose of analyses, the responses were recoded, with those who responded “strongly disagree” or “disagree” coded as “trust” and those who responded “agree” and “strongly agree” coded as “mistrust.” Individuals who chose “neutral” were considered “undecided.”

We assessed sex, age, ethnicity, educational attainment, occupation, and household size based on self-reported data. Those who identified as First Nations, Inuit, Metis, Black (African or Caribbean), Filipino, Latin American or Hispanic, Southeast Asian (eg, Vietnamese, Cambodian, Malaysian, and Laotian), Arab, West Asian (eg, Iranian and Afghan), Korean, Japanese, and other prefer to self-describe, and prefer not to answer, were grouped under “other” due to their small sample sizes. Survey questions and response categories for the BC-Mix baseline and the follow-up survey are provided in Multimedia Appendices 1 and 2, respectively.

**Statistical Analyses**

The domain on COVID-19 vaccine hesitancy and receipt was added to the survey on March 8, 2021. The analysis for this study was restricted to survey responses collected between March 8, 2021, and August 8, 2022. The survey data were weighted using 2016 Census data (Statistics Canada [26]) as the reference population. Employing Bethlehem’s weighting adjustment technique [27], age, sex, geography (Health Authority region), and ethnicity were used as auxiliary variables for weighting.

Participant characteristics were summarized using weighted frequencies and percentages, stratified by ethnicity. The distribution of mistrust in COVID-19 vaccines was examined for each ethnicity. We examined the distribution of vaccine receipt by mistrust status for each ethnicity. Logistic regression was used to examine the unadjusted and adjusted association between mistrust and vaccine receipt stratified by ethnicity. Confounders included ethnicity, sex, age, educational attainment, and household size while accounting for sampling design.

Data preparation, descriptive analyses, and all other analyses were performed in SAS software version 9.4 (SAS Institute Inc).

**Ethical Considerations**

Informed consent was obtained from all participants. Ethical approval for this study was provided by the University of British Columbia Behavioral Research Ethics Board (H20-01785).

**Results**

**Participant Profile**

The analysis included 25,640 individuals. Overall, 55% (15,787/28,694) of the participants were male, and 41.7% (11,967/28,694) were older than 55 years. A total of 10.1% of participants identified as South Asian, 11% as Chinese, 62.7% as White, and 16.1% as of “other” ethnicity. Overall, 7.5% (2140/28,694) lived in a household size of >6, and this proportion was highest among those who identified as South Asian (17.2%, 499/2905) and “other” ethnicities (13.7%, 634/4641) compared with White (4.8%, 872/17,993) and Chinese (4.3%, 135/3156); see Table 1.
Table 1. Participant profile: comparison of characteristics by ethnicity (N=25,640).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>South Asian, weighted n (%)</th>
<th>Chinese, weighted n (%)</th>
<th>White, weighted n (%)</th>
<th>Other, weighted n (%)</th>
<th>All, weighted n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1504 (51.7)</td>
<td>1590 (50.4)</td>
<td>10,325 (57.4)</td>
<td>2369 (51.1)</td>
<td>15,787 (55)</td>
</tr>
<tr>
<td>Female</td>
<td>1402 (48.3)</td>
<td>1566 (49.6)</td>
<td>7667 (42.6)</td>
<td>2272 (48.9)</td>
<td>12,907 (45)</td>
</tr>
<tr>
<td><strong>Age (years)</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>1012 (34.8)</td>
<td>709 (22.5)</td>
<td>3736 (20.8)</td>
<td>1615 (34.8)</td>
<td>7072 (24.6)</td>
</tr>
<tr>
<td>35-54</td>
<td>977 (33.6)</td>
<td>1162 (36.8)</td>
<td>5705 (31.7)</td>
<td>1812 (39)</td>
<td>9655 (33.6)</td>
</tr>
<tr>
<td>≥55</td>
<td>918 (31.6)</td>
<td>1285 (40.7)</td>
<td>8550 (47.5)</td>
<td>1214 (26.2)</td>
<td>11,967 (41.7)</td>
</tr>
<tr>
<td><strong>Educational attainment</strong>&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below high school</td>
<td>59 (2)</td>
<td>58 (1.8)</td>
<td>370 (2.1)</td>
<td>181 (3.9)</td>
<td>669 (2.3)</td>
</tr>
<tr>
<td>Below bachelor</td>
<td>834 (28.7)</td>
<td>658 (20.9)</td>
<td>6910 (38.4)</td>
<td>1586 (34.2)</td>
<td>9988 (34.8)</td>
</tr>
<tr>
<td>University degree</td>
<td>973 (33.5)</td>
<td>1536 (48.7)</td>
<td>5436 (30.2)</td>
<td>1449 (31.2)</td>
<td>9394 (32.7)</td>
</tr>
<tr>
<td>Missing or unknown</td>
<td>1039 (35.8)</td>
<td>903 (28.6)</td>
<td>5276 (29.3)</td>
<td>1426 (30.7)</td>
<td>8644 (30.1)</td>
</tr>
<tr>
<td><strong>Occupation</strong>&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essential workers</td>
<td>559 (19.2)</td>
<td>553 (17.5)</td>
<td>4199 (23.3)</td>
<td>1171 (25.2)</td>
<td>6482 (22.6)</td>
</tr>
<tr>
<td>Nonessential workers</td>
<td>638 (22)</td>
<td>903 (28.6)</td>
<td>4194 (23.3)</td>
<td>1003 (21.6)</td>
<td>6739 (23.5)</td>
</tr>
<tr>
<td>Do not work</td>
<td>401 (13.8)</td>
<td>551 (17.5)</td>
<td>3182 (17.7)</td>
<td>620 (13.4)</td>
<td>4755 (16.6)</td>
</tr>
<tr>
<td>Others</td>
<td>192 (6.6)</td>
<td>227 (7.2)</td>
<td>1166 (6.5)</td>
<td>338 (7.3)</td>
<td>1923 (6.7)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>173 (6)</td>
<td>62 (2)</td>
<td>293 (1.6)</td>
<td>212 (4.6)</td>
<td>740 (2.6)</td>
</tr>
<tr>
<td>Missing or unknown</td>
<td>942 (32.4)</td>
<td>860 (27.2)</td>
<td>4959 (27.6)</td>
<td>1297 (27.9)</td>
<td>8057 (28.1)</td>
</tr>
<tr>
<td><strong>Household size</strong>&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>339 (11.7)</td>
<td>524 (16.6)</td>
<td>3258 (18.1)</td>
<td>681 (14.7)</td>
<td>4802 (16.7)</td>
</tr>
<tr>
<td>2</td>
<td>692 (23.8)</td>
<td>1226 (38.8)</td>
<td>8354 (46.4)</td>
<td>1431 (30.8)</td>
<td>11,702 (40.8)</td>
</tr>
<tr>
<td>3</td>
<td>473 (16.3)</td>
<td>593 (18.8)</td>
<td>2445 (13.6)</td>
<td>719 (15.5)</td>
<td>4231 (14.7)</td>
</tr>
<tr>
<td>4</td>
<td>634 (21.8)</td>
<td>515 (16.3)</td>
<td>2404 (13.4)</td>
<td>741 (16)</td>
<td>4293 (15)</td>
</tr>
<tr>
<td>5</td>
<td>224 (7.7)</td>
<td>163 (5.2)</td>
<td>612 (3.4)</td>
<td>389 (8.4)</td>
<td>1388 (4.8)</td>
</tr>
<tr>
<td>≥6</td>
<td>499 (17.2)</td>
<td>135 (4.3)</td>
<td>872 (4.8)</td>
<td>634 (13.7)</td>
<td>2140 (7.5)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>44 (1.5)</td>
<td>N/A&lt;sup&gt;f&lt;/sup&gt;</td>
<td>48 (0.3)</td>
<td>46 (1)</td>
<td>138 (0.5)</td>
</tr>
</tbody>
</table>

<sup>a</sup>South Asian: n=2906; Chinese: n=3156; White: n=17,992; Other: n=4641; All: n=28,694.
<sup>b</sup>South Asian: n=2907; Chinese: n=3156; White: n=17,991; Other: n=4641; All: n=28,694.
<sup>c</sup>South Asian: n=2905; Chinese: n=3156; White: n=17,993; Other: n=4642; All: n=28,695.
<sup>d</sup>South Asian: n=2905; Chinese: n=3156; White: n=17,993; Other: n=4641; All: n=28,696.
<sup>e</sup>South Asian: n=2905; Chinese: n=3156; White: n=17,993; Other: n=4641; All: n=28,694.
<sup>f</sup>N/A: not applicable.

**Vaccine Receipt by Ethnicity and Trust**

Overall, 76.7% (22,010/28,696) of respondents reported having received at least 1 dose of the COVID-19 vaccines. Vaccine receipt was highest among individuals who identified as Chinese (2718/3156, 86.1%) followed by South Asian (2312/2905, 79.6%), White (13,583/17,991, 75.5%), and those belonging to the “other” ethnicity (3398/4642, 73.2%); see Table 2.

COVID-19 vaccine receipt by vaccine mistrust differed by ethnicity. Overall, 13.7% of the respondents reported mistrust of COVID-19 vaccines (Multimedia Appendix 3). There was a difference by ethnicity in the proportion of people who mistrusted the vaccine but still received it. This proportion was 50.6% among those who identified as South Asian, compared with 40.5% among Chinese, 26.9% among White, and 25.1% among individuals of “other” ethnicities (Table 3).
Table 2. Distribution of COVID-19 vaccine receipt and trust in COVID-19 vaccines by ethnicity in British Columbia, Canada.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>COVID-19 vaccine receipt, n/N (%)</th>
<th>Trust in COVID-19 vaccine, n/N (%)</th>
<th>Trust</th>
<th>Mistrust</th>
<th>Undecided</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Received</td>
<td>Did not receive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asian</td>
<td>2312/2905 (79.6)</td>
<td>595/2917 (20.4)</td>
<td>239/2915 (8.2)</td>
<td>2276/2907 (78.3)</td>
<td>391/2896 (13.5)</td>
</tr>
<tr>
<td>White</td>
<td>13,583/17,991 (75.5)</td>
<td>4408/17,992 (24.5)</td>
<td>2769/17,981 (15.4)</td>
<td>13,499/17,999 (75)</td>
<td>1724/17,958 (9.6)</td>
</tr>
<tr>
<td>Chinese</td>
<td>2718/3156 (86.1)</td>
<td>439/3158 (13.9)</td>
<td>223/3141 (7.1)</td>
<td>2678/3154 (84.9)</td>
<td>255/3148 (8.1)</td>
</tr>
<tr>
<td>Other ethnicity</td>
<td>3398/4642 (73.2)</td>
<td>1243/4638 (26.8)</td>
<td>707/4651 (15.2)</td>
<td>3107/4637 (67)</td>
<td>826/4640 (17.8)</td>
</tr>
<tr>
<td>All</td>
<td>22,010/28,696 (76.7)</td>
<td>6684/28,687 (23.3)</td>
<td>3938/28,745 (13.7)</td>
<td>21,560/28,708 (75.1)</td>
<td>3197/28,802 (11.1)</td>
</tr>
<tr>
<td>Characteristic</td>
<td>South Asian</td>
<td>Did receive vaccine, n (%)</td>
<td>Did not receive vaccine, n (%)</td>
<td>Chinese</td>
<td>Did receive vaccine, n (%)</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>----------------------------</td>
<td>-------------------------------</td>
<td>--------</td>
<td>----------------------------</td>
</tr>
<tr>
<td><strong>COVID-19 vaccine mistrust</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>1914 (82.8)</td>
<td>362 (60.9)</td>
<td>2040 (88.4)</td>
<td>275 (62.7)</td>
<td>89.7</td>
</tr>
<tr>
<td>Mistrust</td>
<td>121 (5.2)</td>
<td>118 (19.9)</td>
<td>90 (3.3)</td>
<td>133 (30.3)</td>
<td>40.5</td>
</tr>
<tr>
<td>Undecided</td>
<td>277 (12)</td>
<td>115 (19.3)</td>
<td>224 (8.2)</td>
<td>31 (7.1)</td>
<td>87.9</td>
</tr>
<tr>
<td><strong>Sex</strong>&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1136 (49.1)</td>
<td>368 (61.9)</td>
<td>1356 (49.9)</td>
<td>233 (53.2)</td>
<td>85.3</td>
</tr>
<tr>
<td>Female</td>
<td>1176 (50.9)</td>
<td>226 (38.1)</td>
<td>1361 (50.1)</td>
<td>205 (46.8)</td>
<td>86.9</td>
</tr>
<tr>
<td><strong>Age (years)</strong>&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>725 (31.4)</td>
<td>287 (48.3)</td>
<td>569 (20.9)</td>
<td>140 (31.9)</td>
<td>80.3</td>
</tr>
<tr>
<td>35-54</td>
<td>793 (34.3)</td>
<td>184 (30.9)</td>
<td>974 (35.8)</td>
<td>188 (42.9)</td>
<td>83.8</td>
</tr>
<tr>
<td>≥55</td>
<td>794 (34.3)</td>
<td>124 (20.8)</td>
<td>1175 (43.2)</td>
<td>110 (25.2)</td>
<td>91.4</td>
</tr>
<tr>
<td><strong>Educational attainment</strong>&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below high school</td>
<td>47 (2)</td>
<td>13 (2.1)</td>
<td>40 (1.5)</td>
<td>18 (4.2)</td>
<td>68.8</td>
</tr>
<tr>
<td>Below bachelor</td>
<td>617 (29.2)</td>
<td>159 (26.7)</td>
<td>583 (21.5)</td>
<td>75 (17.1)</td>
<td>88.6</td>
</tr>
<tr>
<td>University degree</td>
<td>727 (31.4)</td>
<td>247 (41.5)</td>
<td>1337 (49.2)</td>
<td>200 (45.5)</td>
<td>87</td>
</tr>
<tr>
<td>Missing/unknown</td>
<td>863 (37.3)</td>
<td>176 (29.6)</td>
<td>757 (27.9)</td>
<td>146 (33.2)</td>
<td>83.9</td>
</tr>
<tr>
<td><strong>Occupation</strong>&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essential workers</td>
<td>433 (18.7)</td>
<td>126 (21.2)</td>
<td>508 (18.7)</td>
<td>45 (10.3)</td>
<td>91.9</td>
</tr>
<tr>
<td>Nonessential workers</td>
<td>494 (21.4)</td>
<td>144 (24.2)</td>
<td>767 (28.2)</td>
<td>136 (31.1)</td>
<td>84.9</td>
</tr>
<tr>
<td>Do not work</td>
<td>316 (13.7)</td>
<td>86 (14.4)</td>
<td>493 (18.1)</td>
<td>58 (13.3)</td>
<td>89.4</td>
</tr>
<tr>
<td>Others</td>
<td>138 (6)</td>
<td>54 (9.1)</td>
<td>165 (6.1)</td>
<td>61 (14)</td>
<td>73</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>143 (6.2)</td>
<td>31 (5.2)</td>
<td>57 (2.1)</td>
<td>12 (2.5)</td>
<td>91.7</td>
</tr>
<tr>
<td>Missing/unknown</td>
<td>788 (34.1)</td>
<td>154 (25.9)</td>
<td>811 (30.5)</td>
<td>132 (31.1)</td>
<td>84.6</td>
</tr>
<tr>
<td><strong>Household size</strong>&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>289 (12.5)</td>
<td>51 (8.5)</td>
<td>85.1</td>
<td>454 (16.7)</td>
<td>70 (16)</td>
</tr>
</tbody>
</table>
Association Between Mistrust and Vaccine Receipt

In the unadjusted logistic regression model, those who identified as Chinese were significantly more likely to have received the vaccine (odds ratio [OR] 2.01, 95% CI 1.55-2.62) compared with those who identified as White. Although those who identified as South Asians were also more likely to have received the vaccine, this effect was not significant. Individuals who mistrusted COVID-19 vaccines had 94% reduced odds of vaccine receipt (OR 0.06, 95% CI 0.05-0.07) compared with those who trusted vaccines. In the models stratified by ethnicity, mistrust was associated with 81% (OR 0.19, 95% CI 0.09-0.43), 92% (OR 0.08, 95% CI 0.03-0.18), 94% (OR 0.06, 95% CI 0.05-0.07), and 95% (OR 0.05, 95% CI 0.03-0.07) reduced odds of vaccine receipt among those who identified as South Asian, Chinese, White, and those of other ethnicities, respectively. Individuals who were undecided or neutral regarding whether they trusted or mistrusted the vaccine had 70% reduced odds of receiving the vaccine. Among undecided South Asians, reduced odds of receiving the vaccine were 54% (OR 0.46, 95% CI 0.22-0.94) compared with 71% in the White (OR 0.29, 95% CI 0.24-0.34) and 78% in the “other” (OR 0.22, 95% CI 0.16-0.30). In those identifying as Chinese, the reduction in odds was not significant (Table 4).

After adjusting for age, sex, educational attainment, and household size, mistrust was associated with a 93% reduced odds of vaccine receipt (adjusted OR [aOR] 0.07, 95% CI 0.06-0.08). There was some attenuation of the effect of ethnicity, but results were similar for those who identified as Chinese (aOR 1.63, 95% CI 1.20-2.20), those who identified as South Asian (aOR 1.08, 95% CI 0.93-1.25), and those of “other” ethnicities (aOR 1.17, 95% CI 0.86-1.61).

The adjusted magnitude of association between mistrust and vaccine receipt was greatest among individuals of “other” ethnicities (aOR 0.05, 95% CI 0.03-0.07), followed by those who identified as White (aOR 0.06, 95% CI 0.05-0.07), those who identified as Chinese (aOR 0.08, 95% CI 0.03-0.20), and those who identified as South Asian (aOR 0.17, 95% CI 0.07-0.38). A significant association was observed between those who were undecided and those who received the vaccine among individuals identifying as White (OR 0.30, 95% CI 0.25-0.36) and of “other” ethnicity (OR 0.22, 95% CI 0.16-0.31; Table 5).

Other covariates associated with vaccine receipt included female sex (OR 1.59, 95% CI 1.45-1.75), age >55 years compared with 18-34 years (OR 1.88, 95% CI 1.64-2.15), and a university education compared with below high school education (OR 1.71, 95% CI 1.22-2.38; Table 5).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>South Asian</th>
<th>Chinese</th>
<th>White</th>
<th>Other Ethnicity</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Received vaccine, n (%)</td>
<td>Did not receive vaccine, n (%)</td>
<td>Row</td>
<td>Received vaccine, n (%)</td>
<td>Did not receive vaccine, n (%)</td>
<td>Row</td>
</tr>
<tr>
<td>2</td>
<td>543 (23.5)</td>
<td>149 (25.1)</td>
<td>78.4</td>
<td>1068 (39.3)</td>
<td>158 (36.1)</td>
</tr>
<tr>
<td>3</td>
<td>390 (16.9)</td>
<td>83 (14)</td>
<td>82.4</td>
<td>513 (18.9)</td>
<td>80 (18.2)</td>
</tr>
<tr>
<td>4</td>
<td>506 (21.9)</td>
<td>128 (21.5)</td>
<td>79.8</td>
<td>465 (17.1)</td>
<td>51 (11.5)</td>
</tr>
<tr>
<td>5</td>
<td>198 (8.6)</td>
<td>26 (4.3)</td>
<td>88.5</td>
<td>120 (4.4)</td>
<td>44 (9.9)</td>
</tr>
<tr>
<td>≥6</td>
<td>353 (15.2)</td>
<td>147 (24.7)</td>
<td>70.6</td>
<td>99 (3.6)</td>
<td>36 (8.3)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>34 (1.5)</td>
<td>11 (1.8)</td>
<td>75.8</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

a Vaccine received variable (proportion of people who received the vaccine among those who trusted it, mistrusted it, or were undecided).

b South Asian; received vaccine: n=2312, did not receive: n=595; Chinese; received vaccine: n=2718, did not receive: n=439; White; received vaccine: n=13,583, did not receive: n=4408; Other: received vaccine: n=3398, did not receive: n=1243; All: received vaccine: n=22,010, did not receive: n=6684.

c South Asian; received vaccine: n=2312, did not receive: n=594; Chinese; received vaccine: n=2717, did not receive: n=438; White; received vaccine: n=13,584, did not receive: n=4408; Other: received vaccine: n=3397, did not receive: n=1243; All: received vaccine: n=22,010, did not receive: n=6683.

d South Asian; received vaccine: n=2312, did not receive: n=595; Chinese; received vaccine: n=2718, did not receive: n=438; White; received vaccine: n=13,584, did not receive: n=4408; Other: received vaccine: n=3397, did not receive: n=1243; All: received vaccine: n=22,010, did not receive: n=6683.

e South Asian; received vaccine: n=2312, did not receive: n=595; Chinese; received vaccine: n=2717, did not receive: n=439; White; received vaccine: n=13,583, did not receive: n=4408; Other: received vaccine: n=3398, did not receive: n=1243; All: received vaccine: n=22,010, did not receive: n=6684.

f South Asian; received vaccine: n=2312, did not receive: n=595; Chinese; received vaccine: n=2718, did not receive: n=437; White; received vaccine: n=13,584, did not receive: n=4408; Other: received vaccine: n=3399, did not receive: n=1243; All: received vaccine: n=22,010, did not receive: n=6684.

g South Asian; received vaccine: n=2313, did not receive: n=594; Chinese; received vaccine: n=2719, did not receive: n=438; White; received vaccine: n=13,585, did not receive: n=4408; Other: received vaccine: n=3397, did not receive: n=1243; All: received vaccine: n=22,010, did not receive: n=6684.
Table 4. Unadjusted association between mistrust and vaccine receipt by ethnicity in British Columbia.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Unadjusted odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>South Asian</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>N/A^a</td>
</tr>
<tr>
<td>Chinese</td>
<td>N/A</td>
</tr>
<tr>
<td>Other ethnicity</td>
<td>N/A</td>
</tr>
<tr>
<td>South Asian</td>
<td>N/A</td>
</tr>
<tr>
<td>Mistrust in vaccines</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.19 (0.09-0.43)</td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
</tr>
<tr>
<td>Undecided</td>
<td>0.46 (0.22-0.94)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Reference</td>
</tr>
<tr>
<td>Female</td>
<td>1.68 (1.03-2.75)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>Reference</td>
</tr>
<tr>
<td>35-54</td>
<td>1.71 (0.90-3.23)</td>
</tr>
<tr>
<td>≥55</td>
<td>2.54 (1.30-4.96)</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
</tr>
<tr>
<td>Below high school</td>
<td>Reference</td>
</tr>
<tr>
<td>Below bachelor</td>
<td>1.15 (0.19-6.95)</td>
</tr>
<tr>
<td>University degree</td>
<td>0.80 (0.14-4.62)</td>
</tr>
<tr>
<td>Missing/unknown</td>
<td>1.33 (0.23-7.66)</td>
</tr>
<tr>
<td>Household size</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>0.64 (0.22-1.82)</td>
</tr>
<tr>
<td>3</td>
<td>0.82 (0.25-2.71)</td>
</tr>
<tr>
<td>4</td>
<td>0.69 (0.23-2.08)</td>
</tr>
<tr>
<td>5</td>
<td>1.35 (0.33-5.57)</td>
</tr>
<tr>
<td>≥6</td>
<td>0.42 (0.14-1.27)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>0.55 (0.05-6.26)</td>
</tr>
</tbody>
</table>

^aN/A: not applicable.
Table 5. Adjusted association between mistrust and vaccine receipt by ethnicity in British Columbia.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>South Asian</th>
<th>Chinese</th>
<th>White</th>
<th>Other ethnicity</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Reference</td>
</tr>
<tr>
<td>Chinese</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1.63 (1.20-2.20)</td>
</tr>
<tr>
<td>Other ethnicity</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1.08 (0.93-1.25)</td>
</tr>
<tr>
<td>South Asian</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1.17 (0.86-1.61)</td>
</tr>
<tr>
<td>Mistrust in vaccines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.17 (0.07-0.38)</td>
<td>0.08 (0.03-0.20)</td>
<td>0.06 (0.05-0.07)</td>
<td>0.05 (0.03-0.07)</td>
<td>0.07 (0.06-0.08)</td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Undecided</td>
<td>0.47 (0.21-1.07)</td>
<td>0.81 (0.31-2.11)</td>
<td>0.30 (0.25-0.36)</td>
<td>0.22 (0.16-0.31)</td>
<td>0.32 (0.27-0.37)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.79 (1.05-3.06)</td>
<td>0.84 (0.46-1.54)</td>
<td>1.14 (1.03-1.27)</td>
<td>1.39 (1.07-1.81)</td>
<td>1.24 (1.11-1.38)</td>
</tr>
<tr>
<td>Female</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-34</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>35-54</td>
<td>1.97 (0.97-3.99)</td>
<td>1.36 (0.64-2.87)</td>
<td>1.03 (0.84-1.26)</td>
<td>1.10 (0.78-1.55)</td>
<td>1.18 (1.00-1.40)</td>
</tr>
<tr>
<td>≥55</td>
<td>2.78 (1.42-5.46)</td>
<td>2.16 (0.97-4.82)</td>
<td>1.54 (1.28-1.86)</td>
<td>1.75 (1.23-2.50)</td>
<td>1.78 (1.52-2.10)</td>
</tr>
<tr>
<td>Educational attainment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below high school</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Below bachelor</td>
<td>0.83 (0.11-6.36)</td>
<td>2.70 (0.68-10.81)</td>
<td>0.96 (0.62-1.50)</td>
<td>0.74 (0.31-2.06)</td>
<td>0.98 (0.65-1.48)</td>
</tr>
<tr>
<td>University degree</td>
<td>0.57 (0.08-4.26)</td>
<td>2.20 (0.61-7.86)</td>
<td>1.07 (0.68-1.66)</td>
<td>0.91 (0.65-1.50)</td>
<td>1.04 (0.69-1.57)</td>
</tr>
<tr>
<td>Missing/unknown</td>
<td>1.07 (0.14-7.98)</td>
<td>1.83 (0.51-6.52)</td>
<td>1.26 (0.81-1.97)</td>
<td>1.04 (0.40-2.71)</td>
<td>1.24 (0.82-1.87)</td>
</tr>
<tr>
<td>Household size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>2</td>
<td>0.51 (0.18-1.46)</td>
<td>1.03 (0.48-2.25)</td>
<td>0.88 (0.75-1.03)</td>
<td>0.98 (0.65-1.50)</td>
<td>0.88 (0.75-1.03)</td>
</tr>
<tr>
<td>3</td>
<td>0.77 (0.23-2.58)</td>
<td>1.10 (0.42-2.88)</td>
<td>0.85 (0.69-1.08)</td>
<td>1.36 (0.81-2.27)</td>
<td>0.94 (0.76-1.16)</td>
</tr>
<tr>
<td>4</td>
<td>0.61 (0.20-1.87)</td>
<td>1.50 (0.59-3.82)</td>
<td>0.86 (0.69-1.08)</td>
<td>1.07 (0.65-1.77)</td>
<td>0.91 (0.74-1.12)</td>
</tr>
<tr>
<td>5</td>
<td>1.21 (0.27-5.47)</td>
<td>0.43 (0.13-1.36)</td>
<td>0.84 (0.61-1.16)</td>
<td>0.85 (0.43-1.68)</td>
<td>0.81 (0.58-1.12)</td>
</tr>
<tr>
<td>≥6</td>
<td>0.45 (0.15-1.39)</td>
<td>0.70 (0.26-1.88)</td>
<td>0.62 (0.46-0.83)</td>
<td>1.03 (0.60-1.76)</td>
<td>0.69 (0.53-0.89)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.66 (0.26-1.70)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

**Discussion**

This study assessed the role of self-reported mistrust of COVID-19 vaccines in vaccine receipt among different ethnic groups in BC, Canada. We found that mistrust was lower among individuals who identified as Chinese and South Asian. Vaccine uptake was higher among people identifying as Chinese compared with people identifying as White. Among those who identified as South Asian, vaccine receipt was higher than the White population, but this difference was not statistically significant. Mistrust of the vaccine was significantly associated with a reduced odds of vaccine receipt, regardless of race and ethnicity. However, the magnitude of association between mistrust and vaccine receipt differed by ethnicity, with vaccine trust playing the least important role in vaccine receipt among those who identified as South Asian. Only 8.2% of those who identified as South Asian reported mistrust (Multimedia Appendix 3), and despite that, vaccine receipt among this group was still the highest (50.6%) compared with 40.5% among those who identified as Chinese, 26.9% among White, and 25.1% among individuals of “other” ethnicities. Indecisiveness about whether the vaccine was to be trusted or not also had a significant association with vaccine receipt. In the “undecided” group, those who identified as Chinese and South Asians had the highest proportion of vaccine receipt (88% and 71%, respectively) compared with White (66%) and those identifying as of “other” ethnicities (61%).
Our results contrast with findings from other studies that have reported a higher prevalence of vaccine hesitancy and lower vaccine receipt in racialized groups. For instance, although the United Kingdom became the first country to approve COVID-19 vaccines for emergency use in December 2020 and took the lead in vaccinating its population, it faced elevated vaccine hesitancy among Black, South Asian, and other ethnic minorities [9,28-31]. Similarly, US-based studies reported greater vaccine hesitancy and lower vaccine receipt among African Americans and Hispanics compared with the White American population [32-34].

Several reasons have been cited for this widespread hesitancy and lower vaccine receipt among ethnic groups. Unknown future effects, side effects of vaccines, a lack of trust in government and pharmaceutical industry, and historical inequities and racism have been implicated in vaccine hesitancy among people of color in both the United Kingdom and United States [9,32,34,35]. Contrary to this, overall vaccine acceptance among people designated as a visible minority in Canada has been relatively high with 74.8% reporting being “very” or “somewhat willing” to receive COVID-19 vaccines [2]. Of the 11 distinct ethnic groups surveyed in Canada, the South Asian group had one of the highest proportions of vaccine acceptance (82.5%) third in line to Japanese at 87.6% and Koreans at 85.5% compared with 77.7% in those who were not a visible minority [2].

In our study, mistrust of the vaccine as well as indecisiveness whether to trust the vaccine or not was significantly associated with a reduced odds of vaccine receipt regardless of race and ethnicity. The timing of our survey may explain some of these trends. The BC-Mix survey was launched in September 2020, with the vaccine questionnaire being added in March 2021—approximately 3 months after nationwide vaccination in Canada began. Limited information about the novel SARS-CoV-2 virus and concerns over safety and efficacy of a vaccine developed in a very short time may have generated a considerable number of people indicating reluctance to get vaccinated as reported in studies from the United States and France [36-38]. This may, to some degree, also explain a high prevalence of indecisiveness in our study, which, as our results demonstrate, had a significant association with willingness to get vaccinated. Dearth of accurate information, misinformation or mixed messages coupled with lack of culturally tailored, and targeted messaging in the early months of the pandemic may have further fueled skepticism or indecisiveness among various population groups—specifically the already racialized and marginalized communities. However, despite this, mistrust and indecisiveness appeared to have less of an impact on vaccine receipt among those who identified as Chinese (88%) and South Asian (70%).

There may be several potential explanatory factors for why people who identified as South Asian received a vaccine despite reporting vaccine mistrust. First, the decision to get vaccinated despite mistrust may be due to targeted interventions by BC Health Authorities to increase vaccine uptake within the South Asian community. These interventions may have, to some degree, overridden the initial fears and mistrust at the individual level. Second, a higher risk perception of COVID-19 severe outcomes, perhaps from having witnessed the high burden of disease and mortality among community members in both BC and Ontario [39] coupled with messaging from primary care physicians and community-based health workers framing vaccines as a “lifesaver,” may have driven many to opt for the vaccine despite their initial mistrust [40]. Third, subjective norms could be another explanation whereby members of the South Asian community, mistrustful of the vaccines, felt pressured to “give in” as other members of the family or one’s social circle appeared to be strongly supportive of vaccination [40]. Fourth, South Asian grandparents are 8 times more likely to live with their grandchildren (National Household Survey 2011) [41], and this seems to be reflected in our sample, where 17.2% of South Asian individuals (highest compared with any other ethnic group) lived in a household size of 6 or greater. Fear of exposing the older adult in the family to COVID-19 may have been another factor pushing those who were mistrustful of the vaccine to get vaccinated. Lastly, fear of consequences for not following government regulations, such as those related to international travel or accessing public facilities, may have convinced some to get vaccinated despite mistrust [40]. Future research should focus on some of these psychosocial and behavioral factors that could explain greater acceptance and receipt of COVID-19 vaccines among the South Asian population.

Contrasting findings have been observed in UK and US studies. In the UK Household Longitudinal Study, higher vaccine hesitancy (among those who had not yet received a vaccine) was seen in most minority ethnic groups compared with the White British or Irish group. The highest odds were seen in the Black or Black British group (OR 13.42, 95% CI 6.86-26.24) and the Pakistani and Bangladeshi South Asian subgroups (OR 2.54, 95% CI 1.19-5.44), and adjustment for covariates made relatively little difference to these associations [29]. In the United States, Wu et al [35] examined the mechanisms behind COVID-19 vaccine acceptance in 3 Asian American ethnic groups (East Asian, Southeast Asian, and South Asian), including how sociodemographic characteristics and racism predicted COVID-19 vaccine perceptions. Trustworthiness of public health agency’s recommendation for COVID-19 vaccines was not associated with ethnicity. When predicting COVID-19 vaccine safety concerns, the South Asian group (compared with non-Asian) reported significantly higher safety concerns, which was mediated by racism [35].

Despite some methodological differences, contrasting findings of vaccine hesitancy and receipt among the South Asian population in Canada compared with the United Kingdom raises thought-provoking questions as to why the Canadian South Asian population might be more accepting and trusting of COVID-19 vaccines than their British counterparts. This finding may partly be explained by subgroup differences that exist among South Asian diaspora. It is important to note that South Asians (people originating from or having ancestry in India, Pakistan, Bangladesh, Nepal, Sri Lanka, Maldives, and Bhutan) are not a monolithic group but have considerable heterogeneity based on country of origin, religion, culture, and language. In the United Kingdom, while Indians comprise the biggest South Asian subgroup (similar to Canada), people of Pakistani and
Bangladeshi origin also have a significant presence. In fact, when put together, the group comprising people with Pakistani and Bangladeshi origins surpass Indians in numbers [42]. Significant differences in health behaviors and health outcomes for chronic diseases have been reported among various South Asian subgroups based on religion and country of origin [43]. Research exploring vaccine receipt and hesitancy among various subgroups of ethnic denominations in the United Kingdom reveals greater hesitancy and mistrust among the South Asian population originating from predominantly Muslim countries such as Bangladesh and Pakistan compared with India [29,44].

In the province of BC, Indians, and specifically Indians of Punjabi origin who are settled in the province of Punjab and identify predominantly with Sikh religion, constitute the largest South Asian subgroup, whereas the number of Pakistanis and, especially, Bangladeshis is relatively small [45]. Although we do not have a subgroup breakdown of the South Asian sample, it is possible that our study sample predominantly constitutes South Asian individuals originating from India, and thus our findings may be reflective of the subgroup differences highlighted by UK-based studies where South Asians of Indian origin are more accepting of the vaccine than Pakistanis or Bangladeshis [29,44]. Vaccine acceptance has been a challenge in many Muslim countries where rejection of immunization is widely prevalent on the grounds that the vaccine contains animal byproducts that are not halal (kosher) or unfounded claims that it may cause infertility [46-48]. Subgroup differences based on country of origin and religion require further investigations to understand vaccine receipt and hesitancy among the South Asian population, which may inform effective vaccine receipt strategies for COVID-19 and other vaccines in future.

Another plausible explanation for the difference in vaccination receipt between the South Asian population in the United Kingdom compared with Canada could be greater anti-Muslim sentiment and social exclusion experienced by the Muslim community in the United Kingdom [49]. This may therefore result in greater mistrust between political leadership and Muslims. The South Asian population in the United Kingdom, specifically Muslims, have been targets of increased hate crimes associated with the war on terror after the events of September 11, 2001, and high-profile police action after the 2005 London bombings, which resulted in alienation of the Muslim community and greater distrust of the government [50]. Although anti-Muslim sentiment in Canada has also increased, a 2016 national survey examining the relationship between Canadian Muslims and Canadian society at large revealed that Muslims, nearly 7 in 10 of whom are immigrants, were generally happy to be in Canada and 84% of Muslims surveyed believed that they were treated better in Canada than their coreligionists in other Western countries, a number that has increased since 2006 [51]. The majority also expressed optimism regarding their relationship with the government. In contrast, a 2016 British examination about the state of Muslim citizens’ opinions on a broad range of topics revealed that relations between Britain’s political establishment and the leadership of the Muslim community are at a nadir [50]. Higher vaccine acceptance among the South Asian population in Canada may thus partly be explained by a higher degree of social inclusion experienced by one of the major subgroups (ie, Muslims) within the South Asian diaspora compared with Muslims in the United Kingdom. This however is speculative and needs to be investigated in future research as an understanding of the contexts and fluidities underlying the knowledge and beliefs and lived experiences about health and immunization in diverse communities is critical for designing effective interventions [52].

Higher vaccine receipt among the South Asian population in Canada could also be related to potential differences in vaccine program administration. For instance, ethnic minorities in the United Kingdom were more likely to receive an influenza vaccine administered by general practitioners in communities with whom they may have a trusting relationship, while COVID-19 vaccines were administered through mass vaccination centers or hospital hub [53]. Moreover, residential segregation driven by systemic racism may have resulted in additional barriers in terms of travel time and transport-related costs to accessing centralized vaccination sites for some communities in the United Kingdom [53]. Although a somewhat similar vaccination setup was rolled out in BC, working at a grassroots level to get the populations most impacted by the pandemic onboard may explain lower resistance and higher levels of vaccine receipt among these groups. For instance, in BC, the greatest concentration of the South Asian population is in the Fraser Health region where they make up almost 15% of the population [54]. The South Asian Health Institute was established in 2013 to improve health for South Asian people and reduce chronic disease burden. This institute works closely with Fraser Health Authority (a regional health authority), community leaders and stakeholders to promote health and well-being and has ongoing programs targeting lifestyle behaviors like diet, physical activity as well as mental health, and drug toxicity. During the pandemic, South Asian Health Institute and several South Asian community volunteer groups and clinic partners were able to leverage their pre-pandemic established relationships and networks within the South Asian community and were at the forefront in reaching out to the community via “in-reach” clinics set up in community centers and places of worship [55]. These clinics were staffed with South Asian team members providing support for registration with the provincial COVID-19 vaccination program. Moreover, translation services for COVID-19 information were available in 145 languages including Hindi, Urdu, Punjabi, Tamil, and Chinese [55]. These measures removed many barriers around language, transportation, technological issues, health system navigation, and vaccine hesitancy. This may partly explain high vaccine receipt even among groups that were mistrustful of the vaccine.

Limitations

This study has some limitations. Because this was a web-based survey in English, it may have excluded many South Asians and Chinese with limited digital literacy and fluency in English. Thus, the findings cannot be generalized to these population groups. The findings of our survey may be subject to social desirability bias; given the self-reported nature of the data, it is possible that participants responded in ways that they deemed socially acceptable. We were also unable to provide further
disaggregation of ethnicity to examine differences among certain ethnic groups and subgroups (eg, Indian, Pakistani, and Bangladeshi groups in the South Asian diaspora). Additionally, we did not investigate the source or cause of mistrust that limits our ability to identify modifiable factors for future vaccination strategies. Despite these limitations, this is one of the largest surveys examining vaccine receipt and mistrust among South Asians in Canada.

Conclusions
Factors that influence vaccine acceptance are multifaceted and often, as research shows, shaped by complex, sociocultural, historical, institutional, clinical, community, and governmental levels [56]. In contrast to findings among South Asians in the United Kingdom, our study found higher levels of vaccine acceptance, and similar levels of vaccine uptake, compared with individuals identifying as White. We also found high vaccine receipt among those South Asians who mistrusted the vaccine. The contrasting results may have a range of explanations—including important differences between the South Asian populations in the United Kingdom and Canada in terms of country of origin, religion, racism, social exclusion, alienation, and consequent mistrust in the government and systems—all of which have an impact on vaccination uptake. Future research needs to focus on disentangling the potential influence of these factors on vaccine receipt. An understanding of factors that facilitate vaccine receipt among the South Asian population residing in Canada will enable researchers and health professionals to tailor this information and use it for developing targeted interventions and informed national vaccination programs for South Asians residing in regions with elevated vaccine hesitancy.

Data Availability
The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
NZJ participated in advisory boards and has spoken for AbbVie and Gilead, not related to current work. The other authors declare no conflicts of interest.

Multimedia Appendix 1
BC COVID-19 population mixing patterns baseline survey (BC-Mix).
[PDF File (Adobe PDF File), 558 KB - publichealth_v10i1e48466_app1.pdf ]

Multimedia Appendix 2
BC COVID-19 population mixing patterns follow-up survey (BC-Mix).
[PDF File (Adobe PDF File), 440 KB - publichealth_v10i1e48466_app2.pdf ]

Multimedia Appendix 3
Mistrust in COVID-19 vaccine by ethnicity in British Columbia.
[DOCX File , 13 KB - publichealth_v10i1e48466_app3.docx ]

References


49. Two-thirds of Muslims consider leaving UK. The Guardian. URL: https://www.theguardian.com/uk/2005/jul/26/politics.july7 [accessed 2023-03-23]


**Abbreviations**

- aOR: adjusted odds ratio
- BC: British Columbia
- OR: odds ratio

---

©Bushra Mahmood, Prince Adu, Geoffrey McKee, Aamir Bharmal, James Wilton, Naveed Zafar Janjua. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 16.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Durability of the Effectiveness of Heterologous COVID-19 Vaccine Regimens in Thailand: Retrospective Cohort Study Using National Registration Data

Ponlagrit Kumwichar¹, MD, PhD; Chittawan Poonsiri², PharmD; Siobhan Botwright², MA; Natchalaikorn Sirichumroonwit³, MD; Bootsakorn Loharjun³, MD, MSc; Supharerk Thawillarp³, MD, MS, DrPH; Nontawit Cheewaruangroj⁵, PhD; Amorn Chokchaisiripakdee⁵, MS; Yot Teerawattananon⁵, MD, PhD; Virasakdi Chongsuvivatwong¹, MD, PhD

¹Department of Epidemiology, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand
²Health Intervention and Technology Assessment Program, Ministry of Public Health, Nonthaburi, Thailand
³Department of Medical Services, Institute of Medical Research and Technology Assessment, Ministry of Public Health, Nonthaburi, Thailand
⁴Department of Disease Control, Ministry of Public Health, Nonthaburi, Thailand
⁵Big Data Institute, Ministry of Digital Economy and Society, Bangkok, Thailand

Corresponding Author:
Virasakdi Chongsuvivatwong, MD, PhD
Department of Epidemiology
Faculty of Medicine
Prince of Songkla University
15 Kanchanavanit Road, Hat Yai
Songkhla, 90110
Thailand
Phone: 66 74451165
Email: cvirasak@medicine.psu.ac.th

Abstract

Background: The durability of heterologous COVID-19 vaccine effectiveness (VE) has been primarily studied in high-income countries, while evaluation of heterologous vaccine policies in low- and middle-income countries remains limited.

Objective: We aimed to evaluate the duration during which the VE of heterologous COVID-19 vaccine regimens in mitigating serious outcomes, specifically severe COVID-19 and death following hospitalization with COVID-19, remains over 50%.

Methods: We formed a dynamic cohort by linking records of Thai citizens aged ≥18 years from citizen vital, COVID-19 vaccine, and COVID-19 cases registry databases between May 2021 and July 2022. Encrypted citizen identification numbers were used to merge the data between the databases. This study focuses on 8 common heterologous vaccine sequences: CoronaVac/ChAdOx1, ChAdOx1/BNT162b2, CoronaVac/ChAdOx1/ChAdOx1, CoronaVac/ChAdOx1/BNT162b2, BBIBP-CorV/BBIBP-CorV/BNT162b2, ChAdOx1/ChAdOx1/BNT162b2, and ChAdOx1/mRNA-1273. Nonimmunized individuals were considered for comparisons. The cohort was stratified according to the vaccination status, age, sex, province location, month of vaccination, and outcome. Data analysis employed logistic regression to determine the VE, accounting for potential confounders and durability over time, with data observed over a follow-up period of 7 months.

Results: This study includes 52,580,841 individuals, with approximately 17,907,215 and 17,190,975 receiving 2- and 3-dose common heterologous vaccines (not mutually exclusive), respectively. The 2-dose heterologous vaccinations offered approximately 50% VE against severe COVID-19 and death following hospitalization with COVID-19 for 2 months; however, the protection significantly declined over time. The 3-dose heterologous vaccinations sustained over 50% VE against both outcomes for at least 8 months, as determined by logistic regression with durability time-interaction modeling. The vaccine sequence consisting of CoronaVac/CoronaVac/ChAdOx1 demonstrated >80% VE against both outcomes, with no evidence of VE waning. The final monthly measured VE of CoronaVac/CoronaVac/ChAdOx1 against severe COVID-19 and death following hospitalization at 7 months after the last dose was 82% (95% CI 80.3%-84%) and 86.3% (95% CI 83.6%-84%), respectively.

Conclusions: In Thailand, within a 7-month observation period, the 2-dose regimens could not maintain a 50% VE against severe and fatal COVID-19 for over 2 months, but all of the 3-dose regimens did. The CoronaVac/CoronaVac/ChAdOx1 regimen
showed the best protective effect against severe and fatal COVID-19. The estimated durability of 50% VE for at least 8 months across all 3-dose heterologous COVID-19 vaccine regimens supports the adoption of heterologous prime-boost vaccination strategies, with a primary series of inactivated virus vaccine and boosting with either a viral vector or an mRNA vaccine, to prevent similar pandemics in low- and middle-income countries.

Introduction

The effectiveness of COVID-19 vaccine is typically determined through randomized controlled trials prior to market approval [1]. Most clinical trials on COVID-19 vaccines focused on homologous vaccination [2], whereas most vaccination programs in low- and middle-income countries (LMICs) use heterologous vaccination for the primary series [3-7]. Recently, evidence has consistently shown that the vaccine effectiveness (VE) of a 3-dose homologous messenger ribonucleic acid (mRNA) vaccine regimen against severe COVID-19 appears to wane at 6 months after the third dose. However, evidence regarding the durability of 3-dose heterologous vaccination regimens remains limited [8-11]. Most research on the durability of heterologous COVID-19 VE have been conducted in high-income countries [12-14]. Although real-world VE studies provide information about the protective effects of heterologous regimens, considerable variations exist across studies because of differences in the vaccine combinations used, unmeasured confounders, and differences in the vaccination eligibility criteria over time and across countries [15]. To our knowledge, no population-based studies have been conducted on the durability of VE of the heterologous prime-boost regimens primarily used in LMICs, which involve a primary series of inactivated vaccines followed by an mRNA or a viral vector vaccine.

In Thailand, a heterologous vaccination strategy was implemented without strong evidence of VE due to supply constraints [16]. The primary series of vaccinations was initially administered using inactivated virus vaccines, followed by viral vector vaccines, whereas mRNA vaccines were mainly used as a booster dose [7,17,18]. Although test-negative case-control studies have been conducted of heterologous vaccination effectiveness in Thailand [7,17,18], these studies are limited in size and follow-up period and therefore were unable to determine the temporal relationship [19]. Thailand is not unique in this regard. To the best of our knowledge, no LMIC has modeled VE durability or waning over time in population-based cohort studies.

Thailand has maintained a national citizen registry linked to national health databases since 1993 [20]. Specialized registries have been established for COVID-19 testing, COVID-19 cases, and COVID-19 vaccination [21]. By linking these records, we examined a dynamic cohort of the Thai population to assess the durability of VE of heterologous vaccine sequences, providing insights for future pandemics. This study aims to evaluate the duration during which the VE of heterologous vaccine sequences against severe COVID-19 and death following hospitalization for COVID-19 (fatal COVID-19) remains over 50%.

Methods

Study Design

This was a dynamic cohort study in which participants entered or exited the cohort at different timepoints during the study period, depending on when they met the eligibility criteria for entering or leaving the cohort [22]. The time frame of this study was from May 2021 (when the COVID-19 vaccination started in Thailand) to July 2022 (when the compulsory registration of all COVID-19 cases ended). We assumed that all individuals residing in the same province would have a similar risk of exposure to SARS-CoV-2 during the pandemic.

Setting

Figure 1A displays the time series of notified SARS-CoV-2 infection, severe COVID-19, and deaths following hospitalization with COVID-19, while Figure 1B shows the circulating variants. Figure 1A shows that the national trends in COVID-19 case notifications and COVID-19 vaccine administration from May 2021 to June 2022. In Figure 1B, from May 2021 to May 2022, B.1.1.7 (Alpha) was rapidly replaced by B.1617.2 (Delta), which caused a peak in severe and fatal COVID-19 cases. The Delta variant has since been replaced by the B.1.1.529 (Omicron) variant, which has a higher infection rate but reduced severity [23]. A marked reduction was noted in the reporting of COVID-19 cases after the reimbursement protocol for COVID-19 was integrated into the universal health coverage program [24]. Therefore, no data analysis was conducted beyond July 2022. Figure 1C shows that the first vaccine to be rolled out was CoronaVac (Sinovac [SV], Sinovac Biotech), for which the rate of uptake remained steady until the end of 2021. ChAdOx1 nCoV-19 (University of Oxford/AstraZeneca [AZ]) was rolled out in June 2021 and was the main vaccine used during the Delta variant outbreak. Simultaneously, a limited volume of BBIBP-CorV (Sinopharm [SP], National Biotech Group, Beijing Institute of Biological Products) became available via national academies and the private sector. BNT162b2 (Pfizer/BioNTech Inc [PZ]) was introduced at the end of 2021 and was expected to become the main vaccine used in 2022. The market share of mRNA-1273 (Moderna [MN]) has remained low.
**Figure 1.** The overall situation of the COVID-19 outbreak during the recruitment period of the dynamic cohort. (A) Daily number of cases of COVID-19 notified (top 2 rows) and COVID-19 deaths (bottom row). (B) SARS-CoV-2 variants circulating in Thailand (pink: Alpha; green: Delta; lilac: Omicron). (C) Daily number of vaccine doses administered by product during the study period.

**Data Sources**

Data were retrieved from national databases of the Thai Ministry of Public Health (MoPH) and the Ministry of Interior. The core data set is the civil registration database, which contains all the citizen identification records. In Thailand, citizens must show their citizen identification cards to access public services such as government-funded health care. COVID-19 vaccine procurement and delivery in both the public and private sectors were monitored by the MoPH through the Mor Prom database, which holds details, including the location of vaccination, citizen identification code, vaccine type, and vaccine batch. Another MoPH database, CO-WARD, recorded all reverse transcription polymerase chain reaction (RT-PCR) test results and clinical outcomes, including severity and death. COVID-19 testing,
quarantine, treatment, and vaccination costs were all recorded for reimbursement through national health insurance schemes. In addition, all deaths are recorded by law in a separate national Ministry of Interior (Thailand) database. All the databases were linked using encrypted citizen identification in the analysis.

Participants

In this study, the participants were all Thai citizens aged 18 years or older because the 2021 vaccination policy only included the adult population. At the start of the study period (May 1, 2021), all participants were nonimmunized, meaning that they had neither received a vaccination nor had a prior SARS-CoV-2 infection. Upon vaccination, participants moved into a group defined by the sequence of vaccines received: SV-AZ, AZ-PZ, SV-SV-AZ, SV-AZ-AZ, SV-AZ-PZ, SP-SP-PZ, AZ-AZ-PZ, and AZ-AZ-MN. Thus, the participant vaccination status could change over time sequentially from nonimmunized to 1, 2, and 3 doses. Participants were excluded from the cohort if they had evidence of SARS-CoV-2 infection (positive antigen test kit [ATK] or RT-PCR) or if they died from any cause. Otherwise, they were followed up until the end of the study period (July 31, 2022).

Exposures and Outcomes

Table 1 lists the outcomes and exposure variables. Outcomes were defined as untested/uninfected or infected (positive RT-PCR results). The infected group was further classified as nonsevere or severe (oxygen saturation ≤94% on room air or rapid progressive pneumonia based on the doctor’s decision) [25] and fatal infection (any hospitalized death with SARS-CoV-2 infection). The assessment of severity was assessed daily, without missing data [26]. We omitted the evaluation of VE in asymptomatic patients or patients with nonsevere COVID-19 because of underreporting issues. Additionally, our analysis omitted cases of severe COVID-19 that resulted in death prior to hospitalization. This decision was based on our data exploration, which showed that such scenarios accounted for less than 1% of COVID-19–related deaths. On December 1, 2021, the national policy changed such that individuals who tested positive using an ATK were defined as having SARS-CoV-2 infection, whereas previously, the case definition required confirmation using RT-PCR [27]. Owing to this change in the classification of cases during the study period, we included only severe COVID-19 cases and deaths following hospitalization for COVID-19 (fatal COVID-19) as the outcomes of interest. The primary independent variable in this study was the vaccine regimen. Due to the presence of multiple categories of vaccination sequences, our analysis focused on the 8 most prevalent heterologous sequences. Other vaccine sequences administered to an insufficient number of individuals (see Data Availability) were excluded from this report. The control group consisted of nonimmunized individuals. Additional variables for confounder adjustment encompassed age, sex, province of residence, and the calendar month of infection (a proxy for the predominant variant during each period). The results were consolidated based on the time elapsed 2 months following the receipt of the last vaccine dose, as evidence suggests that the decline in the immune response against the SARS-CoV-2 virus starts 3 months after the final dose [28].
Table 1. Variables used in the data analysis to determine the durability of heterologous COVID-19 vaccine regimens in Thailand.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Abbreviation</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe COVID-19</td>
<td>Y₁</td>
<td>No evidence of SARS-CoV-2 infection or SARS-COV-2 infection without severe COVID-19 (reference)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Severe COVID-19</td>
</tr>
<tr>
<td>Fatal COVID-19</td>
<td>Y₂</td>
<td>No evidence of SARS-CoV-2 infection or SARS-COV-2 infection without death (reference)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Death following hospitalization with COVID-19</td>
</tr>
<tr>
<td><strong>Exposures</strong></td>
<td>V</td>
<td>No immunization (reference)</td>
</tr>
<tr>
<td>Vaccine sequence</td>
<td></td>
<td>SV&lt;sup&gt;a&lt;/sup&gt;-AZ&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AZ-PZ&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV-SV-AZ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV-AZ-AZ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SV-AZ-PZ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP-SP&lt;sup&gt;d&lt;/sup&gt;-PZ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AZ-AZ-PZ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AZ-AZ-MN&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Stratification factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age group (years)</td>
<td>x₁</td>
<td>18-40 (reference)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41-60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41-60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>61-80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>over 80</td>
</tr>
<tr>
<td>Sex</td>
<td>x₂</td>
<td>Female (reference)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Province of residence</td>
<td>x₃</td>
<td>Bangkok (capital, reference)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other provinces (76 provinces)</td>
</tr>
<tr>
<td>Month of infection&lt;sup&gt;f&lt;/sup&gt;</td>
<td>x₄</td>
<td>July 2021 (reference)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>August 2021-July 2022 (11 months)</td>
</tr>
<tr>
<td>Time since 2 months after the last dose of each vaccine sequence (months)&lt;sup&gt;g&lt;/sup&gt;</td>
<td>T</td>
<td>0 (2 months after the last dose of each vaccine sequence, reference)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-5</td>
</tr>
</tbody>
</table>

<sup>a</sup>SV: Sinovac (CoronaVac).
<sup>b</sup>AZ: AstraZeneca (ChAdOx1).
<sup>c</sup>PZ: Pfizer/BioNTech (BNT162b2).
<sup>d</sup>SP: Sinopharm (BBIBP-CorV).
<sup>e</sup>MN: Moderna (mRNA-1273).
<sup>f</sup>Month of infection is a categorical variable.
<sup>g</sup>Time elapsed since 2 months after the last dose of each vaccine sequence was treated as a categorical variable for the Mantel-Haenszel method and as a continuous variable in the logistic regression model (1 month was assumed to be 30 calendar days).

**Data Manipulation Methods**

The encrypted citizen identification is used to merge data from each database. Individuals were classified according to their vaccination sequence at the end of each month and further stratified according to sex, age group, province of residence, SARS-CoV-2 infection status, and severity of infection. Once infected (positive ATK or RT-PCR results), the participants were excluded from the analysis for subsequent months.

**Data Analysis**

**Estimation of VE Over Time**

The variables employed in our analysis (Table 1) were used in stratified cross-tabulation and logistic regression analyses. The outcome variables were severe COVID-19 (Y₁) and death following COVID-19 (Y₂). They were analyzed separately and reported. The exposure variable was the vaccine sequence, for which we included two 2-dose and five 3-dose regimens. Other vaccine sequences had small sample sizes and were omitted.
In this study, 2 analytical methods were employed to evaluate VE. The first method, the Mantel-Haenszel risk ratio [29], calculated VE\textsubscript{i} as \( VE\textsubscript{i} = 1 – aRR \), where \( aRR \) is the adjusted risk ratio, adjusted for other variables (x) and t represents a categorical time variable (2-7 months after the last dose), as indicated by data observed within 7 months since the last dose. In this study, T+2 was used as the measure of t, based on the assumption that VE begins to wane 2 months after the last dose. The second method, logistic regression, assesses VE durability using the formula: \( VE\textsubscript{i} = 1 – aOR \), where \( aOR \) is the adjusted odds ratio, adjusted for other variables (x), and t is a continuous time variable measuring the months after the last dose. This approach allows for projecting aOR to calculate \( VE\textsubscript{i} \) more than 7 months postvaccination. Given the rarity of severe and fatal COVID-19 cases (incidence less than 1 per 1000), logistic regression was used to estimate the aOR of the vaccine, and the aOR was used to estimate the VE. This is based on the assumption that the odds ratio (OR) approximates the relative risk in rare event scenarios [30-32]. The hypothesis posits that VE may either persist or wane over time. Performing both analyses over a 7-month period aims to determine the concordance between the 2 methods, particularly whether aOR aligns with adjusted risk ratio in a rare outcome situation.

**Stratified Cross-Tabulation and Pooling Effect (Mantel-Haenszel Method)**

Each outcome and exposure variable were cross-tabulated using all the stratification factors, as shown in Table 1. Eventually, 51,744 tables (2x2) per month (T) were obtained for each vaccine sequence and outcome. The Mantel-Haenszel relative risk [33] of each specific vaccine sequence was calculated using 2x2 tables. By subtracting the relative risk values of all the vaccine sequences from unity, we determined the effectiveness of the vaccine sequences for each month. All VE points were plotted against the month since the last dose of each vaccine sequence to illustrate the waning of VE\textsubscript{i} over time.

**Logistic Regression**

The primary objective of the regression analysis was to estimate the time since the last vaccination dose at which VE decreased to 50%—the minimum threshold recommended by the World Health Organization [31]. According to the variables listed in Table 1, the main independent variables of interest were vaccine sequence (V), time since 2 months after the last dose (T), and their interaction (VT). Other independent variables with their categories were all denoted as \( x \). Using the abbreviations in Table 1, we constructed the model to predict an individual’s probability of obtaining the outcome using the following formula for each vaccine sequence:

\[
\hat{\beta}_V + \hat{\beta}_X i + \hat{\beta}_0 \]

where \( i \) indexes the set of outcomes, \( \hat{\beta}_X \) is the coefficient for each category of other variables, and \( \hat{\beta}_0 \) denotes the intercept term. From the model, \( \hat{\beta}_V \) would be the natural logarithm (ln) of OR\textsubscript{i} at T=0. For each incremental month, the ln(OR\textsubscript{i}) would change by \( \hat{\beta}_VT \). Given that VE\textsubscript{i} is VE against \( Y_i \) of each vaccine sequence, the VE\textsubscript{i} for any T is A statistically significant \( \hat{\beta}_VT \) indicates a significant change in the VE\textsubscript{i} over time. A variance inflation factor of greater than 5 was considered to indicate collinearity among noninteraction variables [34]. The following formula was used to obtain the estimated time at which VE\textsubscript{i} declined to 50% (\( T_{VE=50%} \)): \( T_{VE=50%} = (\ln(0.5) – \hat{\beta}_X)/\hat{\beta}_VT \). Hence, the number of months after the last dose of each vaccine sequence at which VE\textsubscript{i} declined to 50% was \( T_{VE=50%}+2 \). Additionally, a booster dose may be required.

**Subgroup Analysis**

We conducted a subgroup analysis by age group, utilizing the logistic regression model to examine the consistency of VE durability for each age group in relation to the overall population, where \( x \) represents variables other than the age group. All analyses were performed using epiDisplay version 3.5.0.2 [35], car version 3.1 [36], and tidyverse version 1.3.1 [37] packages in R version 4.1.1 (R Core Team 2021; R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing). \( P<.05 \) was considered statistically significant.

**Ethics Approval**

This study was approved by the human research ethics committee of Songkla University (REC 65-373-18-4). The waiver of consent for accessing encrypted data was granted by the MoPH and Ministry of Interior (Thailand).

**Results**

**Participants**

A dynamic cohort of 52,580,841 individuals was established throughout the study period, beginning in May 2021. The cohort size decreased over the course of the study, as the participants exited because of SARS-CoV-2 infection (positive ATK or RT-PCR). For the 8 vaccine regimens included in our analysis, by the end of the study, 17,907,215 and 17,190,975 individuals who received 2 and 3 doses, respectively (not mutually exclusive), remained in the cohort. Thus, 45.73% (24,048,933/52,580,841) of the cohort comprised individuals who either received vaccine sequences other than the 8 primary sequences or received no vaccination at all. Table 2 describes the characteristics of individuals receiving each vaccine sequence. The columns are arranged in a chronological order based on the date on which the Thai government policy included the vaccine regimen, starting with the inactivated virus vaccine, followed by the viral vector vaccine, and finally the mRNA vaccine.
In May 2021, there were 44,424,291 nonimmunized individuals. This number decreased over time as vaccination campaigns continued. At the end of the study period, this number was reduced to 76,402. The sex distribution across all vaccine types

### Table 2. Characteristics of individuals aged ≥18 years who completed heterologous COVID-19 vaccination from May 2021 to May 2022.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>SV-AZ</th>
<th>AZ-PZ</th>
<th>SV-SV-AZ</th>
<th>SV-AZ-AZ</th>
<th>SV-AZ-PZ</th>
<th>SP-d-SP-PZ</th>
<th>AZ-AZ-PZ</th>
<th>AZ-AZ-MN</th>
<th>Nonimmunized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (N) f</td>
<td>15,949,194</td>
<td>1,958,021</td>
<td>2,403,520</td>
<td>1,862,878</td>
<td>4,703,404</td>
<td>1,951,546</td>
<td>5,166,290</td>
<td>1,103,337</td>
<td>44,424,291</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8,414,839 (52.8)</td>
<td>968,649 (49.5)</td>
<td>1,331,232 (55.4)</td>
<td>1,020,155 (54.8)</td>
<td>2,584,869 (55)</td>
<td>1,068,301 (54.7)</td>
<td>2,862,671 (55.4)</td>
<td>635,662 (57.6)</td>
<td>23,267,161 (52.4)</td>
</tr>
<tr>
<td>Male</td>
<td>7,534,355 (47.2)</td>
<td>989,372 (50.5)</td>
<td>1,072,288 (44.6)</td>
<td>842,723 (45.2)</td>
<td>2,118,535 (45)</td>
<td>883,245 (45.3)</td>
<td>2,303,619 (44.6)</td>
<td>467,675 (42.4)</td>
<td>21,157,130 (47.6)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>46.2 (16.2)</td>
<td>49.7 (16.9)</td>
<td>42.3 (12.0)</td>
<td>47.1 (16.1)</td>
<td>47.5 (16.1)</td>
<td>39.5 (14.1)</td>
<td>47.2 (16.6)</td>
<td>47.2 (16.6)</td>
<td>46.3 (17.3)</td>
</tr>
<tr>
<td><strong>Age group (years), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-40</td>
<td>6,287,541 (39.4)</td>
<td>864,171 (44.1)</td>
<td>1,040,358 (43.3)</td>
<td>681,877 (36.6)</td>
<td>165,3476 (35.2)</td>
<td>1,106,757 (56.7)</td>
<td>1,791,961 (34.7)</td>
<td>454,076 (41.2)</td>
<td>19,669,319 (44.3)</td>
</tr>
<tr>
<td>40-60</td>
<td>6,998,580 (43.9)</td>
<td>752,975 (38.5)</td>
<td>1,270,408 (52.9)</td>
<td>84,7059 (45.5)</td>
<td>2,125,381 (45.2)</td>
<td>713,478 (36.6)</td>
<td>1,895,300 (36.7)</td>
<td>376,090 (34.1)</td>
<td>17,215,748 (38.7)</td>
</tr>
<tr>
<td>60-80</td>
<td>2,502,226 (15.7)</td>
<td>309,146 (15.8)</td>
<td>89,667 (3.7)</td>
<td>314,308 (16.9)</td>
<td>875,393 (18.6)</td>
<td>121,323 (6.2)</td>
<td>1,388,652 (26.9)</td>
<td>256,191 (23.2)</td>
<td>6,970,081 (15.7)</td>
</tr>
<tr>
<td>Over 80</td>
<td>160,847 (1)</td>
<td>31,729 (1.6)</td>
<td>3087 (0.1)</td>
<td>19,634 (1)</td>
<td>49,154 (1)</td>
<td>9988 (0.5)</td>
<td>90,377 (1.7)</td>
<td>16,980 (1.5)</td>
<td>569,143 (1.3)</td>
</tr>
<tr>
<td><strong>Regional area, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangkok (capital)</td>
<td>254,872 (1.6)</td>
<td>86,010 (4.4)</td>
<td>384,195 (16)</td>
<td>38,383 (2.1)</td>
<td>83,496 (1.8)</td>
<td>95,015 (4.9)</td>
<td>1,304,126 (25.2)</td>
<td>358,327 (32.5)</td>
<td>3,675,053 (8.3)</td>
</tr>
<tr>
<td>Central</td>
<td>2,927,216 (18.4)</td>
<td>301,783 (15.4)</td>
<td>546,683 (22.7)</td>
<td>503,522 (27)</td>
<td>832,144 (17.7)</td>
<td>573,932 (29.4)</td>
<td>1,479,139 (28.6)</td>
<td>352,876 (32)</td>
<td>9,954,471 (22.4)</td>
</tr>
<tr>
<td>North</td>
<td>1,885,484 (11.8)</td>
<td>485,370 (24.8)</td>
<td>150,490 (6.3)</td>
<td>295,738 (15.9)</td>
<td>701,329 (14.9)</td>
<td>257,491 (13.2)</td>
<td>306,847 (6.3)</td>
<td>49,952 (4.5)</td>
<td>4,477,914 (10.1)</td>
</tr>
<tr>
<td>Northeast</td>
<td>6,874,988 (43.1)</td>
<td>697,679 (35.6)</td>
<td>557,002 (23.2)</td>
<td>553,144 (29.7)</td>
<td>2,343,599 (49.8)</td>
<td>522,451 (28.3)</td>
<td>1,100,220 (21.3)</td>
<td>174,410 (15.8)</td>
<td>15,085,451 (34)</td>
</tr>
<tr>
<td>East</td>
<td>1,025,189 (6.4)</td>
<td>110,809 (5.7)</td>
<td>159,441 (6.6)</td>
<td>177,323 (9.5)</td>
<td>231,917 (4.9)</td>
<td>187,462 (9.6)</td>
<td>392,273 (7.6)</td>
<td>80,371 (7.3)</td>
<td>3,226,324 (7.2)</td>
</tr>
<tr>
<td>West</td>
<td>723,149 (4.5)</td>
<td>59,807 (3)</td>
<td>118,212 (4.9)</td>
<td>113,776 (6.1)</td>
<td>118,322 (2.5)</td>
<td>88,224 (4.5)</td>
<td>207,480 (4)</td>
<td>40,709 (3.7)</td>
<td>2,059,025 (4.6)</td>
</tr>
<tr>
<td>South</td>
<td>2,258,296 (14.2)</td>
<td>216,564 (11.1)</td>
<td>487,497 (20.3)</td>
<td>180,992 (9.7)</td>
<td>392,597 (8.4)</td>
<td>196,971 (10.1)</td>
<td>376,205 (7.3)</td>
<td>46,692 (4.2)</td>
<td>5,946,053 (13.4)</td>
</tr>
</tbody>
</table>

| Interval duration between doses (weeks), median (IQR) |       |       |          |          |          |            |          |          |              |
| Interval between 1st and 2nd doses (weeks) | 3.4 (3-4) | 5.1 (4.8-8.6) | 3 (3-3.4) | 3.4 (3-4) | 3.4 (3-4) | 3 (3-4) | 12 (11.6-12) | 12 (11-12) | N/A h |
| Interval between 2nd and 3rd doses (weeks) | N/A | N/A | 16 (14-19) | 19 (16-22) | 20 (17-25) | 16 (12-20) | 17 (15-20) | 17.1 (15-20) | N/A |

aSV: Sinovac (CoronaVac).
bAZ: AstraZenenca (ChAdOx1).
cPZ: Pfizer/BioNTech (BNT162b2).
dSP: Sinopharm (BBIBP-CorV).
eMN: Moderna (mRNA-1273).
fNonimmunized population (reference) in May 2021.
gIndividuals who completed vaccination using a different vaccine sequence were not included.
hN/A: not applicable.

https://publichealth.jmir.org/2024/1/e48255
remained relatively balanced. The number of individuals in the 18-40 years and 40-60 years age groups differed across the vaccine regimens. The interval between doses 1 and 2 of the vaccine was approximately within 1 month if dose 1 was SV or SP and 2-3 months if dose 1 was AZ, as per national guidelines [25]. The duration between doses 2 and 3 varied between 1 and 6 months. Amid the emergence of the Omicron variant, PZ and MN were the predominant vaccines used.

Effectiveness of the Vaccines

The effectiveness of each heterologous vaccine regimen over time is shown in Figure 2, Figure 3, and Table 3. The estimated VEs over time from the logistic regression are illustrated in Figures S1-S2 of Multimedia Appendix 1. Table S1 of Multimedia Appendix 1 shows no significant collinearity (variance inflation factor<5) among the noninteraction variables in all logistic regression models. The 2-dose heterologous vaccination provided approximately 50% of the initial protection against severe and fatal COVID-19. After dose 2, protection against severe COVID-19 declined to less than 50% effectiveness at 5.9 months for SV-AZ and 4.7 months for AZ-PZ. The protective effect against death following COVID-19 was too low to compute a \( \text{TVE}=50\%+2 \).

All 3-dose heterologous vaccine sequences provided better protection than 2-dose heterologous vaccinations. SV-SV-AZ and SV-AZ-AZ provided persistent protection against both outcomes by over 80% and 60%, respectively, from the initial period to 7 months after dose 3 without evidence of waning. There was no statistical significance of the waning of VE against death between the SP-SP-PZ and AZ-AZ-MN groups. The \( \text{TVE}=50\%+2 \) protective effect against severe and fatal COVID-19 for each other 3-dose heterologous vaccine sequence varied from 8.7 to 14.6 months. Figures 2 and 3 demonstrate the concordance in estimating VE using logistic regression within 7 months following the last dose for all vaccine sequences, as illustrated in Figures S1-S2 of Multimedia Appendix 1.

Figure 2. Vaccine effectiveness of 2-dose vaccine sequences at preventing (A) severe and (B) fatal COVID-19 estimated using the Mantel-Haenszel risk ratio method with 95% CI ribbon from July 2021 to July 2022. For each month, the risk ratios of having the outcomes were stratified by the number of months after dose 2 and adjusted by sex, age group, province, and month of infection. AZ: AstraZeneca (ChAdOx1); PZ: Pfizer/BioNTech (BNT162b2); SV: Sinovac (CoronaVac).
Figure 3. Vaccine effectiveness of 3-dose vaccine sequences at preventing (A) severe and (B) fatal COVID-19 estimated using the Mantel-Haenszel risk ratio method with 95% CI ribbon from July 2021 to July 2022. For each month, the risk ratios of having the outcomes were stratified by the number of months after dose 3 and adjusted by sex, age groups, province, and month of infection. For SV-SV-AZ, SV-AZ-AZ, and SP-SP-PZ vaccine sequences, the vaccine effectiveness at 2 and 3 months after dose 3 was excluded due to an insufficient number of outcomes, resulting in CIs with an infinite range. AZ: AstraZeneca (ChAdOx1); MN: Moderna (mRNA-1273); PZ: Pfizer/BioNTech (BNT162b2); SP: Sinopharm (BBIBP-CorV); SV: Sinovac (CoronaVac).
### Table 3. Vaccine effectiveness of each vaccine combination at 2 and 7 months after the last dose of vaccine.

<table>
<thead>
<tr>
<th>Vaccine sequence</th>
<th>VE at 2 months after the last dose (95% CI)</th>
<th>VE at 7 months after the last dose (95% CI)</th>
<th>P value of the vaccine sequence-time interaction</th>
<th>Months since the last dose to VE=50%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2-dose vaccination</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VE against severe COVID-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV&lt;sup&gt;b&lt;/sup&gt; + AZ&lt;sup&gt;c&lt;/sup&gt;</td>
<td>55.7 (53.6-57.7)</td>
<td>47.8 (45.7-49.8)</td>
<td>&lt;.001</td>
<td>5.7</td>
</tr>
<tr>
<td>AZ + PZ&lt;sup&gt;d&lt;/sup&gt;</td>
<td>54.8 (5.3-58.8)</td>
<td>43.6 (39.2-47.7)</td>
<td>.01</td>
<td>4.3</td>
</tr>
<tr>
<td><strong>VE against fatal COVID-19</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV + AZ</td>
<td>47.4 (44.1-5.5)</td>
<td>45.8 (42.9-45.8)</td>
<td>.36</td>
<td>Omitted&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>AZ + PZ</td>
<td>54.0 (48.0-59.3)</td>
<td>5.9 (45.9-55.4)</td>
<td>.58</td>
<td>Omitted&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>3-dose vaccination</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VE against severe COVID-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV-SV-AZ</td>
<td>81.2 (78.6-83.5)</td>
<td>82.3 (8.3-84.0)</td>
<td>.55</td>
<td>Omitted&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>SV-AZ-AZ</td>
<td>76.2 (72.3-79.5)</td>
<td>6.0 (61.9-69.6)</td>
<td>.11</td>
<td>Omitted&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>SV-AZ-PZ</td>
<td>76.0 (73.7-78.4)</td>
<td>58.5 (55.0-61.8)</td>
<td>&lt;.001</td>
<td>8.7</td>
</tr>
<tr>
<td>SP&lt;sup&gt;g&lt;/sup&gt;-SP-PZ</td>
<td>79.8 (76.2-82.9)</td>
<td>68.2 (63.9-71.9)</td>
<td>.02</td>
<td>11.9</td>
</tr>
<tr>
<td>AZ-AZ-PZ</td>
<td>86.5 (85.3-87.7)</td>
<td>71.6 (69.6-73.6)</td>
<td>&lt;.001</td>
<td>1.8</td>
</tr>
<tr>
<td>AZ-AZ-MN&lt;sup&gt;h&lt;/sup&gt;</td>
<td>88.8 (86.5-9.8)</td>
<td>72.4 (68.3-75.9)</td>
<td>&lt;.001</td>
<td>1.3</td>
</tr>
<tr>
<td>VE against fatal COVID-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV-SV-AZ</td>
<td>86.2 (82.8-89.0)</td>
<td>86.3 (83.6-88.6)</td>
<td>.98</td>
<td>Omitted&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>SV-AZ-AZ</td>
<td>72.0 (66.3-76.6)</td>
<td>61.1 (55.4-66.2)</td>
<td>.22</td>
<td>Omitted&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>SV-AZ-PZ</td>
<td>77.9 (74.6-8.8)</td>
<td>6.3 (55.8-64.4)</td>
<td>.002</td>
<td>9.0</td>
</tr>
<tr>
<td>SP-SP-PZ</td>
<td>76.5 (7.8-81.1)</td>
<td>73.2 (68.5-77.3)</td>
<td>.64</td>
<td>Omitted&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>AZ-AZ-PZ</td>
<td>88.8 (87.3-9.1)</td>
<td>79.7 (77.7-81.6)</td>
<td>&lt;.001</td>
<td>14.6</td>
</tr>
<tr>
<td>AZ-AZ-MN</td>
<td>92.4 (89.8-94.3)</td>
<td>87.3 (8.4-89.7)</td>
<td>.18</td>
<td>Omitted&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>VE: vaccine effectiveness at a particular month estimated by the vaccine sequence-time interaction formula using the β coefficients adjusted for sex, age group, living province, and month of vaccination.

<sup>b</sup>SV: Sinovac (CoronaVac).

<sup>c</sup>AZ: AstraZeneca (ChAdOx1).

<sup>d</sup>PZ: Pfizer/BioNTech (BNT162b2).

<sup>e</sup>T<sub>VE=50%</sub> was omitted because the initial vaccine effectiveness was not significantly greater than 50% protective effectiveness.

<sup>f</sup>T<sub>VE=50%</sub> was omitted due to no evidence of vaccine effectiveness waning within 1 year.

<sup>g</sup>SP: Sinopharm (BBIBP-CorV).

<sup>h</sup>MN: Moderna (mRNA-1273).

### Subgroup Analysis

Figures S3-S18 of Multimedia Appendix 1 display the decline in VE over time, stratified by age group, and Table 2 provides a comprehensive summary of these figures. This table presents the effectiveness of the 2- and 3-dose COVID-19 vaccine sequences stratified by age group in relation to severe COVID-19 cases and COVID-19-related fatalities. The initial VE was compared with the VE 7 months after administration of the last dose. These findings indicate that VE is generally lower in people aged ≥80 years. The VE of distinct vaccine sequences varied based on the age group and outcome; however, the P values suggest that these differences were not statistically significant. The SV-SV-AZ regimen provided at least 6 months of durable VE against both severe and fatal COVID-19 in all age groups, including in those aged over 80 years (Figures S7-S8 of Multimedia Appendix 1). Neither of the dose regimens included in our analysis (SV-AZ and AZ-PZ) provided protective efficacy (>50% VE). Conversely, the VE of each 3-dose sequence increased (>50%) and subsequently waned over time, with a significant P value for the interaction term. Those older (age 61-80 years and ≥80 years) experienced reduction in VE to below 50% against severe or fatal COVID-19.
within 7 months of the last dose, except for the SV-SV-AZ sequence.

Discussion

Principal Findings

In Thailand, as in many other LMICs, most people receive heterologous vaccine regimens for primary series and booster doses. The main 2-dose regimens delivered in Thailand (SV-AZ and AZ-PZ) exhibited suboptimal protection against severe COVID-19, with VE decreasing to <50% after 5-7 months, as confirmed by the Mantel-Haenszel method and logistic regression. Furthermore, the initial effectiveness of these vaccines against COVID-19-related mortality was <50%. In contrast, the 3-dose regimen provided superior initial protection against both severe and fatal COVID-19, with the VE remaining >50% for at least 8 months, as determined by logistic regression with durability time-interaction modeling. However, among individuals aged ≥80 years, all vaccine combinations conferred a comparatively lower level of protection with shorter durability relative to other age groups. For the 2-dose regimen in Thailand, sequential administration of the SV-AZ vaccine was initially favored even after the availability of AZ-AZ because of a shorter recommended interdose interval based on research examining humoral immune responses [3,38]. Nevertheless, a subsequent test-negative case-control study suggested that SV-AZ was ineffective against severe COVID-19 (VE<50%) within 3 months of administering the final dose [17], which is consistent with our findings. Regarding the 3-dose regimens, other studies from Thailand similarly found heterologous vaccines to have a high VE, although they included a different set of regimens in their analysis [18].

In light of a heterologous 3-dose vaccination strategy that provides a minimum of 8 months of sustained VE against severe and fatal COVID-19 cases, the need for annual immunization has become a topic of ongoing debate [39,40], particularly given the financial burden of regular COVID-19 vaccination campaigns. Information from this study can potentially help governments identify appropriate target groups and vaccination frequencies for future COVID-19 booster vaccination policies. The durability of immunity acquired through vaccination is known to be nonuniform across different age groups due to immunosenescence [41]. Consequently, older individuals may not develop the same level of immunity against COVID-19 as younger adults, rendering vaccines less effective in preventing severe and fatal COVID-19 [42-45]. Our data show that VE for individuals aged >80 years waned after 6 months. Only the SV-SV-AZ regimen exceptionally provided more than 50% effectiveness against both severe and fatal COVID-19 over 6 months. Therefore, individuals aged >80 years receiving other regimens might need a booster dose approximately every 6 months during the pandemic era.

Limitations

This study has a few limitations that are primarily related to the use of national registry data. We were unable to identify information on unmeasured confounders such as occupations, comorbidities, and SARS-CoV-2 variants. In Thailand, health care workers were prioritized at the start of the campaign, and as it would be expected that health care workers might have different risks of developing severe COVID-19, this may have led to overestimates or underestimates of VE of the SV-AZ regimen in our study [46]. Similarly, we could not control for the comorbidities that increase the risk of developing severe COVID-19, such as diabetes and immunocompromising conditions. Individuals with these comorbidities had priority for receiving vaccination [46]. Lack of control for comorbidities may have led to an underestimation of the VE of all vaccines.

Additionally, we were unable to control for the variants of concern, which is important because there was a switch from Delta to Omicron during the study period. However, partitioning the analysis into windows for the months when Alpha and Delta variants were dominant was not possible owing to an inadequate number of cases in the vaccinated population. Most of the Thai population only received their first dose of vaccine after the Delta variant surged. Booster doses were administered mainly in 2022 during the Omicron variant–dominant period. Therefore, in this study, the evaluation of VE applies mainly to protection against the Delta variant.

Another limitation of the registry data is that it was not possible to verify whether all the individuals in the cohort were truly uninfected. Therefore, it is likely that there were many individuals with hybrid immunity and undetected natural infections in the cohort. Furthermore, this study did not account for reinfection, which might lead to a survival bias, particularly in older populations, as the remaining population is likely to be healthier [47,48]. Finally, we were unable to continue our analysis beyond July 2022 when the CO-WARD registry database ended. We also did not track individuals after any evidence of SARS-CoV-2 infection; thus, we do not have data on reinfection. The lack of reinfection data means that this study does not provide information on the protective effect of natural SARS-CoV-2 infection against subsequent infections. Given the transmissibility of the Omicron variant, which frequently causes breakthrough infections in vaccinated individuals but is less virulent [49], it is possible that by 2024, natural immunity acquired from natural infection with the Omicron variant will protect most people from reinfection. Therefore, another study is needed to clarify whether an additional booster dose is needed for individuals younger than 80 years.

Conclusion

Our findings show evidence of a significant waning of vaccine protection in Thailand, particularly for the main heterologous 2-dose regimens (SV-AZ and AZ-PZ). Overall, all 3-dose regimens demonstrated higher VE than 2-dose regimens and maintained durability of VE against severe and fatal COVID-19, with over 50% effectiveness for at least 8 months. Only the SV-SV-AZ regimen consistently showed effectiveness across all age groups. Individuals older than 80 years who received other regimens should receive a booster dose at 6 months after their last dose. Within a 7-month follow-up period, the heterologous prime-boost vaccination strategies, which involve a primary series of an inactivated virus vaccine and a booster with either a viral vector or mRNA vaccine, is proven to prevent severe and fatal COVID-19 in situations of vaccine shortages. This approach is particularly pertinent in scenarios, wherein
viral vector and mRNA vaccines are not readily available during the early stages of a pandemic. However, estimates of VE over a year are needed for vaccine regimens delivered in Thailand to inform future vaccination policies. The ongoing emergence of SARS-CoV-2 variants and the resulting herd immunity warrant the continuation of vaccine development and monitoring of VE.

Acknowledgments
The researchers would like to sincerely thank the Ministry of Public Health and the Bureau of Registration Administration, Department of Provincial Administration, Ministry of Interior for providing us with a large valuable data set. This study was partially supported by the Thailand Science Research and Innovation through the 2024 fundamental fund to the Health Intervention and Technology Assessment Program, Ministry of Public Health.

Data Availability
The aggregated data and R codes are shared by GitHub in [50]. To accommodate readers who are not familiar with R coding, we have also provided an Excel file that summarizes the results for each time period in [51].

Authors' Contributions
PK, NC, ST, and AC accessed the underlying data. PK, NC, and AC scrutinized the foundational data and took responsibility for the integrity of the data and accuracy of the data analysis. PK, YT, and VC were responsible for deciding to submit this manuscript. PK and VC conceived and designed this study. PK, NC, and AC curated the data. PK and VC designed the methods and conducted formal analyses. PK, SB, CP, YT, and VC wrote the original draft. SB, CP, NS, BL, ST, NC, AC, and YT reviewed and edited the manuscript. CP and NS were the project administrators and managed the resources. VC supervised this study. All the authors have read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File, 4289 KB - publichealth_v10i1e48255_app1.docx ]

References


against Delta and Omicron variants in prime immunized adults with two doses of the BBIBP-CorV vaccine. Vaccines (Basel) 2022 Jul 03;10(7):1071 [Full text] [doi: 10.3390/vaccines10071071] [Medline: 35891235]


37. Wickham H. tidyverse: easily install and load the Tidyverse. CRAN. URL: https://CRAN.R-project.org/package=tidyverse [accessed 2022-11-29]


Abbreviations

- aOR: adjusted odds ratio
- ATK: antigen test kit
- AZ: AstraZeneca (ChAdOx1 nCoV-19)
LMIC: low- and middle-income country

MN: Moderna (mRNA-1273)

MoPH: Ministry of Public Health

mRNA: messenger ribonucleic acid

PZ: Pfizer/BioNTech (BNT162b2)

RT-PCR: reverse transcription polymerase chain reaction

SP: Sinopharm (BBIBP-CorV)

SV: Sinovac (CoronaVac)

VE: vaccine effectiveness
Predicting COVID-19 Vaccination Uptake Using a Small and Interpretable Set of Judgment and Demographic Variables: Cross-Sectional Cognitive Science Study

Nicole L Vike, PhD; Sumra Bari, PhD; Leandros Stefanopoulos, MSc; Shamal Lalvani, MSc; Byoung Woo Kim, MSc; Nicos Maglaveras, PhD; Martin Block, PhD; Hans C Breiter, MD; Byoung Woo Kim, MSc; Nicos Maglaveras, PhD; Martin Block, PhD; Hans C Breiter, MD; Aggelos K Katsaggelos, PhD

1Department of Computer Science, University of Cincinnati, Cincinnati, OH, United States
2Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, United States
3School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece
4Integrated Marketing Communications, Medill School, Northwestern University, Evanston, IL, United States
5Department of Psychiatry, Massachusetts General Hospital, Harvard School of Medicine, Boston, MA, United States
6Department of Computer Science, Northwestern University, Evanston, IL, United States
7Department of Radiology, Northwestern University, Evanston, IL, United States

*these authors contributed equally

Corresponding Author:
Hans C Breiter, MD
Department of Computer Science
University of Cincinnati
2901 Woodside Drive
Cincinnati, OH, 45219
United States
Phone: 1 617 413 0953
Email: breitehs@ucmail.uc.edu

Abstract

Background: Despite COVID-19 vaccine mandates, many chose to forgo vaccination, raising questions about the psychology underlying how judgment affects these choices. Research shows that reward and aversion judgments are important for vaccination choice; however, no studies have integrated such cognitive science with machine learning to predict COVID-19 vaccine uptake.

Objective: This study aims to determine the predictive power of a small but interpretable set of judgment variables using 3 machine learning algorithms to predict COVID-19 vaccine uptake and interpret what profile of judgment variables was important for prediction.

Methods: We surveyed 3476 adults across the United States in December 2021. Participants answered demographic, COVID-19 vaccine uptake (ie, whether participants were fully vaccinated), and COVID-19 precaution questions. Participants also completed a picture-rating task using images from the International Affective Picture System. Images were rated on a Likert-type scale to calibrate the degree of liking and disliking. Ratings were computationally modeled using relative preference theory to produce a set of graphs for each participant (minimum R²>0.8). In total, 15 judgment features were extracted from these graphs, 2 being analogous to risk and loss aversion from behavioral economics. These judgment variables, along with demographics, were compared between those who were fully vaccinated and those who were not. In total, 3 machine learning approaches (random forest, balanced random forest [BRF], and logistic regression) were used to test how well judgment, demographic, and COVID-19 precaution variables predicted vaccine uptake. Mediation and moderation were implemented to assess statistical mechanisms underlying successful prediction.

Results: Age, income, marital status, employment status, ethnicity, educational level, and sex differed by vaccine uptake (Wilcoxon rank sum and chi-square P<.001). Most judgment variables also differed by vaccine uptake (Wilcoxon rank sum P<.05). A similar area under the receiver operating characteristic curve (AUROC) was achieved by the 3 machine learning frameworks, although random forest and logistic regression produced specificities between 30% and 38% (vs 74.2% for BRF), indicating a lower performance in predicting unvaccinated participants. BRF achieved high precision (87.8%) and AUROC (79%) with moderate to high accuracy (70.8%) and balanced recall (69.6%) and specificity (74.2%). It should be noted that, for BRF,
the negative predictive value was <50% despite good specificity. For BRF and random forest, 63% to 75% of the feature importance came from the 15 judgment variables. Furthermore, age, income, and educational level mediated relationships between judgment variables and vaccine uptake.

Conclusions: The findings demonstrate the underlying importance of judgment variables for vaccine choice and uptake, suggesting that vaccine education and messaging might target varying judgment profiles to improve uptake. These methods could also be used to aid vaccine rollouts and health care preparedness by providing location-specific details (eg, identifying areas that may experience low vaccination and high hospitalization).

(JMIR Public Health Surveill 2024;10:e47979) doi:10.2196/47979

KEYWORDS
reward; aversion; judgment; relative preference theory; cognitive science; behavioral economics; machine learning; balanced random forest; mediation; moderation; mobile phone; smartphone

Introduction

Background

In early 2020, the COVID-19 pandemic wreaked havoc worldwide, triggering rapid vaccine development efforts. Despite federal, state, and workplace vaccination mandates, many individuals made judgments against COVID-19 vaccination, leading researchers to study the psychology underlying individual vaccination preferences and what might differentiate the framework for judgment between individuals who were fully vaccinated against COVID-19 and those who were not (henceforth referred to as vaccine uptake). A better understanding of these differences in judgment may highlight targets for public messaging and education to increase the incidence of choosing vaccination.

Prior Work

Multiple studies have sought to predict an individual’s intention to receive a COVID-19 vaccine or specific variables underlying vaccination choices or mitigation strategies [1-7], but few have predicted vaccine uptake. One such study used 83 sociodemographic variables (with education, ethnicity, internet access, income, longitude, and latitude being the most important predictors) to predict vaccine uptake with 62% accuracy [8], confirming both the importance and limitations of these variables in prediction models. Other studies have compared demographic groups between vaccinated and nonvaccinated persons; Bulusu et al [9] found that young adults (aged 18-35 years), women, and those with higher levels of education had higher odds of being vaccinated. In a study of >12 million persons, the largest percentage of those who initiated COVID-19 vaccination were White, non-Hispanic women between the ages of 50 and 64 years [10]. Demographic variables are known to affect how individuals judge what is rewarding or aversive [11,12] yet are not themselves variables quantifying how individuals make judgments that then frame decisions.

Judgment reflects an individual’s preferences, or the variable extent to which they approach or avoid events in the world based on the rewarding or aversive effects of these events [13-15]. The definition of preference in psychology differs from that in economics. In psychology, preferences are associated with “wanting” and “liking” and are framed by judgments that precede decisions, which can be quantified through reinforcement reward or incentive reward tasks [12,16-21]. In economics, preferences are relations derived from consumer choice data (refer to the axioms of revealed preference [22]) and reflect choices or decisions based on judgments that place value on behavioral options. Economist Paul Samuelson noted that decisions are “assumed to be correlative to desire or want” [23]. In this study, we focused on a set of variables that frame judgment, with the presumption that judgments precede choices [12,20]. Variables that frame judgment can be derived from tasks using operant key-pressing tasks that quantify “wanting” [24-33] or simple rating tasks that are analogous to “liking” [20,34]. Both operant keypress and rating tasks measure variables that quantify the average (mean) magnitude (K), variance (σ), and pattern (ie, Shannon entropy [H]) of reward and aversion judgments [35]. We refer to this methodology and the multiple relationships between these variables and features based on their graphical relationships as relative preference theory (RPT; Figure 1) [18,36]. RPT has been shown to produce discrete, recurrent, robust, and scalable relationships between judgment variables [37] that produce mechanistic models for prediction [33], and which have demonstrated relationships to brain circuitry [24-27,30] and psychiatric illness [28]. Of the graphs produced for RPT, 2 appear to resemble graphs derived with different variables in economics, namely, prospect theory [38] and the mean-variance function for portfolio theory described by Markowitz [39]. Given this graphical resemblance, it is important to note that RPT functions quantifying value are not the same as standard representations of preference in economics. Behavioral economic variables such as loss aversion and risk aversion [38,40-51] are not to be interpreted in the same context given that both reflect biases and bounds to human rationality. In psychology, they are grounded in judgments that precede decisions, whereas in economics, they are grounded in consumer decisions themselves. Going forward, we will focus on judgment-based loss aversion, representing the overweighting of negative judgments relative to positive ones, and judgment-based risk aversion, representing the preference for small but certain assessments over larger but less certain ones (ie, assessments that have more variance associated with them) [38,40-51]. Herein, loss aversion and risk aversion refer to ratings or judgments that precede decisions.

A number of studies have described how risk aversion and other judgment variables are important for individual vaccine choices and hesitancies [52-58]. Hudson and Montelpare [54] found that risk aversion may promote vaccine adherence when people
perceive contracting a disease as more dangerous or likely. Trueblood et al [52] noticed that those who were more risk seeking (as measured via a gamble ladder task) were more likely to receive the vaccine even if the vaccine was described as expedited. Wagner et al [53] described how risk misperceptions (when the actual risk does not align with the perceived risk) may result from a combination of cognitive biases, including loss aversion. A complex theoretical model using historical vaccine attitudes grounded in decision-making has also been proposed to predict COVID-19 vaccination, but this model has not yet been tested [59]. To our knowledge, no study has assessed how well a model comprising variables that reflect reward and aversion judgments predicts vaccine uptake.

Figure 1. Picture-rating task and judgment variable extraction. (A) An example picture from the picture-rating task in which participants were asked to rate how much they liked or disliked an image on a scale from −3 (dislike very much) to +3 (like very much), with 0 being neutral. (B) Visual representation of the x-y plane for relative preference theory (RPT) value function fitting and resulting features extracted. (C) Visual representation of the x-y plane for RPT limit function fitting and resulting features extracted. (D) Visual representation of the x-y plane for RPT trade-off function fitting and resulting features extracted. (E) Each of the 15 features and their abbreviated terms.

Goal of This Study

Given the many vaccine-related issues that occurred during the COVID-19 pandemic (eg, vaccine shortages, hospital overload, and vaccination resistance or hesitancy), it is critical to develop methods that might improve planning around such shortcomings. Because judgment variables are fundamental to vaccine choice, they provide a viable target for predicting vaccine uptake. In addition, the rating methodology used to quantify variables of judgment is independent of methods quantifying vaccine uptake or intent to vaccinate, limiting response biases within the study data.

In this study, we aimed to predict COVID-19 vaccine uptake using judgment, demographic, and COVID-19 precaution (ie, behaviors minimizing potential exposure to COVID-19) variables using multiple machine learning algorithms, including logistic regression, random forest, and balanced random forest (BRF). BRF was hypothesized to perform best given its potential benefits with handling class imbalances [60], balancing both recall and specificity, and producing Gini scores that provide relative variable importance to prediction. In this study, the need for data imbalance techniques was motivated by the importance of the specificity metric, which would reflect the proportion of participants who did not receive full vaccination; without balancing, the model might not achieve similar recall and specificity values. When there is a large difference between recall and specificity, specificity might instead reflect the size of the minority class (those who did not receive full vaccination). In general, random forest approaches have been reported to have benefits over other approaches such as principal component analysis and neural networks, in which the N-dimensional feature space or layers (in the case of neural networks) are complex nonlinear functions, making it difficult to interpret variable importance and relationships to the outcome variable. To provide greater certainty about these assumptions, we performed logistic regression in parallel with random forest and BRF. The 3 machine learning approaches used a small feature set (<20) with interpretable relationships to the predicted variable. Such interpretations may not be achievable in big data approaches that use hundreds to thousands of variables that seemingly add little significance to the prediction models. Interpretation was facilitated by (1) the Gini importance criterion associated with BRF and random forest, which provided a profile of the judgment variables most important for prediction; and (2) mediation and moderation analyses that offered insights into statistical mechanisms among judgment variables, demographic (contextual) variables, and vaccine uptake. Determining whether judgment variables are predictive of COVID-19 vaccine uptake and defining which demographic variables facilitate this prediction presents a number of behavioral targets for vaccine education and messaging—and
potentially identifies actionable targets for increasing vaccine uptake.

More broadly, the prediction of vaccine uptake may aid (1) vaccine supply chain and administration logistics by indicating areas that may need more or fewer vaccines, (2) targeted governmental messaging to locations with low predicted uptake, and (3) preparation of areas that may experience high cases of infection that could ultimately impact health care preparedness and infrastructure. The proposed method could also be applied to other mandated or government-recommended vaccines (eg, influenza and human papillomavirus) to facilitate the aforementioned logistics. Locally, vaccine uptake prediction could facilitate local messaging and prepare health care institutions for vaccine rollout and potential hospital overload. Nationally, prediction might inform public health officials and government communication bodies that are responsible for messaging and vaccine rollout with the goal of improving vaccine uptake and limiting infection and hospital overload.

Methods

Recruitment

Similar recruitment procedures for a smaller population-based study have been described previously [61-63]. In this study, participants were randomly sampled from the general US population using an email survey database used by Gold Research, Inc. Gold Research administered questionnaires in December 2021 using recruitment formats such as (1) customer databases from large companies that participate in revenue-sharing agreements, (2) social media, and (3) direct mail. Recruited participants followed a double opt-in consent procedure that included primary participation in the study as well as secondary use of anonymized, deidentified data (ie, all identifying information was removed by Gold Research before retrieval by the research group) in secondary analyses (refer to the Ethical Considerations section for more detail). During consent procedures, participants provided demographic information (eg, age, race, and sex) to ensure that the sampled participants adequately represented the US census at the time of the survey (December 2021). Respondents were also presented with repeated test questions to screen out those providing random and illogical responses or showing flatline or speeder behavior. Participants who provided such data were flagged, and their data were removed.

Because other components of the survey required an adequate sample of participants with mental health conditions, Gold Research oversampled 15% (60,000/400,000) of the sample for mental health conditions, and >400,000 respondents were contacted to complete the questionnaire. Gold Research estimated that, of the 400,000 participants, >300,000 (≥75%) either did not respond or declined to participate. Of the remaining 25% (100,000/400,000) who clicked on the survey link, >50% (52,000/100,000) did not fully complete the questionnaire. Of the ≥48,000 participants who completed the survey (ie, ≥48,000/400,000, ≥12% of the initial pool of queried persons), those who did not clear data integrity assessments were omitted. Participants who met quality assurance procedures (refer to the following section) were selected, with a limit of 4000 to 4050 total participants.

Eligible participants were required to be aged between 18 and 70 years at the time of the survey, comprehend the English language, and have access to an electronic device (eg, laptop or smartphone).

Ethical Considerations

All participants provided informed consent, which included their primary participation in the study as well as the secondary use of their anonymized, deidentified data (ie, all identifying information removed by Gold Research before retrieval by the research group) in secondary analyses. This study was approved by the Northwestern University institutional review board (approval STU00213665) for the initial project start and later by the University of Cincinnati institutional review board (approval 2023-0164) as some Northwestern University investigators moved to the University of Cincinnati. All study approvals were in accordance with the Declaration of Helsinki. All participants were compensated with US $10 for taking part. Detailed survey instructions have been published previously [61-63].

Quality Assurance and Data Exclusion

Three additional quality assurance measures were used to flag nonadhering participants: (1) participants who indicated that they had ≥10 clinician-diagnosed illnesses (refer to Figure S1 in Multimedia Appendix 1 [18,33,36,64-68] for a list), (2) participants who showed minimal variance in the picture-rating task (ie, all pictures were rated the same or the ratings varied only by 1 point; refer to the Picture-Rating Task section), and (3) inconsistencies between educational level and years of education and participants who completed the questionnaire in <800 seconds.

Data from 4019 participants who passed the initial data integrity assessments were anonymized and then sent to the research team. Data were further excluded if the quantitative feature set derived from the picture-rating task was incomplete or if there were extreme outliers (refer to the RPT Framework section). Using these exclusion criteria, of the 4019 participants, 3476 (86.49%) were cleared for statistical analysis, representing 0.87% (3476/400,000) of the initial recruitment pool. A flowchart of participant exclusion is shown in Figure 2.
Questionnaire
Participants were asked to report their age, sex, ethnicity, annual household income, marital status, employment status, and educational level. Participants were asked to report whether they had received the full vaccination (yes or no responses). At the time of the survey, participants were likely to have received either 2 doses of the Pfizer or Moderna vaccine or 1 dose of the Johnson & Johnson vaccine as per the Centers for Disease Control and Prevention guidelines. Participants were also asked to respond yes (they routinely followed the precaution) or no (they did not routinely follow the precaution) to 4 COVID-19 precaution behaviors: mask wearing, social distancing, washing or sanitizing hands, and not gathering in large groups (refer to Tables S1 and S2 in Multimedia Appendix 1 for the complete questions and sample sizes, respectively). In addition, participants completed a picture-rating task at 2 points during the survey (refer to the Picture-Rating Task section).

Picture-Rating Task
A picture-rating task was administered to quantify participants’ degree of liking and disliking a validated picture set using pictures calibrated over large samples for their emotional intensity and valence [69,70]. Ratings from this task have been mathematically modeled using RPT to define graphical features of reward and aversion judgments. Each feature quantifies a core aspect of judgment, including risk aversion and loss aversion. Judgment variables have been shown to meet the criteria for lawfulness [37] that produce mechanistic models for prediction [33], with published relationships to brain circuitry [24-27,30] and psychiatric illness [28]. A more complete description of these judgment variables and their computation can be found in the RPT Framework section and in Table 1.

For this task, participants were shown 48 unique color images from the International Affective Picture System [69,70]. A total of 6 picture categories were used: sports, disasters, cute animals, aggressive animals, nature (beach vs mountains), and men and women dressed minimally, with 8 pictures per category (48 pictures in total; Figure 1A). These images have been used and validated in research on human emotion, attention, and preferences [69,70]. The images were displayed on the participants’ digital devices with a maximum size of 1204 × 768 pixels. Below each picture was a rating scale from −3 (dislike very much) to +3 (like very much), where 0 indicated indifference (Figure 1A). While there was no time limit for selecting a picture rating, participants were asked to rate the images as quickly as possible and use their first impression. Once a rating was selected, the next image was displayed.
Table 1. Judgment variable descriptions. Function and curve for derivation indicates which function space was used to derive each judgment variable (refer to the work by Markowitz [39,71]; equations are also discussed in the study by Kim et al [18]).

<table>
<thead>
<tr>
<th>Judgment variable</th>
<th>Function and curve for derivation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss aversion</td>
<td>Value function and (K,H) curve</td>
<td>The degree to which one overweighs negative stimuli against positive stimuli</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>Value function and (K,H) curve</td>
<td>The degree to which one prefers an uncertain high-value outcome to something certain but of a lower value</td>
</tr>
<tr>
<td>Loss resilience</td>
<td>Value function and (K,H) curve</td>
<td>Measures one’s preference to accept a certain loss over an uncertain loss. It is similar to risk aversion but in the domain of losses.</td>
</tr>
<tr>
<td>Ante</td>
<td>Value function and (K,H) curve</td>
<td>What one is willing to pay to enter a game of chance (eg, poker)</td>
</tr>
<tr>
<td>Insurance</td>
<td>Value function and (K,H) curve</td>
<td>The amount of security one is willing to acquire to avoid negative outcomes</td>
</tr>
<tr>
<td>Peak positive risk</td>
<td>Limit function and (K, σ) curve</td>
<td>Per the decision utility equation by Markowitz [39,71], this is the peak risk regarding approach choices that must be overcome for approach behavior to occur.</td>
</tr>
<tr>
<td>Peak negative risk</td>
<td>Limit function and (K, σ) curve</td>
<td>Per the decision utility equation by Markowitz [39,71], this is the peak risk regarding avoidance choices that must be overcome for avoidance behavior to occur.</td>
</tr>
<tr>
<td>Reward tipping point</td>
<td>Limit function and (K, σ) curve</td>
<td>Per the decision utility equation by Markowitz [39,71], this is the reward value beyond which approach choices are made.</td>
</tr>
<tr>
<td>Aversion tipping point</td>
<td>Limit function and (K, σ) curve</td>
<td>Per the decision utility equation by Markowitz [39,71], this is the intensity of aversion beyond which avoidance choices are made.</td>
</tr>
<tr>
<td>Total reward risk</td>
<td>Limit function and (K, σ) curve</td>
<td>Total value of reward across the range of risks associated with those positive outcomes</td>
</tr>
<tr>
<td>Total aversion risk</td>
<td>Limit function and (K, σ) curve</td>
<td>The total amount of aversion across the range of risks associated with those negative outcomes</td>
</tr>
<tr>
<td>Reward-aversion trade-off</td>
<td>Value function and (H_H) curve</td>
<td>This angle represents the average bias of information toward approach or avoidance behavior.</td>
</tr>
<tr>
<td>Trade-off range</td>
<td>Value function and (H_H) curve</td>
<td>The variance or bias toward approach versus avoidance behavior and one metric of the range in a person's portfolio of preference</td>
</tr>
<tr>
<td>Reward-aversion consistency</td>
<td>Value function and (H_H) curve</td>
<td>A continuum between how much an individual has conflict versus indifference in their reward-aversion preference—where conflict means that they both like and dislike something and indifference means that they do not like or dislike something</td>
</tr>
<tr>
<td>Consistency range</td>
<td>Value function and (H_H) curve</td>
<td>How much a person swings between conflict and indifference in their preferences; it is a second metric regarding the range in a person's portfolio of preference</td>
</tr>
</tbody>
</table>

RPT Framework

Ratings from the picture-rating task were analyzed using an RPT framework. This framework fits approach and avoidance curves and derives mathematical features from graphical plots (Figures 1B–1D). These methods have been described at length in prior work and are briefly described in this section [11,18,33,36]. More complete descriptions and quality assurance procedures can be found in Multimedia Appendix 1.

At least 15 judgment variables can be mathematically derived from this framework and are psychologically interpretable; they have been validated using both operand keypresses [9,25-27] and picture-rating tasks [11,34]. The 15 judgment variables are loss aversion, risk aversion, loss resilience, ante, insurance, peak positive risk, peak negative risk, reward tipping point, aversion tipping point, total reward risk, total aversion risk, reward-aversion trade-off, trade-off range, reward-aversion consistency, and consistency range. Loss aversion, risk aversion, loss resilience, ante, and insurance are derived from the logarithmic or power-law fit of mean picture ratings (K) versus entropy of ratings (H); this is referred to as the value function (Figure 1B). Peak positive risk, peak negative risk, reward tipping point, aversion tipping point, total reward risk, and total aversion risk are derived from the quadratic fit of K versus the SD of picture ratings (σ); this is referred to as the limit function (Figure 1C). Risk aversion trade-off, trade-off range, risk aversion consistency, and consistency range are derived from the radial fit of the pattern of avoidance judgments (H_L) versus the pattern of approach judgments (H_A); this is referred to as the trade-off function (Figure 1D). Value (Figure S2A in Multimedia Appendix 1), limit (Figure S2B in Multimedia Appendix 1), and trade-off (Figure S2C in Multimedia Appendix 1) functions were plotted for 500 randomly sampled participants, and nonlinear curve fits were assessed for goodness of fit, yielding R², adjusted R², and the associated F statistic for all participants (Figure S2D in Multimedia Appendix 1). Only the logarithmic and quadratic fits are listed in Table S3 in Multimedia Appendix 1. Each feature describes a quantitative component of a participant’s reward and aversion judgment (refer to Table 1 for abbreviated descriptions and Multimedia Appendix 1 for complete descriptions). Collectively, the 15 RPT features will be henceforth referred to as “judgment variables.” The summary statistics for these variables can be found in Table S3 in Multimedia Appendix 1.
Statistical and Machine Learning Analyses

Overview

Wilcoxon rank sum tests, chi-square tests, and Gini importance plotting were performed in Stata (version 17; StataCorp) [72]. Machine learning algorithms were run in Python (version 3.9; Python Software Foundation) [73], where the scikit-learn (version 1.2.2) [74] and imbalanced-learn (version 0.10.1) [75] libraries were used. Post hoc mediation and moderation analyses were performed in R (version 4.2.0; R Foundation for Statistical Computing) [76].

Demographic and Judgment Variable Differences by Vaccination Uptake

Each of the 7 demographic variables (age, income, marital status, employment status, ethnicity, educational level, and sex) was assessed for differences using yes or no responses to receiving the full COVID-19 vaccination (2525/3476, 72.64% yes responses and 951/3476, 27.36% no responses), henceforth referred to as vaccine uptake. Ordinal (income and educational level) and continuous (age) demographic variables were analyzed using the Wilcoxon rank sum test (α=.05). Expected and actual rank sums were reported using Wilcoxon rank sum tests. Nominal variables were analyzed using the chi-square test (α=.05). For significant chi-square results, demographic response percentages were computed to compare the fully vaccinated and not fully vaccinated groups.

Each of the 15 judgment variables was assessed for differences across yes or no responses to vaccine uptake using the Wilcoxon rank sum test (α=.05). The expected and actual rank sums were reported. Significant results (α<.05) were corrected for multiple comparisons using the Benjamini-Hochberg correction, and Q values of <0.05 (Q_{B-H}) were reported.

Prediction Analyses

Logistic regression, random forest, and BRF were used to predict vaccine uptake using judgment, demographic, and COVID-19 precaution variables. Gini plots were produced for random forest and BRF to determine the importance of the judgment variables in predicting COVID-19 vaccination. The BRF algorithm balances the samples by randomly downsampling the majority class at each bootstrapped iteration to match the number of samples in the minority class. To provide greater certainty about the results, random forest and logistic regression were performed to compare with BRF results.

Two sets of BRF, random forest, and logistic regression analyses were run: (1) with the 7 demographic variables and 15 judgment variables included as predictors and (2) with the 7 demographic variables, 15 judgment variables, and 4 COVID-19 precaution behaviors included as predictors. COVID-19 precaution behaviors included yes or no responses to wearing a mask, social distancing, washing hands, and avoiding large gatherings (refer to Table S1 in Multimedia Appendix 1 for more details). The sample sizes for yes or no responses to the COVID-19 precaution behavior questions are provided in Table S2 in Multimedia Appendix 1. For all 3 models, 10-fold cross-validation was repeated 100 times to obtain performance metrics, where data were split for training (90%) and testing (10%) for each of the 10 iterations in cross-validation. The averages of the performance metrics were reported across 100 repeats of 10-fold cross-validation for the test sets. The reported metrics included accuracy, recall, specificity, negative predictive value (NPV), precision, and area under the receiver operating characteristic curve (AUROC). For BRF, the Python toolbox imbalanced-learn was used to build the classifier, where the training set for each iteration of cross-validation was downsamplt but the testing set was unchanged (ie, imbalanced). That is, downsampling only occurred with the bootstrapped samples for training the model, and balancing was not performed on the testing set. The default number of estimators was 100, and the default number of tree splits was 10; the splits were created using the Gini criterion. In separate analyses, estimators were increased to 300, and splits were increased to 15 to test model performance. Using the scikit-learn library, the same procedures used for BRF were followed for logistic regression without downsampling. Logistic regression without downsampling was implemented with a maximum of 100 iterations and optimization using a limited-memory Broyden-Fletcher-Goldfarb-Shanno solver. For logistic regression, model coefficients with respective SEs, z statistics, P values, and 95% CIs were reported.

Relative feature importance based on the Gini criterion (henceforth referred to as Gini importance) was determined from BRF and random forest using the .feature_importances_ attribute from scikit-learn, and results were reported as the mean decrease in the Gini score and plotted in Stata. To test model performance using only the top predictors, two additional sets of BRF analyses were run: (1) with the top 3 features as predictors and (2) with the top 3 features and 15 judgment variables as predictors.

Post Hoc Mediation and Moderation

Given the importance of both judgment variables and demographic variables (refer to the Results section), we evaluated post hoc how age, income, and educational level (ie, the top 3 predictors) might statistically influence the relationship between the 15 judgment variables and COVID-19 vaccine uptake. To identify statistical mechanisms influencing our prediction results, we used mediation and moderation, which can (1) determine the directionality between variables and (2) assess variable influence in statistical relationships. Mediation is used to determine whether one variable, the mediator, statistically improves the relationship between 2 other variables (independent variables [IVs] and dependent variables [DVs]) [77-80]. When mediating variables improve a relationship, the mediator is said to sit in the statistical pathway between the IVs and DVs [77,80,81]. Moderation is used to test whether the interaction between an IV and a moderating variable predicts a DV [81,82].

For mediation, primary and secondary mediations were performed. Primary mediations included each of the 15 judgment behaviors as the IV, each of the 3 demographic variables (age, income, and educational level) as the mediator, and vaccine uptake as the DV. Secondary mediations held the 15 judgment behaviors as the mediator, the 3 demographic variables as the IV, and vaccine uptake as the DV. For moderation, the
moderating variable was each of the 3 demographic variables (age, income, and educational level), the IV was each of the 15 judgment behaviors, and the DV was vaccine uptake. The mathematical procedures for mediation and moderation can be found in Multimedia Appendix 1.

Results

Demographic Assessment

Of the 400,000 persons queried by Gold Research, Inc, 48,000 (12%) completed the survey, and 3476 (0.87%) survived all quality assurance procedures. Participants were predominately female, married, and White individuals; employed full time with some college education; and middle-aged (mean age 51.40, SD 14.92 years; Table 2). Of the 3476 participants, 2525 (72.64%) reported receiving a full dose of a COVID-19 vaccine, and 951 (27.36%) reported not receiving a full dose. Participants who indicated full vaccination were predominately female, married, White individuals, and retired; had some college education; and were older on average (mean age 54.19, SD 14.13 years) when compared to the total cohort. Participants who indicated that they did not receive the full vaccine were also predominately female, married, and White individuals. In contrast to those who received the full vaccination, those not fully vaccinated were predominately employed full time, high school graduates, and of average age (mean age 43.98, SD 14.45 years; median age 45, IQR 32-56 years) when compared to the total cohort. Table 2 summarizes the demographic group sample size percentages for the total cohort, those fully vaccinated, and those not fully vaccinated.

When comparing percentages between vaccination groups, a higher percentage of male individuals were fully vaccinated, and a higher percentage of female individuals were not fully vaccinated (Table 2). In addition, a higher percentage of married, White and Asian or Pacific Islander, and retired individuals indicated receiving the full vaccine when compared to the percentages of those who did not receive the vaccine (Table 2). Conversely, a higher percentage of single, African American, and unemployed individuals indicated not receiving the full vaccine (Table 2).
Table 2. Demographic summary of the sample (N=3476).

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Vaccine uptake</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fully vaccinated (n=2525)</td>
<td>Not fully vaccinated (n=951)</td>
</tr>
<tr>
<td>Age (y), median (IQR)</td>
<td>59 (44-46)</td>
<td>45 (32-56)</td>
</tr>
<tr>
<td>Income level (US $), n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25,000</td>
<td>338 (13.4)</td>
<td>288 (30.3)</td>
</tr>
<tr>
<td>25,000-50,000</td>
<td>621 (24.6)</td>
<td>285 (30)</td>
</tr>
<tr>
<td>50,000-75,000</td>
<td>500 (19.8)</td>
<td>194 (20.4)</td>
</tr>
<tr>
<td>75,000-100,000</td>
<td>428 (17)</td>
<td>98 (10.3)</td>
</tr>
<tr>
<td>100,000-150,000</td>
<td>388 (15.4)</td>
<td>57 (6)</td>
</tr>
<tr>
<td>150,000-300,000</td>
<td>204 (8.1)</td>
<td>26 (2.7)</td>
</tr>
<tr>
<td>&gt;300,000</td>
<td>46 (1.8)</td>
<td>3 (0.3)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1038 (41.1)</td>
<td>286 (30.1)</td>
</tr>
<tr>
<td>Female</td>
<td>1478 (58.5)</td>
<td>660 (69.4)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>9 (0.4)</td>
<td>5 (0.5)</td>
</tr>
<tr>
<td>Race and ethnicity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>109 (4.3)</td>
<td>13 (1.4)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>126 (5)</td>
<td>101 (10.6)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>89 (3.5)</td>
<td>34 (3.6)</td>
</tr>
<tr>
<td>Native American or Alaska Native</td>
<td>11 (0.4)</td>
<td>15 (1.6)</td>
</tr>
<tr>
<td>White</td>
<td>2152 (85.2)</td>
<td>756 (79.5)</td>
</tr>
<tr>
<td>Mixed</td>
<td>15 (0.6)</td>
<td>15 (1.6)</td>
</tr>
<tr>
<td>Other</td>
<td>9 (0.4)</td>
<td>7 (0.7)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>14 (0.6)</td>
<td>10 (1.1)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>478 (18.9)</td>
<td>288 (30.3)</td>
</tr>
<tr>
<td>Married</td>
<td>1426 (56.5)</td>
<td>362 (38.1)</td>
</tr>
<tr>
<td>Divorced</td>
<td>296 (11.7)</td>
<td>120 (12.6)</td>
</tr>
<tr>
<td>Separated</td>
<td>36 (1.4)</td>
<td>28 (2.9)</td>
</tr>
<tr>
<td>Widowed</td>
<td>129 (5.1)</td>
<td>23 (2.4)</td>
</tr>
<tr>
<td>Living with partner</td>
<td>144 (5.7)</td>
<td>125 (13.1)</td>
</tr>
<tr>
<td>Other or prefer not to answer</td>
<td>16 (0.6)</td>
<td>5 (0.5)</td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>254 (10.1)</td>
<td>227 (23.9)</td>
</tr>
<tr>
<td>Employed full time</td>
<td>906 (35.9)</td>
<td>338 (35.5)</td>
</tr>
<tr>
<td>Employed part time</td>
<td>226 (9)</td>
<td>86 (9)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>97 (3.8)</td>
<td>68 (7.2)</td>
</tr>
<tr>
<td>&gt;1 job</td>
<td>3 (0.1)</td>
<td>4 (0.4)</td>
</tr>
<tr>
<td>Retired</td>
<td>931 (36.9)</td>
<td>148 (15.6)</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>108 (4.3)</td>
<td>80 (8.4)</td>
</tr>
<tr>
<td>Educational level, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some high school</td>
<td>44 (1.7)</td>
<td>57 (6)</td>
</tr>
</tbody>
</table>
| High school graduate          | 410 (16.2)     | 324 (34.1)     | 734 (21.1)            |https://publichealth.jmir.org/2024/1/e47979
Analysis of Machine Learning Features

**Demographic Variable Differences by Vaccine Uptake**

Age, income level, and educational level significantly differed between those who did and did not receive the vaccine (Wilcoxon rank sum test $\alpha<.05$; Table 3). Those who indicated full vaccination were, on average, older (median age 59 y), had a higher annual household income (median reported income level US $50,000-$75,000), and had higher levels of education (the median reported educational level was a bachelor’s degree). Chi-square tests revealed that marital status, employment status, sex, and ethnicity also varied by full vaccine uptake (chi-square $\alpha<.05$; Table 3).

Table 3. Wilcoxon rank sum and chi-square test results ($\alpha=.05$). Expected and actual rank sums for each vaccine uptake group are reported. “Higher” indicates that the actual rank sum was higher than expected, and “lower” indicates that the actual rank sum was lower than expected.

<table>
<thead>
<tr>
<th>Demographic variable and full vaccination status</th>
<th>Rank sum</th>
<th>Expected</th>
<th>Higher or lower</th>
<th>Test</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,873,514.5</td>
<td>4,389,712.5</td>
<td>Higher</td>
<td>Rank sum</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>1,169,511.5</td>
<td>1,653,313.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,770,430.5</td>
<td>4,389,712.5</td>
<td>Higher</td>
<td>Rank sum</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>1,272,595.5</td>
<td>1,653,313.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Chi-square</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Chi-square</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Chi-square</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,794,497.5</td>
<td>4,389,712.5</td>
<td>Higher</td>
<td>Rank sum</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>1,248,528.5</td>
<td>1,653,313.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>Chi-square</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

-JMIR PUBLIC HEALTH AND SURVEILLANCE-

aN/A: not applicable.

**Judgment Variable Differences by Vaccine Uptake**

In total, 10 of the 15 judgment variables showed nominal rank differences ($\alpha<.05$), and 9 showed significant rank differences after correction for multiple comparisons ($Q_{Hoch}<0.05$) between those who indicated full vaccination and those who indicated that they did not receive the full vaccination (Table 4). The 10 features included loss aversion, risk aversion, loss resilience, ante, insurance, peak positive risk, peak negative risk, total reward risk, total aversion risk, and trade-off range. Those who indicated full vaccination exhibited lower loss aversion, ante, peak positive risk, peak negative risk, total reward risk, and total aversion risk as well as higher risk aversion, loss resilience, insurance, and trade-off range when compared to the expected rank sum. Those who did not receive the full vaccination exhibited lower risk aversion, loss resilience, insurance, and trade-off range and higher loss aversion, ante, peak positive risk, peak negative risk, total reward risk, and total aversion risk when compared to the expected rank sum.
Table 4. Judgment variable differences by vaccine uptake (yes vs no responses to the full vaccination question). Expected and actual rank sums for each vaccine uptake group are reported from Wilcoxon rank sum tests ($\alpha=.05$). “Higher” indicates that the actual rank sum was higher than expected, and “lower” indicates that the actual rank sum was lower than expected.

<table>
<thead>
<tr>
<th>Judgment variable and full vaccination uptake</th>
<th>Rank sum</th>
<th>Expected</th>
<th>Higher or lower</th>
<th>$P$ value</th>
<th>$Q_{Hoch}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Loss aversion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,716,061.5</td>
<td>1,653,313.5</td>
<td>Higher</td>
<td>.02</td>
<td>.10</td>
</tr>
<tr>
<td>Yes</td>
<td>4,326,964.5</td>
<td>4,389,712.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Risk aversion</strong></td>
<td>&lt;.001</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,551,369.5</td>
<td>1,653,313.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,491,656.5</td>
<td>4,389,712.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Loss resilience</strong></td>
<td>&lt;.001</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,524,220</td>
<td>1,653,313.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,518,806</td>
<td>4,389,712.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ante</strong></td>
<td>&lt;.001</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,803,995</td>
<td>1,653,313.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,239,031</td>
<td>4,389,712.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Insurance</strong></td>
<td>&lt;.001</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,535,572.5</td>
<td>1,653,313.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,507,453.5</td>
<td>4,389,712.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Peak positive risk</strong></td>
<td>&lt;.001</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,801,272</td>
<td>1,653,313.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,241,754</td>
<td>4,389,712.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Peak negative risk</strong></td>
<td>.001</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,738,603</td>
<td>1,653,313.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,304,423</td>
<td>4,389,712.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reward tipping point</strong></td>
<td>.30</td>
<td>N/A²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,625,945</td>
<td>1,653,313.5</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,417,081</td>
<td>4,389,712.5</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Aversion tipping point</strong></td>
<td>.94</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,651,477</td>
<td>1,653,313.5</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,391,549</td>
<td>4,389,712.5</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total reward risk</strong></td>
<td>&lt;.001</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,771,478</td>
<td>1,653,313.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,271,548</td>
<td>4,389,712.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total aversion risk</strong></td>
<td>.002</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,737,001.5</td>
<td>1,653,313.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,306,024.5</td>
<td>4,389,712.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reward-aversion trade-off</strong></td>
<td>.31</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,626,357</td>
<td>1,653,313.5</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,416,669</td>
<td>4,389,712.5</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trade-off range</strong></td>
<td>&lt;.001</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,529,559.5</td>
<td>1,653,313.5</td>
<td>Lower</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4,513,466.5</td>
<td>4,389,712.5</td>
<td>Higher</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reward-aversion consistency</strong></td>
<td>.20</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Machine Learning Results: Predicting Vaccination Uptake

**Prediction Results**

With the inclusion of demographic and judgment variables, the BRF classifier with the highest accuracy (68.9%) and precision (86.7%) in predicting vaccine uptake resulted when the number of estimators was set to 300 and the number of splits was set to 10. With the addition of 4 COVID-19 precaution behaviors, the BRF classifier with the highest accuracy (70.8%) and precision (87.8%) to predict vaccine uptake occurred when the number of estimators was set to 300 and the number of splits was set to 10. It is notable that specificity was consistently >72%, precision was >86%, and the AUROC was >75% but the NPV was consistently <50%. For random forest and logistic regression, recall and accuracy values were higher than those for BRF, but specificity was always <39%, indicating a lower performance in predicting those who did not receive the vaccine. Precision was also lower, yet the AUROC was consistent with that of the BRF results.
Table 5. Predicting COVID-19 vaccine uptake using three machine learning algorithms⁵.

<table>
<thead>
<tr>
<th>Method and features</th>
<th>Splits</th>
<th>Estimators</th>
<th>Accuracy (%)</th>
<th>Recall (%)</th>
<th>Specificity (%)</th>
<th>Precision (%)</th>
<th>NPV (%)</th>
<th>AUROC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BRF⁶</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judgment+demographic</td>
<td>10</td>
<td>100</td>
<td>68.4</td>
<td>67.1</td>
<td>72.1</td>
<td>86.5</td>
<td>45.2</td>
<td>75.2</td>
</tr>
<tr>
<td>Judgment+demographic</td>
<td>15</td>
<td>100</td>
<td>68.3</td>
<td>66.9</td>
<td>72.2</td>
<td>86.5</td>
<td>45.2</td>
<td>75.1</td>
</tr>
<tr>
<td>Judgment+demographic</td>
<td>10</td>
<td>300</td>
<td>68.9</td>
<td>67.5</td>
<td>72.4</td>
<td>86.7</td>
<td>45.7</td>
<td>75.6</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>15</td>
<td>300</td>
<td>68.8</td>
<td>67.5</td>
<td>72.3</td>
<td>86.7</td>
<td>45.7</td>
<td>75.6</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>10</td>
<td>100</td>
<td>70.6</td>
<td>69.2</td>
<td>74.2</td>
<td>87.7</td>
<td>47.6</td>
<td>78.6</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>15</td>
<td>100</td>
<td>70.4</td>
<td>69.0</td>
<td>74.2</td>
<td>87.7</td>
<td>47.6</td>
<td>78.6</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>10</td>
<td>300</td>
<td>70.8</td>
<td>69.6</td>
<td>74.2</td>
<td>87.8</td>
<td>48.0</td>
<td>79.0</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>15</td>
<td>300</td>
<td>70.8</td>
<td>69.5</td>
<td>74.2</td>
<td>87.8</td>
<td>48.0</td>
<td>79.0</td>
</tr>
<tr>
<td><strong>Random forest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judgment+demographic</td>
<td>10</td>
<td>100</td>
<td>74.9</td>
<td>91.6</td>
<td>30.0</td>
<td>77.7</td>
<td>57.7</td>
<td>74.7</td>
</tr>
<tr>
<td>Judgment+demographic</td>
<td>15</td>
<td>100</td>
<td>74.8</td>
<td>91.7</td>
<td>30.1</td>
<td>77.7</td>
<td>57.9</td>
<td>74.7</td>
</tr>
<tr>
<td>Judgment+demographic</td>
<td>10</td>
<td>300</td>
<td>75.0</td>
<td>92.3</td>
<td>29.1</td>
<td>77.6</td>
<td>58.8</td>
<td>75.1</td>
</tr>
<tr>
<td>Judgment+demographic</td>
<td>15</td>
<td>300</td>
<td>75.0</td>
<td>92.3</td>
<td>29.0</td>
<td>77.6</td>
<td>59.0</td>
<td>75.1</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>10</td>
<td>100</td>
<td>77.0</td>
<td>91.7</td>
<td>38.1</td>
<td>79.8</td>
<td>63.6</td>
<td>78.1</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>15</td>
<td>100</td>
<td>77.0</td>
<td>91.6</td>
<td>38.3</td>
<td>79.8</td>
<td>63.6</td>
<td>78.1</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>10</td>
<td>300</td>
<td>77.3</td>
<td>92.1</td>
<td>37.9</td>
<td>79.7</td>
<td>64.8</td>
<td>78.5</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>15</td>
<td>300</td>
<td>77.3</td>
<td>92.1</td>
<td>38.0</td>
<td>79.8</td>
<td>64.8</td>
<td>78.5</td>
</tr>
<tr>
<td><strong>Logistic regression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judgment+demographic</td>
<td>N/A⁷</td>
<td>N/A</td>
<td>74.8</td>
<td>91.6</td>
<td>30.2</td>
<td>77.7</td>
<td>57.8</td>
<td>75.5</td>
</tr>
<tr>
<td>Judgment+demographic+covid_beh</td>
<td>N/A⁷</td>
<td>N/A</td>
<td>77.0</td>
<td>92.1</td>
<td>37.2</td>
<td>79.6</td>
<td>64.0</td>
<td>79.1</td>
</tr>
</tbody>
</table>

⁵A total of 15 judgment variables (Table 4), 7 demographic variables (Table 3), and 4 COVID-19 precaution behavior (covid_beh) variables (Table S1 in Multimedia Appendix 1) were included in balanced random forest, random forest, and logistic regression models to predict COVID-19 vaccine uptake. We used 10-fold cross-validation, where the data were split 90-10 for each of the 10 iterations.

⁶NPV: negative predictive value.

⁷AUROC: area under the receiver operating characteristic curve.

Feature Importance for BRF and Random Forest

Regarding BRF, Gini importance was highest for age, educational level, and income in both BRF classifiers (both without [Figures 3A and 3B] and with [Figures 3C and 3D] inclusion of the COVID-19 precaution behaviors; refer to the clusters outlined in red in Figures 3B and 3D). For both BRF classifiers, the top 3 predictors (age, income, and educational level) had a combined effect of 23.4% on the Gini importance for prediction. Following these predictors, the 15 judgment variables had similar importance scores for both BRF classifiers (range 0.037-0.049; refer to the clusters outlined in black in Figures 3B and 3D). Regarding BRF, Gini importance was highest for age, educational level, and income in both BRF classifiers (both without [Figures 3A and 3B] and with [Figures 3C and 3D] inclusion of the COVID-19 precaution behaviors; refer to the clusters outlined in red in Figures 3B and 3D). For both BRF classifiers, the top 3 predictors (age, income, and educational level) had a combined effect of 23.4% on the Gini importance for prediction. Following these predictors, the 15 judgment variables had similar importance scores for both BRF classifiers (range 0.037-0.049; refer to the clusters outlined in black in Figures 3B and 3D). These 15 predictors had a combined effect of 62.9% to 68.7% on the Gini importance for prediction, indicating that judgment variables were collectively the most important for prediction outcomes. The least important features for predicting vaccination status were demographic variables regarding employment status, marital status, ethnicity, sex, and the 4 COVID-19 precaution behaviors. These predictors only contributed 7.3% to the Gini importance for prediction. As a follow-up analysis, BRF analyses were run using the top 3 features from both the Gini importance plots (age, educational level, and income; Table S4 in Multimedia Appendix 1) and the top 3 features plus 15 judgment variables (Table S5 in Multimedia Appendix 1). The results did not outperform those presented in Table 5. For random forest, the Gini importance was highest for age and educational level (Figure 4). These top 2 predictors had a combined effect of 16.5% to 16.8% for the 2 models (Figures 4A and 4C). Following these predictors, the 15 judgment variables and the income variable had similar Gini importance, with a combined effect of 69.4% to 75.5% for Gini importance. The least important predictors mirrored those of the BRF results.
Figure 3. Balanced random forest (BRF) Gini importance plots. (A) Table of predictor ranks for the model including 15 judgment and 7 demographic variables. (B) Gini plot of the ranks in (A); the red box outlines the top 3 features (age, income, and educational level), and the black box outlines the 15 judgment variables. (C) Table of predictor ranks for the model including 15 judgment, 7 demographic, and 4 COVID-19 precaution behavior variables. (D) Gini plot of the ranks in (C); the red box outlines the top 3 features (age, income, and educational level), and the black box outlines the 15 judgment variables.

<table>
<thead>
<tr>
<th>Number</th>
<th>Predictor</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>Age</td>
<td>0.12096</td>
</tr>
<tr>
<td>21</td>
<td>Education level</td>
<td>0.05845</td>
</tr>
<tr>
<td>20</td>
<td>Income</td>
<td>0.05491</td>
</tr>
<tr>
<td>19</td>
<td>Consistency range</td>
<td>0.04869</td>
</tr>
<tr>
<td>18</td>
<td>Peak negative risk</td>
<td>0.04853</td>
</tr>
<tr>
<td>17</td>
<td>Loss resilience</td>
<td>0.04825</td>
</tr>
<tr>
<td>16</td>
<td>Reward-aversion consistency</td>
<td>0.04753</td>
</tr>
<tr>
<td>15</td>
<td>Total reward risk</td>
<td>0.04686</td>
</tr>
<tr>
<td>14</td>
<td>Peak positive risk</td>
<td>0.04664</td>
</tr>
<tr>
<td>13</td>
<td>Total aversion risk</td>
<td>0.04651</td>
</tr>
<tr>
<td>12</td>
<td>Risk aversion</td>
<td>0.04613</td>
</tr>
<tr>
<td>11</td>
<td>Reward tipping point</td>
<td>0.04557</td>
</tr>
<tr>
<td>10</td>
<td>Aversion tipping point</td>
<td>0.04532</td>
</tr>
<tr>
<td>9</td>
<td>Loss aversion</td>
<td>0.04484</td>
</tr>
<tr>
<td>8</td>
<td>Anxie</td>
<td>0.04420</td>
</tr>
<tr>
<td>7</td>
<td>Trade-off range</td>
<td>0.04356</td>
</tr>
<tr>
<td>6</td>
<td>Reward-aversion trade-off</td>
<td>0.04259</td>
</tr>
<tr>
<td>5</td>
<td>Insurance</td>
<td>0.04177</td>
</tr>
<tr>
<td>4</td>
<td>Employment status</td>
<td>0.03525</td>
</tr>
<tr>
<td>3</td>
<td>Marital status</td>
<td>0.02355</td>
</tr>
<tr>
<td>2</td>
<td>Ethnicity</td>
<td>0.01084</td>
</tr>
<tr>
<td>1</td>
<td>Sex</td>
<td>0.00908</td>
</tr>
</tbody>
</table>

Figure 4. Random forest (RF) Gini importance plots. (A) Table of predictor ranks for the model including 15 judgment and 7 demographic variables. (B) Gini plot of the ranks in (A); the red box outlines the top 2 features (age and educational level), and the black box outlines the 15 judgment variables. (C) Table of predictor ranks for the model including 15 judgment, 7 demographic, and 4 COVID-19 precaution behavior variables. (D) Gini plot of the ranks in (C); the red box outlines the top 2 features (age and educational level), and the black box outlines the 15 judgment variables.

<table>
<thead>
<tr>
<th>Number</th>
<th>Predictor</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>Age</td>
<td>0.10999</td>
</tr>
<tr>
<td>21</td>
<td>Education level</td>
<td>0.05845</td>
</tr>
<tr>
<td>20</td>
<td>Income</td>
<td>0.05025</td>
</tr>
<tr>
<td>19</td>
<td>Consistency range</td>
<td>0.05015</td>
</tr>
<tr>
<td>18</td>
<td>Reward-aversion consistency</td>
<td>0.04914</td>
</tr>
<tr>
<td>17</td>
<td>Risk aversion</td>
<td>0.04982</td>
</tr>
<tr>
<td>16</td>
<td>Aversion tipping point</td>
<td>0.04819</td>
</tr>
<tr>
<td>15</td>
<td>Peak negative risk</td>
<td>0.04891</td>
</tr>
<tr>
<td>14</td>
<td>Total reward risk</td>
<td>0.04857</td>
</tr>
<tr>
<td>13</td>
<td>Peak positive risk</td>
<td>0.04763</td>
</tr>
<tr>
<td>12</td>
<td>Total aversion risk</td>
<td>0.04689</td>
</tr>
<tr>
<td>11</td>
<td>Reward tipping point</td>
<td>0.04657</td>
</tr>
<tr>
<td>10</td>
<td>Reward-aversion trade-off</td>
<td>0.04647</td>
</tr>
<tr>
<td>9</td>
<td>Anxie</td>
<td>0.04537</td>
</tr>
<tr>
<td>8</td>
<td>Income</td>
<td>0.04534</td>
</tr>
<tr>
<td>7</td>
<td>Loss aversion</td>
<td>0.04532</td>
</tr>
<tr>
<td>6</td>
<td>Trade-off range</td>
<td>0.04288</td>
</tr>
<tr>
<td>5</td>
<td>Insurance</td>
<td>0.04252</td>
</tr>
<tr>
<td>4</td>
<td>Employment status</td>
<td>0.03334</td>
</tr>
<tr>
<td>3</td>
<td>Marital status</td>
<td>0.02383</td>
</tr>
<tr>
<td>2</td>
<td>Ethnicity</td>
<td>0.01320</td>
</tr>
<tr>
<td>1</td>
<td>Sex</td>
<td>0.00839</td>
</tr>
</tbody>
</table>
Logistic Regression Model Statistics

Both model 1 (demographic and judgment variables) and model 2 (demographic, judgment, and COVID-19 precaution behavior variables) were significant ($P<.001$). The model statistics are provided in Tables 6 (model 1) and 7 (model 2). In model 1, age, income, marital status, employment status, sex, educational level, ante, aversion tipping point, reward-aversion consistency, and consistency range were significant ($\alpha<.05$). In model 2, age, income, marital status, employment status, sex, educational level, risk aversion, ante, peak negative risk, mask wearing, and not gathering in large groups were significant ($\alpha<.05$).

Table 6. Logistic regression model 1 results (demographic and judgment variables predict vaccine uptake).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable coefficient</th>
<th>P value</th>
<th>z statistic (SE; 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.039</td>
<td>&lt;.001</td>
<td>11.70 (0.003; 0.033 to 0.046)</td>
</tr>
<tr>
<td>Income</td>
<td>0.262</td>
<td>&lt;.001</td>
<td>7.91 (0.033; 0.197 to 0.326)</td>
</tr>
<tr>
<td>Marital status</td>
<td>−0.061</td>
<td>.03</td>
<td>−2.21 (0.028; −0.116 to −0.007)</td>
</tr>
<tr>
<td>Employment status</td>
<td>0.048</td>
<td>.04</td>
<td>2.03 (0.023; 0.002 to 0.093)</td>
</tr>
<tr>
<td>Sex</td>
<td>−0.244</td>
<td>.008</td>
<td>−2.65 (0.092; −0.425 to −0.064)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.017</td>
<td>.66</td>
<td>0.44 (0.038; −0.057 to 0.091)</td>
</tr>
<tr>
<td>Educational level</td>
<td>0.232</td>
<td>&lt;.001</td>
<td>6.84 (0.034; 0.166 to 0.299)</td>
</tr>
<tr>
<td>Loss aversion</td>
<td>0.002</td>
<td>.78</td>
<td>0.28 (0.009; −0.015 to 0.019)</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>−0.605</td>
<td>.06</td>
<td>−1.89 (0.320; −1.232 to 0.023)</td>
</tr>
<tr>
<td>Loss resilience</td>
<td>0.353</td>
<td>.13</td>
<td>1.51 (0.235; −0.107 to 0.813)</td>
</tr>
<tr>
<td>Ante</td>
<td>−1.079</td>
<td>.04</td>
<td>−2.08 (0.519; −2.096 to −0.061)</td>
</tr>
<tr>
<td>Insurance</td>
<td>0.364</td>
<td>.45</td>
<td>0.75 (0.483; −0.582 to 1.311)</td>
</tr>
<tr>
<td>Peak positive risk</td>
<td>−0.176</td>
<td>.40</td>
<td>−0.85 (0.208; −0.584 to 0.231)</td>
</tr>
<tr>
<td>Peak negative risk</td>
<td>−0.383</td>
<td>.05</td>
<td>−1.93 (0.199; −0.773 to 0.006)</td>
</tr>
<tr>
<td>Reward tipping point</td>
<td>0.126</td>
<td>.26</td>
<td>1.14 (0.110; −0.090 to 0.342)</td>
</tr>
<tr>
<td>Aversion tipping point</td>
<td>−0.499</td>
<td>.03</td>
<td>−2.19 (0.228; −0.947 to −0.052)</td>
</tr>
<tr>
<td>Total reward risk</td>
<td>−0.020</td>
<td>.52</td>
<td>−0.64 (0.032; −0.083 to 0.042)</td>
</tr>
<tr>
<td>Total aversion risk</td>
<td>0.114</td>
<td>.09</td>
<td>1.76 (0.065; −0.013 to 0.240)</td>
</tr>
<tr>
<td>Reward-aversion trade-off</td>
<td>−0.002</td>
<td>.63</td>
<td>−0.49 (0.004; −0.009 to 0.006)</td>
</tr>
<tr>
<td>Trade-off range</td>
<td>0.012</td>
<td>.12</td>
<td>1.54 (0.008; −0.002 to 0.028)</td>
</tr>
<tr>
<td>Reward-aversion consistency</td>
<td>−0.367</td>
<td>.047</td>
<td>−1.98 (0.185; −0.730 to −0.004)</td>
</tr>
<tr>
<td>Consistency range</td>
<td>−0.399</td>
<td>.04</td>
<td>−2.06 (0.194; −0.779 to −0.0019)</td>
</tr>
</tbody>
</table>

Overall model: $P<.001$; pseudo-$R^2$=0.149; log-likelihood=−1736.8; log-likelihood null=−2039.7.
Table 7. Logistic regression model 2 results (demographic, judgment, and COVID-19 precaution variables predict vaccine uptake).  

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable coefficient</th>
<th>P value</th>
<th>z statistic (SE; 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.040</td>
<td>&lt;.001</td>
<td>11.33 (0.003; 0.033 to 0.046)</td>
</tr>
<tr>
<td>Income</td>
<td>0.310</td>
<td>&lt;.001</td>
<td>8.82 (0.035; 0.214 to 0.379)</td>
</tr>
<tr>
<td>Marital status</td>
<td>−0.064</td>
<td>.03</td>
<td>−2.18 (0.029; −0.121 to −0.006)</td>
</tr>
<tr>
<td>Employment status</td>
<td>0.049</td>
<td>.04</td>
<td>2.02 (0.024; 0.002 to 0.097)</td>
</tr>
<tr>
<td>Sex</td>
<td>−0.244</td>
<td>.01</td>
<td>−2.55 (0.096; −0.432 to −0.056)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.012</td>
<td>.75</td>
<td>0.32 (0.039; −0.063 to 0.088)</td>
</tr>
<tr>
<td>Educational level</td>
<td>0.264</td>
<td>&lt;.001</td>
<td>7.42 (0.036; 0.194 to 0.333)</td>
</tr>
<tr>
<td>Loss aversion</td>
<td>0.008</td>
<td>.39</td>
<td>0.87 (0.009; −0.010 to 0.026)</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>−0.663</td>
<td>.045</td>
<td>−2.00 (0.331; −1.311 to −0.014)</td>
</tr>
<tr>
<td>Loss resilience</td>
<td>0.216</td>
<td>.37</td>
<td>0.89 (0.243; −0.259 to 0.691)</td>
</tr>
<tr>
<td>Ante</td>
<td>−1.207</td>
<td>.03</td>
<td>−2.23 (0.542; −2.270 to −1.144)</td>
</tr>
<tr>
<td>Insurance</td>
<td>0.422</td>
<td>.40</td>
<td>0.84 (0.503; 0.564 to 1.408)</td>
</tr>
<tr>
<td>Peak positive risk</td>
<td>0.079</td>
<td>.72</td>
<td>0.36 (0.220; −0.353 to 0.511)</td>
</tr>
<tr>
<td>Peak negative risk</td>
<td>−0.434</td>
<td>.04</td>
<td>−2.09 (0.208; −0.842 to −0.026)</td>
</tr>
<tr>
<td>Reward tipping point</td>
<td>0.166</td>
<td>.22</td>
<td>1.24 (0.134; −0.097 to 0.430)</td>
</tr>
<tr>
<td>Aversion tipping point</td>
<td>−0.412</td>
<td>.08</td>
<td>−1.74 (0.237; −0.876 to 0.052)</td>
</tr>
<tr>
<td>Total reward risk</td>
<td>−0.044</td>
<td>.26</td>
<td>−1.12 (0.039; −0.121 to 0.033)</td>
</tr>
<tr>
<td>Total aversion risk</td>
<td>0.107</td>
<td>.12</td>
<td>1.58 (0.068; −0.026 to 0.240)</td>
</tr>
<tr>
<td>Reward-aversion trade-off</td>
<td>−0.001</td>
<td>.78</td>
<td>−0.28 (0.004; −0.009 to 0.007)</td>
</tr>
<tr>
<td>Trade-off range</td>
<td>0.014</td>
<td>.09</td>
<td>1.68 (0.008; −0.002 to 0.029)</td>
</tr>
<tr>
<td>Reward-aversion consistency</td>
<td>0.254</td>
<td>.20</td>
<td>1.29 (0.197; −0.133 to 0.641)</td>
</tr>
<tr>
<td>Consistency range</td>
<td>0.064</td>
<td>.76</td>
<td>0.31 (0.206; −0.340 to 0.468)</td>
</tr>
<tr>
<td>Mask wearing</td>
<td>−0.875</td>
<td>&lt;.001</td>
<td>−5.31 (0.165; −1.198 to −0.552)</td>
</tr>
<tr>
<td>Social distancing</td>
<td>−0.384</td>
<td>.06</td>
<td>−1.92 (0.201; −0.778 to 0.009)</td>
</tr>
<tr>
<td>Hand hygiene</td>
<td>−0.155</td>
<td>.44</td>
<td>−0.78 (0.198; −0.543 to 0.234)</td>
</tr>
<tr>
<td>Not gathering in large groups</td>
<td>−0.680</td>
<td>&lt;.001</td>
<td>−4.31 (0.158; −0.990 to −0.371)</td>
</tr>
</tbody>
</table>

aOverall model: P<.001; pseudo-R^2=0.206; log-likelihood=−1620.0; log-likelihood null=−2039.7.

**Post Hoc Mediation and Moderation**

Because judgment variables and demographic variables (age, income, and educational level) were important predictors, we evaluated post hoc whether demographics statistically mediated or moderated the relationship between each of the 15 judgment variables and binary responses to COVID-19 vaccination.

For primary mediations, age significantly mediated the statistical relationship between 11 judgment variables and vaccine uptake (α<.05; Table 8), income mediated 8 relationships α< <.05; Table 8), and educational level mediated 9 relationships (α<.05; Table 8). In total, 7 judgment variables overlapped across the 3 models: loss resilience, ante, insurance, peak positive risk, peak negative risk, risk aversion trade-off, and consistency range. Of these, 5 significantly differed between vaccine uptake (those fully vaccinated and those not): loss resilience, ante, insurance, peak positive risk, and peak negative risk (Table 3). Thus, 2 judgment features did not differ by vaccine uptake but were connected with uptake by significant mediation.

For the secondary mediation analyses, 5 judgment variables mediated the statistical relationship between age and vaccine uptake; these variables overlapped with the 11 findings of the primary mediation analyses. Furthermore, 4 judgment variables mediated the statistical relationship between income and vaccine uptake; these variables overlapped with the 8 findings of the primary mediation analyses. Finally, 4 judgment variables mediated the statistical relationship between educational level and vaccine uptake; these variables overlapped with the 9 findings of the primary mediation analyses. In all secondary analyses, approximately half of the judgment variables were involved in mediation as compared to the doubling of judgment variable numbers observed in the primary mediation analyses. In the secondary mediation analyses, the same 4 judgment variables were found in both primary and secondary mediation results, indicating a mixed mediation framework.
Table 8. Mediation and moderation results ($\alpha=.05$). Mediator and moderator variables appear in boldface.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Mediator variable</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mediator</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk aversion</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Loss resilience</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Ante</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>Peak positive risk</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Peak negative risk</td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Aversion tipping point</td>
<td></td>
<td>.004</td>
</tr>
<tr>
<td>Total reward risk</td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Risk aversion trade-off</td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>Trade-off range</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Consistency range</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss resilience</td>
<td></td>
<td>.005</td>
</tr>
<tr>
<td>Ante</td>
<td></td>
<td>.002</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Peak positive risk</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Peak negative risk</td>
<td></td>
<td>.2</td>
</tr>
<tr>
<td>Risk aversion trade-off</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Risk aversion consistency</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Consistency range</td>
<td></td>
<td>.007</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk aversion</td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td>Loss resilience</td>
<td></td>
<td>.009</td>
</tr>
<tr>
<td>Ante</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Insurance</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Peak positive risk</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Peak negative risk</td>
<td></td>
<td>.009</td>
</tr>
<tr>
<td>Risk aversion trade-off</td>
<td></td>
<td>.004</td>
</tr>
<tr>
<td>Risk aversion consistency</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Consistency range</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Loss resilience</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td><strong>Ante</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td>.08</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Insurance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>.02</td>
</tr>
</tbody>
</table>
From the moderation analyses, only 2 interactions out of a potential 45 were observed. Age interacted with risk aversion trade-off, and income interacted with loss resilience to statistically predict vaccine uptake ($\alpha < .05$; Table 8). The 2 moderation results overlapped with the mediation results, indicating mixed mediation-moderation relationships [78,80,81].

**Discussion**

**Principal Findings**

Relatively few studies have sought to predict COVID-19 vaccine uptake using machine learning approaches [8,59]. Given that a small set of studies has assessed the psychological basis that may underlie vaccine uptake and choices [6,52,53,56,58,59,83], but none have used computational cognition variables based on reward and aversion judgment to predict vaccine uptake, we sought to assess whether variables quantifying human judgment predicted vaccine uptake. This study found that 7 demographic and 15 judgment variables predicted vaccine uptake. Other machine learning approaches (random forest and logistic regression) produced higher accuracies but lower specificities, indicating a lower prediction of those who did not receive the vaccine. The BRF also had challenges predicting the negative class, as demonstrated by the relatively low NPV despite having higher specificity than random forest and logistic regression. Feature importance analyses from both BRF and random forest showed that the judgment variables collectively dominated the Gini importance scores. Furthermore, demographic variables acted as statistical mediators in the relationship between judgment variables and vaccine uptake. These mediation findings support the interpretation of the machine learning results that demographic factors, together with judgment variables, predict COVID-19 vaccine uptake.

**Interpretation of Judgment Differences Between Vaccinated and Nonvaccinated Individuals**

Those who were fully vaccinated had lower values for loss aversion, ante, peak positive risk, peak negative risk, total reward risk, and total aversion risk, along with higher values for risk aversion, loss resilience, insurance, and trade-off range (refer to Table 1 for variable descriptions). Lower loss aversion corresponds to less overweighting of bad outcomes relative to good ones [84] and a potential willingness to obtain a vaccine with uncertain outcomes. A lower ante suggests that individuals are less willing to engage in risky behaviors surrounding potential infection, which is also consistent with the 4 other judgment variables that define relationships between risk and value (peak positive risk, peak negative risk, total reward risk, and total aversion risk). In participants who indicated full vaccination, lower peak positive risk and peak negative risk were related to individuals having a lower risk that they must overcome to make a choice to either approach or avoid, as per the decision utility equation by Markowitz [39,71]. The lower total reward risk and total aversion risk indicate that the interactions between reward, aversion, and the risks associated with them did not scale significantly; namely, higher reward was not associated with higher risk, and higher negative outcomes were not associated with the uncertainty of them. For these participants, the ability of the vaccine to increase the probability of health and reduce the probability of harm from illness did not have to overcome high obstacles in their vaccine choice. Higher risk aversion in vaccinated participants suggests that these participants viewed contracting COVID-19 as a larger risk and, therefore, were more likely to receive the full dose. These findings are consistent with those of a study by Lepinteur et al [58], who found that risk-averse individuals were more likely to accept the COVID-19 vaccination, indicating that the perceived risk of contracting COVID-19 was greater than any risk from the vaccine. Hudson and Montelpare [54] also found that risk aversion may promote vaccine adherence when people perceive contracting a disease as more dangerous or likely. Higher loss resilience in the vaccinated group was also consistent with the perspective that vaccination would improve...
their resilience and act as a form of insurance against negative consequences. The higher trade-off range suggests that vaccinated individuals have a broader portfolio of preferences and are more adaptive to bad things occurring, whereas a lower trade-off indicates a restriction in preferences and less adaptability in those who did not receive the vaccine.

**Comparison of Prediction Algorithms**

When testing these judgment variables (with demographic and COVID-19 precaution behavior variables) in a BRF framework to predict vaccine uptake, we observed a high AUROC of 0.79, where an AUROC of 0.8 is often the threshold for excellent model performance in machine learning [85,86]. The similarity of our reported recall and specificity values with the BRF suggests a balance between predicting true positives and true negatives. The high precision indicates a high certainty in predicting those who were fully vaccinated. The BRF model was successful in identifying those who received the full vaccine (positive cases; indicated by high precision and moderate recall) and those who did not (negative cases; indicated by the specificity). However, NPV was low, indicating a higher rate of false prediction of those who did not receive a full dose counterbalanced by a higher specificity that reflects a higher rate of predicting true negatives. These observations are reflected in the moderate accuracy, which measures the number of correct predictions. A comparison of random forest, logistic regression, and BRF revealed that random forest and logistic regression models produced less balance between recall (high) and specificity (low), which could be interpreted as a bias toward predicting the majority class (ie, those who received the vaccine). That being said, the NPV for BRF was lower than that for random forest and logistic regression, where a low NPV indicates a low probability that those predicted to have not received the vaccine truly did not receive the vaccine when taking both classes into account. Together, the results from all 3 machine learning approaches reveal challenges in predicting the negative class (ie, those who did not receive the vaccine).

Overall, the 3 models achieved high accuracy, recall, precision, and AUROC. BRF produced a greater balance between recall and specificity, and the outcome of the worst-performing metric (ie, NPV) was still higher than the specificities for the random forest and logistic regression models.

**Feature Importance**

Of the 3 prediction algorithms, random forest and BRF had very similar Gini importance results, whereas logistic regression elevated most demographic variables and a minority of judgment variables. This observation could be due to the large variance in each of the judgment variables, which could present challenges for achieving a good fit with logistic regression. In contrast, the demographic and COVID-19 precaution variables had low variance and could be more easily fit in a linear model, hence their significance in the logistic regression results. In comparison to logistic regression, decision trees (eg, BRF and random forest) use variable variance as additional information to optimize classification, potentially leading to a higher importance of judgment variables over most demographic and all COVID-19 precaution variables.

Focusing on the model with balanced recall and specificity (ie, the BRF classifiers [with and without COVID-19 precaution behaviors]), the top predictors were 3 demographic variables (age, income, and educational level), with distributions that varied by vaccine uptake in manners consistent with those of other reports. Namely, older individuals, those identifying as male and White individuals, and those who indicated a higher income and educational level corresponded to those who were or intended to be vaccinated [2,5,87]. Despite their saliency, these 3 variables together only contributed 23% to the prediction, corresponding to approximately one-third of the contribution from the 15 judgment variables (63%-69%). The individual Gini importance scores for the 15 judgment variables only ranged from 0.039 to 0.049 but were the dominant set of features behind the moderate accuracy, high precision, and high AUROC. The 18% difference between the accuracy and precision measures suggests that variables other than those used in this study may improve prediction, including contextual variables that may influence vaccine choices. Variables may include political affiliation [7], longitude and latitude [8], access to the internet [8], health literacy [54], and presence of underlying conditions [9]. Future work should seek to include these types of variables.

In the second BRF classifier, the 4 COVID-19 precaution behaviors only contributed 6.6% to the prediction. This low contribution could be due to these variables being binary, unlike the other demographic variables, which included a range of categories. In addition, COVID-19 precaution behaviors are specific to the context of the COVID-19 pandemic and do not promote interpretation beyond their specific context. The 15 judgment variables represent a contrast to this as they are empirically computed from a set of functions across many picture categories. An individual with higher risk aversion will generally tolerate higher amounts of uncertainty regarding a potential upside or gain as opposed to settling for what they have. This does not depend on what stimulus category they observe or the stimulus-response condition. Instead, it is a general feature of the bounds to their judgment and is part of what behavioral economists such as Kahneman consider as bounds to human rationality [84].

**Mechanistic Relationships Between Judgment and Demographic Variables**

The Gini score plots were clear sigmoid-like graphs (Figure 3), with only 3 of the 7 demographic variables ranking above the judgment variables. This observation was consistent in both BRF classifiers (with and without COVID-19 precaution behaviors), raising the possibility of a statistically mechanistic relationship among the top 3 demographic variables, the 15 judgment variables, and vaccine uptake. Indeed, we observed 28 primary mediation effects and 13 secondary mediation effects in contrast to 2 moderation relationships, which also happened to overlap with mediation findings, suggesting mixed mediation-moderation relationships [81,88]. The observation that most judgment variables were significant in mediation relationships but not in moderation relationships argues that prediction depended on the directional relationship between judgment and demographic variables to predict vaccine uptake. Furthermore, there were more significant primary mediations

https://publichealth.jmir.org/2024/1/e47979
(when judgment variables were the IVs) compared to secondary mediations, suggesting the importance of judgment variables as IVs and demographic variables as mediators. Mathematically, judgment variables (IVs) influenced vaccine uptake (DV), and this relationship was stronger when demographic variables were added to the equation. The 13 secondary mediations all overlapped with the 28 primary mediations, where demographic variables were IVs and judgment variables were mediators, suggesting that demographic variables influenced vaccine uptake (DV) and that this relationship became stronger with the addition of judgment variables. This overlap of primary and secondary mediations for 4 of the judgment variables suggests that both judgment and demographic variables influenced the choice of being vaccinated within a mixed mediation framework because adding either one of them to the mediation model regressions made the relationships stronger [49]. The lack of moderation results and a considerable number of overlapping primary and secondary mediation results imply that the relationship between judgment variables and vaccine uptake did not depend purely on their interaction with age, income, or educational level (ie, moderation) but, instead, depended on the direct effects of these 3 demographic variables to strengthen the relationship between judgment variables and vaccine uptake. This type of analysis of statistical mechanisms is helpful for understanding contextual effects on our biases and might be important for considering how best to target or message those with higher loss aversion, ante, peak positive risk, peak negative risk, total reward risk, and total aversion risk (ie, in those who were not fully vaccinated).

**Model Utility**

The developed model is automatable and may have applications in public health. The picture-rating task can be deployed on any smart device or computer, making it accessible to much of the US population or regional populations. The ratings from this task can be automatically processed, and the results can be stored in local or national databases. This method of data collection is novel in that persons cannot bias their responses as the rating task can be automatically processed, and the results can be stored in a general population sample to verify these findings. Government and public health bodies can access these data to determine predicted vaccine uptake rates locally or nationally, which can be used to (1) prepare vaccine rollouts and supply chain demand, (2) prepare health care institutions in areas that may experience low vaccine adherence and potentially higher infection rates, and (3) determine which areas may need more targeted messaging to appeal to specific judgment profiles. For use case 3, messaging about infection risks or precaution behaviors could be framed to address those with lower risk aversion, who, in this study, tended to forgo vaccination. Given that such individualized data would not be available a priori, it would be more plausible to collect data from similarly sized cohorts in geographic regions of concern to obtain regional judgment behavior profiles and, thus, target use cases 1 to 3. Further development of this model with different population samples might also improve our understanding of how certain judgment variables may be targeted with different types of messaging, offering a means to potentially improve vaccine uptake. This model might also be applied to other mandated or recommended vaccines such as those for influenza or human papillomavirus, ultimately improving preparation and messaging efforts. However, future work would be needed to model these varying vaccine choices.

Given the use of demographic variables in the proposed model, specific demographic populations could be assessed or considered for messaging. If particular demographic groups are predicted to have a low vaccine uptake rate, messaging can be targeted to those specific groups. For example, we observed that a higher percentage of female individuals were not fully vaccinated when compared to male individuals. This could be related to concerns about the COVID-19 vaccine affecting fertility or pregnancy. To improve uptake in this population, scientifically backed messaging could be used to confirm the safety of the vaccine in this context. Lower rates of vaccination have been reported in Black communities, which was also observed in this study. Researchers have identified targetable issues related to this observation, which include engagement of Black faith leaders and accessibility of vaccination clinics in Black communities, to name a few [89].

In summary, this model could be used to predict vaccine uptake at the local and national levels and further assess the demographic and judgment features that may underlie these choices.

**Limitations**

This study has a number of limitations that should be considered. First, there are the inherent limitations of using an internet survey—namely, the uncontrolled environment in which participants provide responses. Gold Research, Inc, and the research team applied stringent exclusion criteria, including the evaluation of the judgment graphs given that random responses produce graphs with extremely low $R^2$ fits (eg, <0.1). This was not the case in our cohort of 3476 participants, but this cannot perfectly exclude random or erroneous responses to other questionnaire components. Second, participants with mental health conditions were oversampled to meet the criteria for other survey components not discussed in this paper. This oversampling could potentially bias the results, and future work should use a general population sample to verify these findings. Third, demographic variability and the resulting confounds are inherent in population surveys, and other demographic factors not collected in this study may be important for prediction (eg, religion and family size). Future work might consider collecting a broader array of demographic factors to investigate and include in predictive modeling. Fourth, we used a limited set of 7 demographic variables and 15 judgment variables; however, a larger set of judgment variables is potentially computable and could be considered for future studies. There is also little information on how post–COVID-19 effects, including socioeconomic effects, affect COVID-19 vaccination choices.

**Conclusions**

To our knowledge, there has been minimal research on how biases in human judgment might contribute to the psychology underlying individual vaccination preferences and what differentiates individuals who were fully vaccinated against COVID-19 from those who were not. This population study of several thousand participants demonstrated that a small set of
demographic variables and 15 judgment variables predicted vaccine uptake with moderate to high accuracy and high precision and AUROC, although a large range of specificities was achieved depending on the classification method used. In an age of big data machine learning approaches, this study provides an option for using fewer but more interpretable variables. Age, income, and educational level were independently the most important predictors of vaccine uptake, but judgment variables collectively dominated the importance rankings and contributed almost two-thirds to the prediction of COVID-19 vaccination for the BRF and random forest models. Age, income, and educational level significantly mediated the statistical relationship between judgment variables and vaccine uptake, indicating a statistically mechanistic relationship grounding the prediction results. These findings support the hypothesis that small sets of judgment variables might provide a target for vaccine education and messaging to improve uptake. Such education and messaging might also need to consider contextual variables (ie, age, income, and educational level) that mediate the effect of judgment variables on vaccine uptake. Judgment and demographic variables can be readily collected using any digital device, including smartphones, which are accessible worldwide. Further development and use of this model could (1) improve vaccine uptake, (2) better prepare vaccine rollouts and health care institutions, (3) improve messaging efforts, and (4) have applications for other mandated or government-recommended vaccines.

Acknowledgments
The authors thank Carol Ross, Angela Braggs-Brown, Tom Talavage, Eric Nauman, and Marc Cahay at the University of Cincinnati (UC) College of Engineering and Applied Sciences, who significantly impacted the transfer of research funding to UC. Funding for this work was provided in part to HCB by the Office of Naval Research (awards N00014-21-1-2216 and N00014-23-1-2396) and to HCB from a Jim Goetz donation to the UC College of Engineering and Applied Sciences. Finally, the authors thank the anonymous reviewers for their constructive input, which substantially improved the manuscript. The opinions expressed in this paper are those of the authors and are not necessarily representative of those of their respective institutions.

Data Availability
The data set and corresponding key used in this study are available in Multimedia Appendix 2.

Conflicts of Interest
A provisional patent has been submitted by the following authors (NLV, SB, HCB, SL, LS, and AKK): “Methods of predicting vaccine uptake,” provisional application # 63/449,460.

Multimedia Appendix 1
Supplementary material.
[DOCX File, 10932 KB - publichealth_v101e47979_app1.docx]

Multimedia Appendix 2
Data set.
[XLSX File (Microsoft Excel File), 610 KB - publichealth_v101e47979_app2.xlsx]

References


34. Thompson JC, Chaplin TM. Adolescent brain response to reward is associated with a
delay discounting as potential risk factors for cigarette smoking and other substance use. Prev Med 2022 Dec;165(Pt B):107270 [Medline: 36152818]


delay discounting as potential risk factors for cigarette smoking and other substance use. Prev Med 2022 Dec;165(Pt B):107270 [Medline: 36152818]


47. Thrailkill EA, DeSarno M, Higgins ST. Intersections between environmental reward availability, loss aversion, and delay


49. Hansen A, Turpyn CC, Mauro K, Thompson JC, Chaplin TM. Adolescent brain response to reward is associated with a
bias toward immediate reward. Dev Neuropsychol 2020 May;32(2):397-409 [FREE Full text] [doi: 10.1017/brainawac456] [Medline: 30837014]


53. Trueblood JS, Sussman AB, O’Leary D. The role of preference in responses to messaging about COVID-19 vaccine


72. StataCorp. Stata statistical software: release 17. StataCorp LLC. College Station, TX: StataCorp LLC; 2021. URL: https://www.stata.com/unix/stata/ [accessed 2022-04-23]


Abbreviations

- AUROC: area under the receiver operating characteristic curve
- BRF: balanced random forest
- DV: dependent variable
- IV: independent variable
- NPV: negative predictive value
- RPT: relative preference theory

©Nicole L Vike, Sumra Bari, Leandros Stefanopoulos, Shamal Lalvani, Byoung Woo Kim, Nicos Maglaveras, Martin Block, Hans C Breiter, Aggelos K Katsaggelos. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 18.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Motivators and Demotivators for COVID-19 Vaccination Based on Co-Occurrence Networks of Verbal Reasons for Vaccination Acceptance and Resistance: Repetitive Cross-Sectional Surveys and Network Analysis

Qiuyan Liao1*, PhD; Jiehu Yuan1*, MA; Irene Oi Ling Wong1,2, PhD; Michael Yuxuan Ni1,3,4, MD, PhD; Benjamin John Cowling1,2,5, PhD; Wendy Wing Tak Lam1,6, PhD

1School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
2World Health Organization Collaborating Center for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)
3State Key Laboratory of Brain and Cognitive Sciences, University of Hong Kong, Hong Kong, China (Hong Kong)
4Urban Systems Institute, University of Hong Kong, Hong Kong, China (Hong Kong)
5Laboratory of Data Discovery for Health, Hong Kong Science and Technology Park, Hong Kong, China (Hong Kong)
6Jockey Club Institute of Cancer Care, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China (Hong Kong)
* these authors contributed equally

Corresponding Author:
Qiuyan Liao, PhD
School of Public Health
Li Ka Shing Faculty of Medicine
The University of Hong Kong
2/F, Patrick Manson Building (North Wing), 7 Sassoon Road, Pokfulam
Hong Kong
China (Hong Kong)
Phone: 852 39179289
Fax: 852 28559528
Email: qyliao11@hku.hk

Abstract

Background: Vaccine hesitancy is complex and multifaced. People may accept or reject a vaccine due to multiple and interconnected reasons, with some reasons being more salient in influencing vaccine acceptance or resistance and hence the most important intervention targets for addressing vaccine hesitancy.

Objective: This study was aimed at assessing the connections and relative importance of motivators and demotivators for COVID-19 vaccination in Hong Kong based on co-occurrence networks of verbal reasons for vaccination acceptance and resistance from repetitive cross-sectional surveys.

Methods: We conducted a series of random digit dialing telephone surveys to examine COVID-19 vaccine hesitancy among general Hong Kong adults between March 2021 and July 2022. A total of 5559 and 982 participants provided verbal reasons for accepting and resisting (rejecting or hesitating) a COVID-19 vaccine, respectively. The verbal reasons were initially coded to generate categories of motivators and demotivators for COVID-19 vaccination using a bottom-up approach. Then, all the generated codes were mapped onto the 5C model of vaccine hesitancy. On the basis of the identified reasons, we conducted a co-occurrence network analysis to understand how motivating or demotivating reasons were mentioned to shape people’s vaccination decisions. Each reason’s eigenvector centrality was calculated to quantify their relative importance in the network. Analyses were also stratified by age group.

Results: The co-occurrence network analysis found that the perception of personal risk to the disease (egicentrality=0.80) and the social responsibility to protect others (egicentrality=0.58) were the most important mentioned reasons that motivate COVID-19 vaccination, while lack of vaccine confidence (egicentrality=0.89) and complacency (perceived low disease risk and low importance of vaccination; egicentrality=0.45) were the most important mentioned reasons that demotivate COVID-19 vaccination. For older people aged ≥65 years, protecting others was a more important motivator (egicentrality=0.57), while the
concern about poor health status was a more important demotivator (egicentrality=0.42); for young people aged 18 to 24 years, recovering life normalcy (egicentrality=0.20) and vaccine mandates (egicentrality=0.26) were the more important motivators, while complacency (egicentrality=0.77) was a more important demotivator for COVID-19 vaccination uptake.

Conclusions: When disease risk is perceived to be high, promoting social responsibility to protect others is more important for boosting vaccination acceptance. However, when disease risk is perceived to be low and complacency exists, fostering confidence in vaccines to address vaccine hesitancy becomes more important. Interventions for promoting vaccination acceptance and reducing vaccine hesitancy should be tailored by age.

(JMIR Public Health Surveill 2024;10:e50958) doi:10.2196/50958

KEYWORDS
COVID-19; vaccination acceptance; vaccine hesitancy; motivators; co-occurrence network analysis

Introduction

Background

Despite strong evidence on the effectiveness of COVID-19 vaccines for preventing severe illness, hospitalization, and deaths associated with COVID-19 [1-3], vaccine hesitancy remains widespread and an important barrier for achieving high population uptake of the vaccines [4]. Hong Kong has learned its painful lesson of low vaccination uptake from the explosive epidemic caused by the Omicron BA.2 variant in February to April 2022, which claimed >10,000 lives of its 7 million people within 3 months, the highest per capita COVID-19 mortality worldwide in 2022 [5]. The devastating impact of this outbreak was mainly attributed to the low uptake of the COVID-19 vaccination among people aged ≥60 years before the pandemic. Administration of the mRNA vaccine BNT162b2 (BioNTech, Fosun Pharma, or Pfizer) and the inactivated vaccine CoronaVac (Sinovac) began in February 2021. To boost vaccine uptake, from May to June 2021, the Hong Kong government invited the business sector to offer various vaccination incentives such as lottery prizes (“lucky draw”) and vaccination leaves [6]. The Hong Kong government also introduced the concept of “vaccine bubble” in April 2021 [7], which was regarded as a preparatory stage for the implementation of the “vaccine pass.” Under the “vaccine bubble” strategy, citizens and staff of premises were encouraged to take the COVID-19 vaccination to enjoy some relaxation of social distancing measures [7], for instance, by allowing more customers per table in the restaurants if all staff and customers were vaccinated. Despite various efforts, by December 23, 2021, shortly before Hong Kong was introduced into the community, only 49% of persons aged ≥60 years had received at least 2 doses of COVID-19 vaccines in Hong Kong, and only 7% had received a third dose [8], an important reason for a much higher mortality rate, particularly among older people during the Omicron outbreak. In February 2022, the “vaccine pass” was implemented, which required all eligible persons to receive at least 1 dose of the COVID-19 vaccine to gain access to specified premises [9].

The reasons why people accepted or refused the COVID-19 vaccines are multifaceted, including cultural value, social norms, perceived disease risk, confidence in vaccines, convenience, and trust in government and the health system [10,11]. In view of the complexity of vaccine hesitancy and its determinants, the Strategic Advisory Group of Experts Working Group on Vaccine Hesitancy developed the 3C model, which proposes that vaccine hesitancy is shaped by 3 categories of determinants: *complacency, convenience, and confidence* [12]. The 3C model was later extended to include 2 more Cs: *risk calculation and collective responsibility* [13]. This model provides a good summary of the main categories of determinants of people’s vaccination acceptance and has been widely used to understand people’s vaccination acceptance or hesitancy [14]. On the basis of the original 3C and the extended 5C models, the determinants of different categories are interconnected to shape vaccine hesitancy [12,13]. However, existing studies mainly focused on the association of vaccination acceptance with each determinant of the 5C categories [15,16]. How these determinants are interconnected in shaping the vaccination outcome (acceptance or resistance) remains underexplored. Moreover, most studies that used the 3C or 5C model measured determinants of COVID-19 vaccine hesitancy using questionnaire surveys [17-19]; this may fail to capture the diverse and evolving context-specific factors during the COVID-19 vaccination program such as vaccination incentives [20,21] and vaccine mandates [22,23]. In summary, the multiple facets of vaccine hesitancy mean that people may accept or refuse the vaccine due to multiple psychosocial reasons and the emerging vaccination policies, which has been confirmed in qualitative research [24]. However, most current research that investigated the reasons for vaccination acceptance or hesitancy mainly described the frequency of each reason [25,26] and focused on only 1 vaccination outcome (either acceptance or hesitancy) [24,27]. There is little quantitative evidence on the co-occurrence of reasons for shaping different vaccination outcomes.

Network analysis is a promising statistical technique for depicting the multiple and complex interconnections between variables within a system and has been widely applied for characterizing the networks of psychological symptoms and attitudes [28-31]. A network, which is characterized by its nodes (eg, variables) and edges (the associations between variables), offers powerful visualization for the interactions among multiple factors in shaping a phenomenon. Network analysis is more exploratory and flexible than multivariate regression models. It adopts a complex system perspective and hence assumes that any variables (eg, reasons for vaccination acceptance or resistance) within the system can potentially interconnect with each other. Network analysis for the co-occurrence of reasons for adopting a behavior can quantify the relative importance of reasons based on not only how frequently a reason works solely...
but also how frequently it is mentioned with other reasons for motivating a behavior [32]. This will help to identify the more important reasons and the most effective combination of motivating reasons or demotivating concerns for intervention development and risk communication design. Therefore, the network approach is promising for studying the multifaceted reasons for vaccine hesitancy and their complex interconnections to guide intervention development.

The application of network analysis in vaccine-related attitudes has recently been expanded [33-38]. However, existing studies mainly included factors that were of specific research interests from the researchers’ perspective for network analysis [33-38]. This may miss some specific concerns from the vaccine receivers’ perspective and cause misalignment between expert risk communication and lay people’s preferences or concerns, potentially rendering the intervention ineffective. In addition, existing studies mainly focused on the networks of factors that influenced willingness to accept COVID-19 vaccines [33-38]. However, determinants and their interactions that shape vaccination acceptance and resistance can be different [39-41]. For instance, while protecting oneself and others was a more salient reason for vaccination acceptance, lacking confidence in the vaccine was a more important reason for vaccination rejection [41]. Such differences can be further complicated by the variability of motivators and demotivators for vaccination acceptance across age groups. For example, younger adults were found to refuse a COVID-19 vaccine mainly due to conspiracy beliefs and lacking trust in the government [42], while the top reasons for older adults refusing a COVID-19 vaccine were their chronic condition and old age [27,40]. However, such investigations overlooked the interconnections between motivators and demotivators. Distinguishing the networks of reasons for accepting and refusing a vaccine in different age groups is important for informing the development of vaccine risk communication that is tailored to the audience’s underlying vaccine-related attitudes. Furthermore, these studies, which used the network analysis, neither investigated the reasons relating to the 5C model and the co-occurrence of reasons nor quantified their relative importance for influencing vaccine hesitancy.

Objective
In this study, we collected participants’ diverse verbal reasons for accepting or refusing a COVID-19 vaccine using open-ended questions through population-based telephone surveys conducted in Hong Kong during the COVID-19 pandemic and categorized the reasons based on the 5C model of vaccine hesitancy for network analysis. We aimed to provide quantitative evidence on the co-occurrence of reasons and their relative importance for both vaccine acceptance and resistance, as well as how the relative importance of motivators and demotivators for COVID-19 vaccination could differ by age.

Methods
Participants
Repeated cross-sectional surveys were conducted on a weekly or monthly basis to monitor acceptance of the COVID-19 vaccine among general adults after the COVID-19 vaccination campaign was launched in early 2021 in Hong Kong [43-45]. In each round, we recruited Hong Kong adults aged ≥18 years using random-digital-dialed telephone interviews with a ratio of 1:1 for landlines and mobile phones. The surveys were implemented by a local survey company that had been demonstrated to have the capacity to provide high-quality population-based surveys using random digital calls. All the calls were made during working hours and nonworking hours to avoid oversampling of nonworking people. Each sampled telephone number was called up to 5 times at different times and on different days before being replaced with a new one. For each landline-based call, if there were >1 eligible member in the household, the “next birthday rule” (the person whose next birthday is closest to the survey date) was adopted to determine the person to be interviewed. Individuals with linguistic and cognitive difficulties in completing a telephone interview were excluded. The target sample size was an alternative of 500 or 1000 for each round. In each round, core study measures such as the uptake of COVID-19 vaccines and the intention to receive a COVID-19 vaccine were retained throughout, while additional study measures were rotated to maintain a feasible length of the questionnaire for a telephone interview. Data from a total of 19 survey rounds that collected reasons for accepting a COVID-19 vaccine and 9 survey rounds that collected reasons for being hesitant or resistant about taking a COVID-19 vaccine were used for this analysis. Participants’ demographic distributions were compared between the accepting group and the resistant group using the Pearson chi-square test. All the surveys included were conducted between March 2021 and July 2022. The details of each survey round can be found in Multimedia Appendix 1.

Ethical Considerations
The study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practices. As our surveys were conducted over the telephone, written informed consent was not feasible, but verbal informed consent was obtained from all participants before the interview started. The data sets did not contain any personally identifiable information. This study received ethics approval from the institutional review board (reference UW 20-095).

Study Instruments
To explore the reasons for accepting, hesitating, or refusing a COVID-19 vaccine, in 16 survey rounds conducted between March 1, 2021, and August 20, 2021, participants who had completed at least 1 dose of the COVID-19 vaccine were asked about the major reasons that motivated them to receive the vaccine. The Hong Kong government started recommending a vaccine booster (the third vaccine dose) for general adults who had received 2 doses of COVID-19 vaccines for at least 6 months on November 23, 2021 [46]. Therefore, in 3 additional survey rounds that were conducted after this date (between December 6, 2021, and January 13, 2022), participants who had received a vaccine booster were asked about the reasons that motivated them to receive a vaccine booster. The open-ended responses from these 19 survey rounds, in which participants were asked about their reasons for taking a COVID-19 vaccine or a vaccine booster, were coded as reasons for vaccination acceptance. In 9 survey rounds conducted between December...
6, 2021, and July 14, 2022, we identified resistant participants based on their vaccination status and intention. First, participants were asked about the number of doses of COVID-19 vaccines they had received. Those who had received none, 1 dose, or 2 doses were then asked about how likely they would receive the first, second, or third vaccine dose, respectively, in the next 3 months (rated on a scale from “definitely not” to “very unlikely,” “unlikely,” “unsure,” “likely,” “very likely,” and “definitely yes”). Those who answered “definitely not,” “very unlikely,” “unlikely,” and “unsure” to the vaccination intention question were defined as the resistant group. These participants were asked about the major reasons for being hesitant about or rejecting COVID-19 vaccination. The verbal responses from these 9 survey rounds in which participants were asked about the reasons for refusing the 2-dose primary COVID-19 vaccination series or a vaccine booster were coded as reasons for vaccination resistance. The vaccination resistance reasons were collected at a later phase of the vaccination program and during the Omicron wave when COVID-19–confirmed cases soared rapidly (Multimedia Appendix 2). Therefore, our data capture reasons from the most resistant group, given that the high vaccination uptake rate had been achieved in the population and the disease risk was relatively high, but participants still reported being hesitant or resistant regarding vaccination. The data collection period and number of daily COVID-19 cases are presented in Multimedia Appendix 2. The participants were asked to provide reasons that first came to mind, and then the interviewer jotted down notes of participants’ statements and asked follow-up questions of “any other reasons” to encourage participants to give >1 reason for their vaccination decision. Across all survey rounds, participants were also asked about their sociodemographics including age, sex, educational attainment, employment status, and history of chronic conditions.

Coded Reasons for Accepting or Resisting a COVID-19 Vaccine

The participants’ verbal responses to the open-ended questions regarding reasons for vaccination acceptance and resistance were extracted and independently coded by 2 researchers (QL and JY). We combined the bottom-up and top-down approaches to generate the main categories of these qualitative verbal responses. First, basic categories were independently generated inductively by the 2 researchers by making sense of the meanings of the participants’ verbal responses (bottom-up approach). Following the initial coding, the 5C categories (complacency, confidence, convenience, calculation, and collective responsibility) of vaccine hesitancy [13] were used to guide the clustering of the initial codes to generate the main categories (top-down approach). Although the 5C model was used as a guide for generating categories of motivators and demotivators, the 2 researchers remained open to allow new categories that were not covered by the 5C model to emerge from the data. Following completion of the initial coding, each coder manually checked 10% of the codes generated by another to ensure consistency and consensus on all codes. Any inconsistencies in coding were solved by reviewing the raw data and having joint discussions between the 2 coders (QL and JY). A total of 52 responses that were coded as the reasons for vaccination acceptance and 23 responses that were coded as the reasons for vaccination resistance were excluded from the final analyses because their meanings were judged ambiguous by both coders. To facilitate interpretation, we finally mapped the coded reasons onto the 5C categories but allowed new categories to emerge. This was aimed at facilitating data interpretation and minimizing conceptual or meaning overlap between the coded reasons. For the mapping, each of the 2 coders (QL and JY) first reached a consensus on the definition of each of the 5C categories and then mapped each reason to 1 C. Any inconsistencies in mapping were solved by joint discussions and reviewing the original definitions and relevant literature. Finally, for reasons that cannot be mapped to any of the existing 5C categories, extended categories of the 5C categories would be proposed.

Network Analysis

Network analysis was conducted separately for vaccination acceptance and vaccination resistance in Python using the Networkx Package (Python Software Foundation) [47]. We first determined the nodes included for network estimation and visualization. For either the acceptance or the resistance network, each reason category of vaccination acceptance or resistance, respectively, represents 1 node in the respective network. An edge between 2 nodes indicates that the 2 reasons were comentioned by participants. However, instead of directly modeling the frequency of comentioning, the edges were normalized using a weighted method following previous work [32]. Specifically, if participants mention only 1 reason for their vaccination acceptance or resistance, a standard weight of “1” will be assigned to that reason. If participants mention 2 reasons, each of the 2 reasons will get a standard weight of “1” because there is only 1 unique pair of reasons. If participants mention ≥3 reasons, the weight between each pair of reasons will be “1” divided by the number of pair combinations among the reasons. For instance, if 3 reasons (k1, k2, and k3) are comentioned by participants, the edge value for each pair of the 3 reasons (k1 – k2, k1 – k3, or k2 – k3) would be one-third so that their total weight will still be equal to “1.” This method ensured that each participant contributed equally to the whole network, regardless of how many reasons they had mentioned for vaccination acceptance or resistance. Following this method, a symmetric adjacency matrix was constructed for both reasons of vaccination acceptance and resistance for network analysis. Therefore, if we have k reasons, then an edge can be expressed as E = {E[i][j] = 0,1,...,k}, which is the ith row and jth column of the matrix. The weight of the edge will depend on the number of participants accepting or refusing a COVID-19 vaccine due to reasons i and j. The more the number of participants that comention the 2 reasons, the greater the weight of their edge will be. The adjacency matrix was then used to calculate each node’s eigenvector centrality—an index used for quantifying the relative importance of the node based on nodes’ co-occurrence patterns in the network [48]. Specifically, a higher normalized eigenvector centrality value indicates a relatively greater importance of the reason, which accounts for not only how frequently the node occurred (ie, reason was mentioned by participants) but also how frequently the node was linked to other influential nodes (ie, reason was comentioned with other
frequently mentioned reasons). Mathematically, the eigenvector centrality can be obtained from the following equation: \( AX = \lambda X \).

In this equation, \( A \) represents the adjacency matrix. The eigenvector \( X \) was calculated using 500 power iteration process, given the constant eigenvalue \( \lambda \).

Following the same methodological procedure, we constructed the co-occurrence networks stratified by age subgroups and obtained eigencentrality values for each reason across age groups. A normalized eigencentrality measure can be used for direct comparisons of the relative importance of the same reasons for vaccination acceptance and resistance across different age groups.

Results

Participants

The survey cooperation rate was around 66.7%, ranging from 51.9% to 74.4%. Survey response rate was defined as the proportions of participants who completed the interviews against those who were contacted and eligible (Multimedia Appendix 1). Overall, in 19 survey rounds, 5559 participants provided reasons for accepting a COVID-19 vaccine, and in 9 survey rounds, 982 participants provided reasons for being resistant (including hesitant) about taking a COVID-19 vaccine; these reasons were used for reason co-occurrence network analysis. Comparisons of the demographics of participants who accepted the vaccine and those who were resistant to receiving the vaccine showed that the vaccination resistance group was more likely to be younger, have higher educational attainment, be unemployed, and have at least 1 chronic condition (Table 1).

Table 1. Participants’ demographic characteristics in the COVID-19 vaccination accepting group and resistant group between March 1, 2021, and July 14, 2022.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Accepting group (n=5559)</th>
<th>Resistant group (n=982)</th>
<th>Differences(^a), ( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3103 (55.8)</td>
<td>551 (56.1)</td>
<td>( 0.87 )</td>
</tr>
<tr>
<td>Male</td>
<td>2456 (44.2)</td>
<td>431 (43.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Age groups (years), n (%)</strong></td>
<td></td>
<td></td>
<td>(&lt;.001)</td>
</tr>
<tr>
<td>18-24</td>
<td>289 (5.2)</td>
<td>129 (13.1)</td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>584 (10.5)</td>
<td>206 (21)</td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>946 (17)</td>
<td>146 (14.9)</td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>1079 (19.4)</td>
<td>118 (12.)</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>1028 (18.5)</td>
<td>123 (12.5)</td>
<td></td>
</tr>
<tr>
<td>≥65</td>
<td>1498 (26.9)</td>
<td>225 (22.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Educational attainment, n (%)</strong></td>
<td></td>
<td></td>
<td>(&lt;.001)</td>
</tr>
<tr>
<td>≤Primary</td>
<td>711 (12.8)</td>
<td>130 (13.2)</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>2555 (46)</td>
<td>359 (36.6)</td>
<td></td>
</tr>
<tr>
<td>≥Tertiary</td>
<td>2207 (39.7)</td>
<td>474 (48.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Employment status, n (%)</strong></td>
<td></td>
<td></td>
<td>(&lt;.001)</td>
</tr>
<tr>
<td>Employed</td>
<td>2773 (49.9)</td>
<td>476 (48.5)</td>
<td></td>
</tr>
<tr>
<td>Students, home makers, or retirees</td>
<td>2524 (45.3)</td>
<td>434 (44.2)</td>
<td></td>
</tr>
<tr>
<td>Unemployed(^b)</td>
<td>175 (3.1)</td>
<td>56 (5.7)</td>
<td></td>
</tr>
<tr>
<td>Chronic condition (yes), n (%)</td>
<td>1546 (27.8)</td>
<td>279 (28.4)</td>
<td>(&lt;.001)</td>
</tr>
</tbody>
</table>

\(^a\)Differences in demographic distributions between participants who provided reasons for accepting a COVID-19 vaccine and those provided reasons for being hesitant or resistant about taking a COVID-19 vaccine.

\(^b\)Unemployed group included unemployed persons or who reported that they were seeking for jobs at the survey time.

Coded Categories of Reasons for Vaccination Acceptance and Resistance

On the basis of participants’ verbal responses to the questions asking for reasons for vaccination acceptance and resistance, 10 main categories of reasons for vaccination acceptance and vaccination resistance, respectively, were generated (Multimedia Appendix 3). The 10 main reasons for vaccination acceptance included disease risk (eg, worry about the COVID-19 risk; n=3689), protecting others (n=2354), back to normal life (n=877), confidence in vaccines (n=703), vaccine mandates (eg, complying with the vaccine pass; n=548), progovernment (eg, to support the government; n=140), social norms (eg, following significant others’ opinions; n=100), convenience...
The 10 main reasons for vaccination resistance included lack of vaccine confidence (n=501), complacency (eg, perceiving no need for vaccination or low risk of disease; n=376), poor health status (eg, “I have chronic diseases”; n=116), vaccine mandates (eg, opposing the vaccine pass; n=52), distrust in government (n=27), lack of social support (eg, “no one take me to the vaccination site”; n=18), no incentives (eg, no “lucky draw”; n=12), inconvenience (n=11), social norms (eg, “friends/family advised me not to take the vaccination”; n=7), and medical preference (eg, preferring Chinese medicine; n=6).

The coded reasons are mapped onto the 5C categories of vaccine hesitancy in Table 2. Both coders decided that 3 reasons including incentives, vaccine mandates, and medical preference were more specific to the contexts of COVID-19 vaccination and the underlying values, hence one more “C”—context was included to cover these reasons. Some reasons were mapped to >1 C to accommodate the multiple psychological antecedents of the reasons. For instance, poor health status as a reason for rejecting vaccination was mapped to both lack of vaccine confidence and calculation because it refers to not only the concern about vaccine side effects but also a trade-off between the vaccine side effects and the capability of their physical body to endure the side effects.

Table 2. Thematic coding and mapping the reasons for COVID-19 vaccination acceptance and resistance using an extensive 5Cs of vaccine hesitancy model.

<table>
<thead>
<tr>
<th>Coded reasons</th>
<th>Complacency</th>
<th>Confidence</th>
<th>Convenience</th>
<th>Calculation</th>
<th>Collective responsibility</th>
<th>Context*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reasons for vaccination acceptance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Perceived high) disease risk</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social norms</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back to normal life</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(High) confidence in vaccines</td>
<td></td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progovernment</td>
<td></td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust in experts</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Convenience</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protecting others</td>
<td></td>
<td></td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaccine mandates</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentives</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reasons for vaccination resistance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complacency (low disease risk, no need, and no urgency)</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Negative) social norms</td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of social support</td>
<td></td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of vaccine confidence</td>
<td></td>
<td></td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor health status</td>
<td></td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distrust in government</td>
<td></td>
<td></td>
<td>✓ ●</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inconvenience</td>
<td></td>
<td></td>
<td></td>
<td>✓ ●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical preference (dislike vaccination)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓ ●</td>
<td></td>
</tr>
<tr>
<td>(Dislike) vaccine mandates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓ ●</td>
<td></td>
</tr>
<tr>
<td>No incentives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓ ●</td>
<td></td>
</tr>
</tbody>
</table>

*an addition C (context) was included to accommodate vaccine mandates, incentives, and medical preference.
**Reason Co-Occurrence Network for Vaccination Acceptance**

Figure 1 shows the reason co-occurrence network for vaccination acceptance. Among the 10 accepting reasons, *disease risk* followed by *protecting others* were the most important motivators with high eigencentrality values (*disease risk*=0.80; *protecting others*=0.58). These 2 reasons were also the most frequently comentioned pair of reasons. The third most frequently mentioned reason by the group who accepted vaccination was *back to normal life* (0.12). A higher co-occurrence frequency was also identified among *disease risk, protecting others, back to normal life, confidence in vaccines, and vaccine mandates*, suggesting strong interconnections among these 5 motivators in the network. *Convenience, incentives, and trust* in experts were mainly mentioned as single motivators for vaccination acceptance.

**Figure 1.** Co-occurrence network reasons for COVID-19 vaccination acceptance and their normalized eigencentrality values provided by 5559 participants between March 1, 2021, and January 13, 2022. Larger node size indicates that the reason is more frequently mentioned as a single reason for resisting the vaccine; more thickness of the edge indicates that the 2 reasons were more frequently comentioned as reasons for accepting the vaccine.

Across age subgroups, the importance of *vaccination mandates* (0.26) and *back to normal life* (0.20) was greater for younger people, while the importance of *protecting others* (0.57) was greater for older people for motivating vaccination acceptance (Figure 2). Mean eigencentrality value for each accepting reason across age groups are provided in a table in Multimedia Appendix 4.
Figure 2. Normalized eigencentrality values of different motivators by different age groups for COVID-19 vaccination provided by 5559 participants between March 1, 2021, and January 13, 2022.

The Network for Reasons of Vaccination Resistance

Figure 3 shows the reason co-occurrence network for vaccination resistance. Lack of vaccine confidence (eigencentrality=0.89) followed by complacency (eigencentrality=0.45) were the most important reasons for vaccination resistance. Poor health status (0.09) was also a commonly cited reason among individuals with resistant attitudes toward vaccination, ranking as the third most prominent factor. Lack of vaccine confidence was most frequently comentioned with complacency, followed by poor health status, distrust in government, and dislike of vaccine mandates as demotivators of COVID-19 vaccination. These 5 demotivators formed the central part of the vaccine resistance network. Inconvenience, incentives, and medical preference (eg, dislike medical intervention including vaccination) were mostly given as the sole reasons for vaccination resistance by participants.
Figure 3. Co-occurrence network of reasons for COVID-19 vaccination resistance and their normalized eigencentrality values provided by 982 participants between December 6, 2021, and July 14, 2022. Larger node size indicates that the reason is more frequently mentioned as a single reason for resisting the vaccine; more thickness of the edge indicates that the 2 reasons were more frequently comentioned as reasons for resisting the vaccine.

Across age subgroups, complacency and poor health status were important reasons for younger and older people, respectively, for resisting vaccination (Figure 4). Mean eigencentrality value for each acceptant reason across age groups was provided in Multimedia Appendix 4.
Figure 4. Normalized eigencentrality values of demotivators by different age groups for COVID-19 vaccination provided by 982 participants between December 6, 2021, and July 14, 2022.

Discussion

Principal Findings

We used open-ended questions to elicit more nuanced and richer responses from participants to investigate the reasons for COVID-19 vaccination acceptance and resistance. We linked their verbal reasons to the 5Cs of vaccine hesitancy to improve the theoretical relevance of the findings. Most reasons for vaccination acceptance and resistance can be mapped to the 5Cs: complacency, confidence, convenience, calculation, and collective responsibility. However, one more C, namely, context, was included to accommodate vaccine mandates and incentives, which represent specific contexts of COVID-19 vaccination.

The reason co-occurrence networks showed that while disease risk (lack of complacency) and protecting others (collective or social responsibility) were the most important reasons for vaccination acceptance, lack of vaccine confidence and complacency (i.e., perceived low disease risk and perceived low importance of vaccination) were the most important reasons for vaccination resistance. Overall, a comparison of the networks for vaccination acceptance and resistance indicates that when disease risk and protecting others are salient for motivating vaccination uptake, confidence in vaccines becomes less important. This is consistent with previous findings that perceiving high personal risk of the pandemic could override people’s concern about the uncertain safety of new vaccines in vaccination decision-making [49]. However, our study indicates that when disease risk is perceived to be low and complacency exists, lack of vaccine confidence becomes the dominant reason for demotivating vaccination uptake. This explains why merely communicating vaccine benefits (e.g., protection of oneself or others) may have a minimal effect on vaccine uptake when people perceive a low risk of the disease, particularly among those who are at low risk for severe consequences of the disease [50].

For vaccination acceptance, protecting others was frequently mentioned with disease risk as a motivator for vaccination uptake, suggesting a strong connection between the perceived risk of the disease to themselves and the social responsibility to protect others [51]. Prosocial vaccination (vaccination for protecting others) has been suggested to be an important messaging strategy for promoting COVID-19 vaccination acceptance [52,53]. Our study suggests that prosocial interventions should be combined with messages for promoting perceived personal risk of the disease to enhance its persuasive effects. The age-specific network indicated that protecting others was a more important motivator for older people than it was for younger people, although theoretically, younger people should be more likely to take vaccinations to protect others because the severity of COVID-19 increased with age [54]. Vaccinating the younger population against COVID-19 to block the transmission of SARS-CoV-2 in the community was a common advocacy for promoting the young population’s vaccination.
uptake [55]. However, our results suggest that older people value social responsibility more when making vaccination decisions, which was also identified in other contexts [56,57]. A previous study found that communicating prosocial benefits is effective in increasing multiple vaccination uptake in older adults [58]. Future interventions targeted at this age group should highlight the prosocial benefits and community responsibility to motivate older adults’ vaccination uptake.

Recovering life normalcy was the third important motivator for COVID-19 vaccination acceptance and had a high frequency of co-occurrence with disease risk and protecting others. The COVID-19 vaccination or the vaccine pass had been frequently framed as a strategy to restore “new normalcy,” that is, relaxation of social distancing measures once high vaccination uptake of the population is achieved [59]. However, our study suggests that people’s vaccination motivation of returning to life normalcy is not a salient motivator; it could be simultaneously correlated with their understanding of the disease risk to themselves and others. A recent experimental study found that messages that mention the private health benefit and economic benefit (eg, vaccines enable economic recovery) can increase vaccination intention by 9% compared with the control group [60]. Future vaccination communication can highlight the cobenefits of returning life to normalcy and protecting oneself and others to achieve optimal effect.

Vaccine mandates is another factor in the context domain; it can be either a motivator or a demotivator of vaccination uptake. Although the “vaccine bubble” was framed as a benefit of vaccination for granting people greater life normalcy [7], it was also perceived to be a constraint on people’s freedom in vaccination choice, which impaired public trust in authorities [61,62]. In our coding, we combined “vaccine bubble” and “vaccine pass” as 1 code, namely, “vaccine mandates.” Our study showed that vaccine mandates as a vaccination motivator was connected to disease risk and protecting others, but as a vaccination demotivator, it was connected to lack of vaccine confidence, complacency, and distrust in government. This indicates that whether vaccine mandates act as a motivator or demotivator depends on individuals’ perceived disease risk, confidence in vaccine, and trust in the government. This finding provides practical implications for how to frame vaccine mandates in a vaccination campaign; that is, vaccine mandates, when framed as a strategy for self-protection and prosocial values by highlighting disease risk and a way for themselves and others to enjoy normal social activities, could motivate vaccination uptake. However, framing vaccine mandates as a coercive measure may be regarded as freedom violation and demotivate vaccination uptake [63]. Both life normalcy and vaccine mandates were more important for younger people as motivators for vaccination acceptance than they were for older people, and hence should be the important intervention targets for promoting younger people’s vaccination uptake in the future.

For vaccination resistance, complacency was strongly connected with lack of vaccine confidence, suggesting that when disease risk is perceived to be low, uncertain vaccine safety can become salient in hindering vaccination uptake [49]. Complacency was a more important demotivator for younger people than it was for older people, which is consistent with the fact that COVID-19 is a less severe disease for younger people [54]. Poor health status was the third important demotivator for COVID-19 vaccination, which was strongly connected with lack of vaccine confidence in the network. Such a pattern was more prevalent in older people, suggesting that the poor health concern mainly comprises concern about vaccine side effects or safety in this group. One previous qualitative study suggested that older people who perceived their health status was poor due to aging or chronic diseases tended to perceive themselves to have lower capability to endure the vaccine side effects [40]. Such perception was linked to their value of aging and lack of social support [40]. This highlights the importance of addressing people’s concerns about their health status for addressing vaccine hesitancy through enhancing confidence in vaccines, reshaping the value of aging, and enhancing social support [64].

Both the networks for vaccination acceptance and resistance showed that convenience (or inconvenience) and incentives (or no incentives) mainly represented as single reasons for motivating or demotivating COVID-19 vaccination uptake. In Hong Kong, convenience in accessing the COVID-19 vaccine had been greatly increased by setting up multiple vaccination sites, extending service hours for vaccination services, and offering the vaccines free for all [46]. In addition, the government offered vaccination leaves and relaxation of social distancing measures, while a series of lucky draws were sponsored by the business sector as reward strategies to boost vaccination uptake. Although convenience and incentives (or perceiving no convenience or incentives) can serve as independent cues or heuristics to motivate or demotivate vaccination uptake [65], the low frequency of these reasons and their weak connections with other reasons indicate that they have limited and mainly transient effects on COVID-19 vaccination uptake [66].

Limitations

Our study had several limitations. First, although participants were encouraged to give >1 reason for accepting, refusing, or being hesitant about taking COVID-19 vaccination, the reasons for COVID-19 vaccination acceptance and resistance could not be exhaustively explored through a telephone survey. We assumed that the reasons mentioned by participants were more mentally accessible for them and thereby more salient in influencing their vaccination decision. Nevertheless, there is still a possibility that some underlying reasons for vaccination acceptance or resistance were hidden by the participants, consciously or unconsciously. Second, our network analysis focused on the co-occurrence of reasons but could not determine the causal relationships between the reasons. Third, although 2 coders independently coded the reasons to ensure the reliability of coding, there remains uncertainty in making sense of participants’ verbal responses and the combination of reasons into main categories. Despite this, we strictly documented the procedure and ensured transparency of the decision-making process to improve the trustworthiness of our findings.

Conclusions

Perception of personal risk to disease and the social responsibility to protect others were the most important comotivators, while lack of vaccine confidence and complacency...
were the most important co-demotivators for COVID-19 vaccination. For COVID-19 vaccination acceptance, recovering life normalcy, confidence in vaccines, and vaccine mandates were additional motivators, but these reasons were likely to work based on people’s understanding of the disease risk to themselves and others. For vaccination resistance, perception of poor health status, distrust in the government, and dislike of vaccine mandates were additional demotivators, all of which were linked to lack of vaccine confidence and complacency. Convenience and immediate incentives for vaccination were mainly mentioned as the single reasons for accepting or resisting the COVID-19 vaccination. For older people, protecting others was a more important motivator, while perception of poor health status was a more important demotivator. For younger people, recovering normal life and vaccine mandates were more important motivators, while complacency was a more important demotivator for the COVID-19 vaccination.

Acknowledgments
This study was supported by the Health and Medical Research Fund of the Food and Health Bureau of the Hong Kong SAR Government (grant number: COVID19F04; grant number: COVID19F11), and the Strategic Topic Grants Scheme (grant number: STG4/M-701/23-N) of the Research Grant Council of the Hong Kong Special Administrative Region, China. BJC is supported by an RGC Senior Research Fellowship (grant number: HKU SRFS2021-7S03).

Data Availability
The data set generated during and analyzed during this study is available from the corresponding author on reasonable request.

Authors’ Contributions
QL conceptualized the study, designed the questionnaire, obtained the funding, analyzed the data, and drafted the initial manuscript; JY conceptualized the study, analyzed the data, and drafted the initial manuscript; IOLW designed the questionnaire, coordinated, and supervised data collection, and revised the manuscript; MYN conceptualized the study, designed the questionnaire, obtained the funding, and revised the manuscript; BJC conceptualized the study, designed the questionnaire, obtained the funding, and revised the manuscript; WWTTL conceptualized the study and revised the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Conflicts of Interest
BJC consults for AstraZeneca, Fosun Pharma, GlaxoSmithKline, Haleon, Moderna, Novavax, Pfizer, Roche, and Sanofi Pasteur, and has received research funding from Fosun Pharma. All other authors report no potential conflicts.

Multimedia Appendix 1
An overview of study survey rounds.

Multimedia Appendix 2
Data collection period and number of daily COVID-19 cases.

Multimedia Appendix 3
Coded categories of reasons for COVID-19 vaccination acceptance and resistance.

Multimedia Appendix 4
Mean eigencentrality values of vaccination acceptance and resistance reasons by age groups.

References


17. Liao et alJMIR PUBLIC HEALTH AND SURVEILLANCE


Edited by A Mavragani, T Sanchez; submitted 18.07.23; peer-reviewed by E Bønes, B Eshrati; comments to author 04.10.23; revised version received 21.10.23; accepted 27.02.24; published 22.04.24.

Please cite as:
Liao Q, Yuan J, Wong IOL, Ni MY, Cowling BJ, Lam WWT
Motivators and Demotivators for COVID-19 Vaccination Based on Co-Occurrence Networks of Verbal Reasons for Vaccination Acceptance and Resistance: Repetitive Cross-Sectional Surveys and Network Analysis
JMIR Public Health Surveill 2024;10:e50958
URL: https://publichealth.jmir.org/2024/1/e50958
doi:10.2196/50958
PMID:38648099

©Qiuyan Liao, Jiehu Yuan, Irene Oi Ling Wong, Michael Yuxuan Ni, Benjamin John Cowling, Wendy Wing Tak Lam. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 22.04.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Ischemic Stroke After Bivalent COVID-19 Vaccination: Self-Controlled Case Series Study

Stanley Xu¹, PhD; Lina S Sy¹, MPH; Vennis Hong¹, MPH; Kimberly J Holmquist¹, MPH; Lei Qian¹, PhD; Paddy Farrington², PhD; Katia J Bruxvoort¹, PhD; Nicola P Klein³, MD; Bruce Fireman⁴, MS; Bing Han¹, PhD; Bruno J Lewin¹, MD

¹Department of Research & Evaluation, Southern California Permanente Medical Group, Pasadena, CA, United States
²School of Mathematics and Statistics, The Open University, United Kingdom
³Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, United States
⁴Kaiser Permanente Vaccine Study Center, Kaiser Permanente Northern California, Oakland, CA, United States

Corresponding Author:
Stanley Xu, PhD
Department of Research & Evaluation
Southern California Permanente Medical Group
100 S Los Robles
Pasadena, CA, 91101
United States
Phone: 1 6265643958
Email: stan.xu@kp.org

Abstract

Background: The potential association between bivalent COVID-19 vaccination and ischemic stroke remains uncertain, despite several studies conducted thus far.

Objective: This study aimed to evaluate the risk of ischemic stroke following bivalent COVID-19 vaccination during the 2022-2023 season.

Methods: A self-controlled case series study was conducted among members aged 12 years and older who experienced ischemic stroke between September 1, 2022, and March 31, 2023, in a large health care system. Ischemic strokes were identified using International Classification of Diseases, Tenth Revision codes in emergency departments and inpatient settings. Exposures were Pfizer-BioNTech or Moderna bivalent COVID-19 vaccination. Risk intervals were prespecified as 1-21 days and 1-42 days after bivalent vaccination; all non-risk-interval person-time served as the control interval. The incidence of ischemic stroke was compared in the risk interval and control interval using conditional Poisson regression. We conducted overall and subgroup analyses by age, history of SARS-CoV-2 infection, and coadministration of influenza vaccine. When an elevated risk was detected, we performed a chart review of ischemic strokes and analyzed the risk of chart-confirmed ischemic stroke.

Results: With 4933 ischemic stroke events, we found no increased risk within the 21-day risk interval for the 2 vaccines and by subgroups. However, risk of ischemic stroke was elevated within the 42-day risk interval among individuals aged younger than 65 years with coadministration of Pfizer-BioNTech bivalent and influenza vaccines on the same day; the relative incidence (RI) was 2.13 (95% CI 1.01-4.46). Among those who also had a history of SARS-CoV-2 infection, the RI was 3.94 (95% CI 1.10-14.16). After chart review, the RIs were 2.34 (95% CI 0.97-5.65) and 4.27 (95% CI 0.97-18.85), respectively. Among individuals aged younger than 65 years who received Moderna bivalent vaccine and had a history of SARS-CoV-2 infection, the RI was 2.62 (95% CI 1.13-6.03) before chart review and 2.24 (95% CI 0.78-6.47) after chart review. Stratified analyses by sex did not show a significantly increased risk of ischemic stroke after bivalent vaccination.

Conclusions: While the point estimate for the risk of chart-confirmed ischemic stroke was elevated in a risk interval of 1-42 days among individuals younger than 65 years with coadministration of Pfizer-BioNTech bivalent and influenza vaccines on the same day and among individuals younger than 65 years who received Moderna bivalent vaccine and had a history of SARS-CoV-2 infection, the risk was not statistically significant. The potential association between bivalent vaccination and ischemic stroke in the 1-42-day analysis warrants further investigation among individuals younger than 65 years with influenza vaccine coadministration and prior SARS-CoV-2 infection. Furthermore, the findings on ischemic stroke risk after bivalent COVID-19 vaccination underscore the need to evaluate monovalent COVID-19 vaccine safety during the 2023-2024 season.
ischemic stroke; bivalent COVID-19 vaccine; influenza vaccine; self-controlled case series; coadministration; ischemic; stroke; TIA; transient ischemic attack; ischemia; cardiovascular; COVID-19; SARS-CoV-2; vaccine; vaccines; vaccination; association; correlation; risk; risks; adverse; side effect; subgroup analyses; subgroup analysis; bivalent; influenza; infectious; respiratory; incidence; case series

Introduction

On August 31, 2022, the US Food and Drug Administration (FDA) granted emergency use authorizations for the Pfizer-BioNTech bivalent COVID-19 vaccine for individuals aged 12 years and older and the Moderna bivalent COVID-19 vaccine for individuals aged 18 years and older [1,2]. The bivalent vaccines contain messenger ribonucleic acid (mRNA) components derived from both the original strain of SARS-CoV-2 and the Omicron variant BA.4 and BA.5 sublineages. Designed to be administered as a single booster dose, bivalent COVID-19 vaccines were recommended to be given ≥60 days after either the primary vaccination or a monovalent booster dose [2]. In the context of waning protection from primary vaccination, bivalent vaccines enhanced the immune response and boosted protection against the virus, offering an additional layer of defense for previously vaccinated individuals [3-5].

Safety data for bivalent mRNA COVID-19 vaccines were initially limited. Because the chemical components and production processes between monovalent and bivalent vaccines were similar, the FDA granted emergency use authorizations for the bivalent COVID-19 vaccines based on safety data for monovalent vaccines, as well as limited bivalent safety data from clinical trials [1,2]. To monitor safety postlicensure, a study using V-safe and the Vaccine Adverse Event Reporting System (VAERS) examined bivalent booster vaccinations in individuals aged older than or equal to 12 years and found that the safety profile was similar to that described for monovalent booster vaccinations [6]. A recent study that comprehensively assessed potential adverse events associated with bivalent vaccines using TreeScan in the Vaccine Safety Datalink (VSD) network found no increased risk for a broad range of adverse events [7].

The VSD has monitored COVID-19 vaccine safety since vaccinations began in December 2020 [8]. In late 2022, VSD’s rapid cycle analyses detected a safety signal for ischemic stroke following the Pfizer-BioNTech COVID-19 bivalent booster vaccination among those 65 years and older, particularly among those who had received a bivalent booster dose and a high-dose or adjuvanted influenza vaccine on the same day (coadministration) [9]. The US Centers for Disease Control and Prevention (CDC) and FDA announced this safety signal in January 2023 [10]. This safety signal attenuated as data accumulated [11]. Another cohort study among adults aged 65 years and older reported that those who received the Pfizer-BioNTech bivalent booster had a similar hazard for ischemic stroke encounters compared to those who received the Moderna bivalent booster vaccine, but had a lower hazard than those who received the Pfizer-BioNTech or Moderna monovalent boosters [12]. In another study, compared to monovalent vaccination, bivalent vaccination was not found to be associated with an increased risk of ischemic stroke, hemorrhagic stroke, myocardial infarction, and pulmonary embolism [13]. A self-controlled case series (SCCS) study conducted in England showed no indication of an increased risk of ischemic stroke risk within 21 days following administration of either of the 2 mRNA COVID-19 bivalent vaccines. Similar findings were observed for individuals aged 65 years and older who received the influenza vaccine concurrently with the bivalent COVID-19 vaccines [14]. Another SCCS study conducted in Israel also did not find an increased risk of ischemic stroke following monovalent or bivalent mRNA COVID-19 vaccine boosters in at-risk populations [15]. A study of Medicare beneficiaries aged 65 years or older showed no significantly elevated risk for stroke immediately after receiving either COVID-19 bivalent vaccine [16]. However, among beneficiaries who had a stroke after getting either COVID-19 bivalent vaccine along with a high-dose or adjuvanted influenza vaccine, there was a significant association between vaccination and nonhemorrhagic stroke within 22 to 42 days for the Pfizer-BioNTech COVID-19 bivalent vaccine. Additionally, there was a significant association between vaccination and transient ischemic attack within 1 to 21 days for the Moderna COVID-19 bivalent vaccine.

The objective of this study was to assess the risk of ischemic stroke after bivalent COVID-19 vaccination among individuals enrolled in Kaiser Permanente Southern California (KPSC) using a modified SCCS design. Subgroup analyses were also conducted by age (younger than 65 years vs older than or equal to 65 years), history of SARS-CoV-2 infection, and coadministration of influenza vaccine.

Methods

Study Population and Study Period

We conducted an SCCS study among members aged 12 years and older from KPSC, a large integrated health care system in the United States. The SCCS method is an alternative to standard epidemiological study designs and has been used for evaluating vaccine safety and in fields such as pharmacoepidemiology [17-21]. In an SCCS study, only individuals who have experienced an event are included. Since individuals serve as their own control and the incidence rates of the outcome of interest are compared within individuals, all time-invariant confounding variables are controlled. The SCCS analytic data sets included individuals who experienced ischemic stroke events between September 1, 2022, and March 31, 2023, had completed a COVID-19 vaccine primary series, and had received their last monovalent dose ≥60 days before September 1, 2022. We required KPSC membership on September 1, 2022.
Exposure and Observation Period

The exposure was defined as the administration of the Pfizer-BioNTech bivalent COVID-19 vaccine (for individuals aged 12 years and older) or the Moderna bivalent COVID-19 vaccine (for individuals aged 18 years and older) between September 1, 2022, and March 31, 2023. The observation period for the recipients of a bivalent COVID-19 vaccine started on September 1, 2022, and ended on March 31, 2023, or upon death, receipt of the second bivalent dose, or disenrollment, whichever came first.

To adjust for seasonality, we also included ischemic stroke events occurring among eligible individuals aged 12 years and older who had completed a primary series and had received their last monovalent dose ≥60 days before September 1, 2022, but who did not receive a bivalent vaccine during September 1, 2022, to March 31, 2023 (nonbivalent recipients; NBRs). The observation period for ischemic stroke events among NBR started on September 1, 2022, and ended on March 31, 2023, or upon death or disenrollment, whichever came first.

Outcome

The outcome was defined as the first occurrence of an ischemic stroke event between September 1, 2022, and March 31, 2023 [22]. Ischemic stroke events were identified through medical encounters with an International Classification of Diseases, Tenth Revision (ICD-10) diagnosis code of G45.8, G45.9, or I63.x, where x represents any additional characters in the ICD-10 code range for ischemic stroke, in the emergency departments or inpatient settings. We also looked back 30 days prior to September 1, 2022, to ensure that the episode was incident. We excluded ischemic stroke events due to other possible causes and adjusted the onset date (details in Multimedia Appendix 1). We considered these ischemic stroke events that were identified with ICD-10 codes to be electronically identified ischemic stroke events.

Covariates

We collected demographic variables (age, sex, and race or ethnicity) and the Charlson Comorbidity Index to describe the characteristics of the study population, as well as concomitant influenza vaccination during the study period and history of SARS-CoV-2 infection in the year prior to September 1, 2022.

Statistical Analyses

We assessed the risk of ischemic stroke following the administration of the Pfizer-BioNTech and Moderna bivalent COVID-19 vaccines separately. Demographic characteristics of individuals who experienced ischemic stroke events during the study period were described among Pfizer-BioNTech bivalent vaccine recipients, Moderna bivalent vaccine recipients, and NBRs.

The risk intervals were prespecified as 1-21 days and 1-42 days after administration of bivalent COVID-19 vaccines, with person-time outside of these risk intervals serving as the control interval. The risk intervals started on the day of vaccination (day 1). Because individuals who had ischemic stroke events might be likely to postpone or avoid bivalent vaccination, we used a modified SCCS approach for event-dependent exposures [22]. The modified SCCS used a pseudo-likelihood approach in the counterfactual framework to estimate the relative incidence (RI) and 95% CIs of events comparing the risk intervals to their corresponding control intervals by maximizing a Poisson pseudolikelihood [23]. In the SCCS analyses, ischemic stroke events occurring among eligible individuals who did not receive the bivalent vaccines were included to account for seasonality, by incorporating calendar month into the model [22]. Given that age did not significantly vary during the relatively short observation period of 7 months, it was not adjusted as a time-varying covariate.

Additionally, we performed several subgroup analyses based on age (younger than 65 vs 65 years and older), coadministration of bivalent COVID-19 vaccine with same-day influenza vaccine (yes or no), and history of SARS-CoV-2 infection (confirmed by a positive laboratory test or a COVID-19 diagnosis) within 1 year prior to September 1, 2022.

When a safety signal (ie, the lower bound of the 95% CI for RI exceeded 1.0) was detected in analyses of electronically identified ischemic stroke events, we conducted a chart review among recipients of bivalent COVID-19 vaccines to confirm ischemic stroke events and identify onset date to determine whether confirmed ischemic stroke events fell in the risk or control interval; confirmation rates were then calculated (number of confirmed events divided by the number of electronically identified events reviewed). We did not conduct a chart review on ischemic stroke events among NBR due to the large number of events in this group and limited resources. In analyses of confirmed ischemic stroke events among recipients of bivalent COVID-19 vaccines, we introduced a randomized allocation of confirmed case status to the NBR group. This allocation was guided by the confirmation rates observed among recipients of bivalent COVID-19 vaccines, as outlined by Xu et al [24]. A total of 5 simulated data sets were generated to replicate the allocation process. The SCCS analyses were conducted on each data set, and the resulting estimates were aggregated using Rubin’s [25] rule, which accounts for both the variability within individual data sets and the variability across multiple data sets. Attributable risk (AR) was calculated using the approach described in Farrington et al [26]:

Here, RI is the relative incidence; \( n_0 \) is the number of ischemic stroke events in the risk interval; and \( N \) is the number of recipients of a vaccine or dose. The reciprocal of AR is the number needed to harm (NNH). Analyses were conducted using SAS (version 9.4; SAS Institute) and the SCCS models were fitted with the R package (R Core Team) SCCS [27].

Ethical Considerations

Ethics approval for this study was obtained from the KPSC institutional review board on June 6, 2023. In accordance with 45CFR 46.116, informed consent was waived by the institutional review board because the research activities (secondary analyses of electronic health records data) presented no more than minimal risk to participants. Patients or members of the public were not involved in the design, conduct, reporting, or
dissemination plans of the research. To protect the privacy and confidentiality of human participants, all staff working on the research study were trained in procedures to protect the privacy of medical record information. All research data were stored behind a firewall in a password-protected network within the Department of Research & Evaluation at KPSC. Study participants were not compensated given the observational nature of the study.

Results

Characteristics of Individuals Who Had Ischemic Stroke Events

Table 1 shows the characteristics of individuals who had ischemic stroke events. In total, there were 1057 ischemic stroke events among recipients of the Pfizer-BioNTech bivalent vaccine with a mean length of observation period of 204 (SD 27) days (ranging from 16 to 212 days), 827 ischemic stroke events among recipients of the Moderna bivalent vaccine with a mean length of observation period of 206 (SD 24) days (ranging from 31 to 212 days), and 3049 ischemic stroke events among NBR with a mean length of observation period of 197 (SD 40) days (ranging from 11 to 212 days). Notably, the majority of ischemic stroke events occurred among individuals aged 65 years or older. Those aged younger than 65 years old had fewer comorbidities than those aged 65 years or older (Multimedia Appendix 2).

Table 1. Characteristics of individuals who had ischemic stroke events among members of Kaiser Permanente Southern California during the period from September 1, 2022, to March 31, 2023.

<table>
<thead>
<tr>
<th>Age (years), n (%)</th>
<th>Recipients of Pfizer-BioNTech bivalent COVID-19 vaccine (n=1057)</th>
<th>Recipients of Moderna bivalent COVID-19 vaccine (n=827)</th>
<th>Nonrecipients of bivalent vaccines (n=3049)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-17</td>
<td>1 (0.1)</td>
<td>N/A</td>
<td>6 (0.2)</td>
</tr>
<tr>
<td>18-44</td>
<td>32 (3.0)</td>
<td>15 (1.8)</td>
<td>261 (8.6)</td>
</tr>
<tr>
<td>45-64</td>
<td>245 (23.2)</td>
<td>177 (21.4)</td>
<td>1011 (33.2)</td>
</tr>
<tr>
<td>65-74</td>
<td>618 (58.5)</td>
<td>501 (60.6)</td>
<td>1406 (46.1)</td>
</tr>
<tr>
<td>≥75</td>
<td>161 (15.2)</td>
<td>134 (16.2)</td>
<td>365 (12.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex, n (%)</th>
<th>Recipients of Pfizer-BioNTech bivalent COVID-19 vaccine (n=1057)</th>
<th>Recipients of Moderna bivalent COVID-19 vaccine (n=827)</th>
<th>Nonrecipients of bivalent vaccines (n=3049)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>540 (51.1)</td>
<td>446 (53.9)</td>
<td>1637 (53.7)</td>
</tr>
<tr>
<td>Male</td>
<td>517 (48.9)</td>
<td>381 (46.1)</td>
<td>1412 (46.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race or ethnicity, n (%)</th>
<th>Recipients of Pfizer-BioNTech bivalent COVID-19 vaccine (n=1057)</th>
<th>Recipients of Moderna bivalent COVID-19 vaccine (n=827)</th>
<th>Nonrecipients of bivalent vaccines (n=3049)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>311 (29.4)</td>
<td>259 (31.3)</td>
<td>1186 (38.9)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>470 (44.5)</td>
<td>326 (39.4)</td>
<td>1027 (33.7)</td>
</tr>
<tr>
<td>Non-Hispanic Asian</td>
<td>119 (11.3)</td>
<td>99 (12.0)</td>
<td>270 (8.9)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>139 (13.1)</td>
<td>120 (14.5)</td>
<td>472 (15.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>6 (0.6)</td>
<td>9 (1.1)</td>
<td>37 (1.2)</td>
</tr>
<tr>
<td>Multiple or other</td>
<td>12 (1.1)</td>
<td>14 (1.7)</td>
<td>57 (1.9)</td>
</tr>
<tr>
<td>Length of observation period in days, mean (SD)</td>
<td>204.3 (26.8)</td>
<td>205.7 (23.5)</td>
<td>196.7 (39.9)</td>
</tr>
<tr>
<td>Death, n (%)</td>
<td>50 (4.7)</td>
<td>37 (4.5)</td>
<td>309 (10.1)</td>
</tr>
</tbody>
</table>

*N/A: not applicable.

Risk of Ischemic Stroke Following the Pfizer-BioNTech Bivalent COVID-19 Vaccine

For the Pfizer-BioNTech bivalent COVID-19 vaccine, there were 103 electronically identified ischemic stroke events in the 21-day postvaccination risk interval, 954 events in the control interval, and 3049 events among NBR; the overall RI was 0.90 (95% CI 0.73-1.12; Multimedia Appendix 3). The RI was not significantly above 1 across all subgroup analyses by age, coadministration of influenza vaccine, and history of SARS-CoV-2 infection. In analyses extending the risk interval to 1-42 days following bivalent vaccination, the overall RI was 0.97 (95% CI 0.81-1.15; Table 2). However, in subgroup analyses using the 1-42–day risk interval, we observed an increased risk of ischemic stroke only among individuals younger than 65 years of age who also received an influenza vaccine on the same day. The RI in this subgroup was 2.13 (95% CI 1.01-4.46). Among the subset who also had a documented history of SARS-CoV-2 infection within the year preceding the study period, the RI increased to 3.94 (95% CI 1.10-14.16).
In this 1-42–day risk interval analysis of the specific subgroup of individuals aged younger than 65 years who received bivalent and influenza vaccines on the same day, chart review of the 37 electronically identified ischemic stroke events found that 2 were determined to be hemorrhagic strokes and 11 were subsequently found to not meet the criteria for true ischemic stroke events, yielding a confirmation rate of 65% (n=24). With the verified 24 ischemic stroke events and ischemic stroke events among NBR (not verified through chart review, but adjusted for using a 65% confirmation rate), we proceeded to reevaluate the RI in this subgroup. The number of confirmed ischemic stroke events was graphed over the interval in days between bivalent vaccination and ischemic stroke event (Figure 1). There were 10 events in the 1-42–day risk interval and 14 events in the control interval. Using a risk interval of 1-42 days after coadministration of the Pfizer-BioNTech bivalent vaccine and influenza vaccine, the overall RI derived from analyzing confirmed ischemic stroke events among those aged younger than 65 years was 2.34 (95% CI 0.97-5.65; \(P=0.06\); Table 3). Between September 1, 2022, and March 31, 2023, a total of 117,423 individuals aged younger than 65 years received the Pfizer-BioNTech bivalent vaccine and influenza vaccine on the same day. According to the equation, \(AR=4.88 \times 10^{-5}\) and \(NNH=20,505\) with a risk interval of 1-42 days. Among 21,128 individuals who also had a documented history of SARS-CoV-2 infection within the year preceding the study period, the RI increased to 4.27 (95% CI 0.97-18.85; \(P=0.06\); Table 3); according to the equation, \(AR=1.45 \times 10^{-4}\) and \(NNH=6897\) with a risk interval of 1-42 days. Among the 10 confirmed ischemic stroke events in the risk interval of 1-42 days, the mean age was 58 (SD 5) years, ranging from 48 to 63 years. Among these cases, 2 individuals had a documented history of previous ischemic stroke, and no one died as of March 31, 2023. In addition, 7 received the standard dose, egg-based quadrivalent influenza vaccine, while 3 received an influenza vaccine of unknown formulation.
<table>
<thead>
<tr>
<th>All ages</th>
<th>Risk interval</th>
<th>Control interval</th>
<th>NBRs</th>
<th>Relative incidence (95% CI)</th>
<th>Number of events, n</th>
<th>Relative incidence (95% CI)</th>
<th>Number of events, n</th>
<th>Relative incidence (95% CI)</th>
<th>Number of events, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;65 years old</td>
<td>Risk interval</td>
<td>Control interval</td>
<td>NBRs</td>
<td>Relative incidence (95% CI)</td>
<td>Number of events, n</td>
<td>Relative incidence (95% CI)</td>
<td>Number of events, n</td>
<td>Relative incidence (95% CI)</td>
<td>Number of events, n</td>
</tr>
<tr>
<td>≥65 years old</td>
<td>Risk interval</td>
<td>Control interval</td>
<td>NBRs</td>
<td>Relative incidence (95% CI)</td>
<td>Number of events, n</td>
<td>Relative incidence (95% CI)</td>
<td>Number of events, n</td>
<td>Relative incidence (95% CI)</td>
<td>Number of events, n</td>
</tr>
</tbody>
</table>

### All recipients

| Overall | 212 | 845 | 3049 | 0.97 (0.81-1.15) | 50 | 228 | 1278 | 0.98 (0.69-1.38) | 162 | 617 | 1771 | 0.96 (0.79-1.17) |
| With history of SARS-CoV-2 | 35 | 139 | 565 | 0.98 (0.66-1.47) | 15 | 53 | 292 | 1.22 (0.64-2.33) | 20 | 86 | 273 | 0.83 (0.50-1.40) |
| Without history of SARS-CoV-2 | 177 | 706 | 2484 | 0.97 (0.80-1.17) | 35 | 175 | 986 | 0.91 (0.61-1.38) | 142 | 531 | 1498 | 0.99 (0.80-1.22) |

### Coadministration of influenza vaccine

| Overall | 41 | 126 | 3049 | 0.97 (0.66-1.41) | 15 | 22 | 1278 | 2.13 (1.01-4.46) | 26 | 104 | 1771 | 0.75 (0.47-1.19) |
| With history of SARS-CoV-2 | 9 | 20 | 565 | 1.26 (0.56-2.82) | 6 | 4 | 292 | 3.94 (1.10-14.16) | 3 | 16 | 273 | 0.51 (0.14-1.90) |
| Without history of SARS-CoV-2 | 32 | 106 | 2484 | 0.91 (0.59-1.40) | 9 | 18 | 986 | 1.57 (0.61-4.05) | 23 | 88 | 1498 | 0.80 (0.49-1.31) |

### No coadministration of influenza vaccine

| Overall | 171 | 719 | 3049 | 0.96 (0.80-1.17) | 35 | 206 | 1278 | 0.81 (0.54-1.20) | 136 | 513 | 1771 | 1.01 (0.81-1.25) |
| With history of SARS-CoV-2 | 26 | 119 | 565 | 0.89 (0.56-1.40) | 9 | 49 | 292 | 0.82 (0.37-1.80) | 17 | 70 | 273 | 0.88 (0.50-1.55) |
| Without history of SARS-CoV-2 | 145 | 600 | 2484 | 0.99 (0.80-1.22) | 26 | 157 | 986 | 0.82 (0.52-1.29) | 119 | 443 | 1498 | 1.04 (0.82-1.31) |

---

*NBR: nonbivalent recipient. These were eligible individuals who did not receive a bivalent vaccine but had completed a primary series of COVID-19 vaccination and had their last monovalent dose ≥60 days before September 1, 2022. Inclusion of these events helps to adjust for temporal trends (seasonality). The same NBR population was used in overall bivalent analyses, as well as bivalent analyses stratified by coadministration of influenza vaccine.

Had SARS-CoV-2 infection (ie, SARS-CoV-2 positive laboratory test or a COVID-19 diagnosis) during the year prior (August 31, 2021-August 31, 2022).
Figure 1. Number of confirmed ischemic stroke events over the interval in days between bivalent vaccination and ischemic stroke event among those who received the Pfizer-BioNTech bivalent vaccine and influenza vaccine on the same day among members of Kaiser Permanente Southern California during the period from September 1, 2022, to March 31, 2023.

Table 3. Numbers of confirmed ischemic stroke events among recipients of the Pfizer-BioNTech bivalent COVID-19 vaccine aged <65 years, and relative incidences in the 42 days after coadministration of bivalent and influenza vaccines among members of Kaiser Permanente Southern California during the period from September 1, 2022, to March 31, 2023.

<table>
<thead>
<tr>
<th></th>
<th>Number of events, n</th>
<th>Risk interval</th>
<th>Control interval</th>
<th>NBRs (^b)</th>
<th>Relative incidence (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>10</td>
<td>14</td>
<td>835</td>
<td></td>
<td>2.34 (0.97-5.65)</td>
</tr>
<tr>
<td>With history of SARS-CoV-2(^c)</td>
<td>4</td>
<td>3</td>
<td>190</td>
<td></td>
<td>4.27 (0.97-18.85)</td>
</tr>
<tr>
<td>Without history of SARS-CoV-2</td>
<td>6</td>
<td>11</td>
<td>645</td>
<td></td>
<td>1.76 (0.57-5.45)</td>
</tr>
</tbody>
</table>

\(^a\) Confirmation by chart review.

\(^b\) NBR: nonbivalent recipient. These were eligible individuals who did not receive a bivalent vaccine but had completed a primary series of COVID-19 vaccination and had their last monovalent dose ≥60 days before September 1, 2022. Inclusion of these events helps to adjust for temporal trends (seasonality). A confirmation rate of 65% was applied to ischemic stroke events among NBR.

\(^c\) Had SARS-CoV-2 infection (ie, SARS-CoV-2 positive laboratory test or a COVID-19 diagnosis) during the year prior (August 31, 2021-August 31, 2022).

Risk of Ischemic Stroke Following the Moderna Bivalent COVID-19 Vaccine

When using a risk interval of 21 days following Moderna bivalent vaccination, the overall risk of ischemic stroke was not elevated from the analysis of electronically identified ischemic stroke events (RI=0.91, 95% CI 0.71-1.15), a finding that held true across all subgroup analyses by age, coadministration of influenza vaccine, and history of SARS-CoV-2 infection (Multimedia Appendix 4). However, extending the risk interval to 42 days after Moderna bivalent vaccination showed an increased risk of ischemic stroke among individuals younger than 65 years of age who had a documented history of SARS-CoV-2 infection, with an RI of 2.62 (95% CI 1.13-6.03). This subgroup involved a total of 36 ischemic stroke events among recipients of the Moderna bivalent COVID-19 vaccine (Table 4).

Of the 36 ischemic stroke events, 1 was a hemorrhagic stroke and 12 were not confirmed as true events through medical chart review, yielding a confirmation rate of 64% (n=23). After using a risk interval of 1-42 days following the Moderna bivalent vaccination and applying a 64% confirmation rate to ischemic stroke events among NBR, the RI derived from analyzing confirmed ischemic stroke events among those aged younger than 65 years who had a documented history of SARS-CoV-2 infection was 2.24 (95% CI 0.78-6.47; \(P=0.14\)).
Table 4. Numbers of electronically identified ischemic stroke events and relative incidences in the 42 days after Moderna bivalent COVID-19 vaccination among members of Kaiser Permanente Southern California during the period from September 1, 2022, to March 31, 2023.

<table>
<thead>
<tr>
<th></th>
<th>All ages</th>
<th>&lt;65 years old</th>
<th>≥65 years old</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk interval</td>
<td>Control interval</td>
<td>NBR</td>
</tr>
<tr>
<td>All recipients</td>
<td>161</td>
<td>666</td>
<td>3049</td>
</tr>
<tr>
<td>With history of SARS-CoV-2b</td>
<td>28</td>
<td>111</td>
<td>565</td>
</tr>
<tr>
<td>Without history of SARS-CoV-2</td>
<td>133</td>
<td>555</td>
<td>2484</td>
</tr>
<tr>
<td>Coadministration of influenza vaccine, overall</td>
<td>19</td>
<td>72</td>
<td>3049</td>
</tr>
<tr>
<td>With history of SARS-CoV-2b</td>
<td>3</td>
<td>14</td>
<td>565</td>
</tr>
<tr>
<td>Without history of SARS-CoV-2</td>
<td>16</td>
<td>58</td>
<td>2484</td>
</tr>
<tr>
<td>No coadministration of influenza vaccine, overall</td>
<td>142</td>
<td>594</td>
<td>3049</td>
</tr>
<tr>
<td>With history of SARS-CoV-2b</td>
<td>25</td>
<td>97</td>
<td>565</td>
</tr>
<tr>
<td>Without history of SARS-CoV-2</td>
<td>117</td>
<td>497</td>
<td>2484</td>
</tr>
</tbody>
</table>

aNBR: nonbivalent recipient. These were eligible individuals who did not receive a bivalent vaccine but had completed a primary series of COVID-19 vaccination and had their last monovalent dose ≥60 days before September 1, 2022. Inclusion of these events helps to adjust for temporal trends (seasonality). The same NBR population was used in overall bivalent analyses, as well as bivalent analyses stratified by coadministration of influenza vaccine.

bHad SARS-CoV-2 infection (ie, SARS-CoV-2 positive laboratory test or a COVID-19 diagnosis) during the year prior (August 31, 2021-August 31, 2022).

cN/A: not applicable.

Sex-Stratified Analyses
Further analyses were conducted to examine if the risk of ischemic stroke after bivalent vaccination differed by sex for those analyses indicating a potential increase in risk of ischemic stroke (Multimedia Appendix 5). The risk of ischemic stroke was not significantly increased after bivalent vaccination in the sex-stratified analyses. Due to limited sample sizes, the CIs of RIs were wide. The SAS and R codes are available for preparing data and fitting the event-dependent SCCS models (Multimedia Appendix 6).
Discussion

Principal Findings

These SCCS analyses did not find evidence that the risk of ischemic stroke was elevated during the 1–21–day postvaccination risk interval in both overall and subgroup analyses by age (younger than 65 years vs 65 years or older), prior history of SARS-CoV-2 infection, and coadministration of influenza vaccine, for both Pfizer-BioNTech and Moderna bivalent vaccines. However, based on electronically identified events, the risk of ischemic stroke was increased within the 1–42–day window after vaccination among those aged younger than 65 years who received their Pfizer-BioNTech bivalent vaccine and influenza vaccine on the same day; the risk was even higher among those who also had a documented SARS-CoV-2 infection history. After conducting a chart review of ischemic stroke events, the point estimate for the risk of ischemic stroke was still elevated in a risk interval of 1–42 days for these 2 subgroup analyses but did not meet the threshold for statistical significance (P=0.06).

For Moderna bivalent vaccination, an initial increase in the risk of ischemic stroke emerged within the 1–42–day window after vaccination among those aged younger than 65 years who had a documented SARS-CoV-2 infection history. However, after conducting a chart review of ischemic stroke events, the RI was 2.24 but was no longer statistically significantly elevated possibly due to a decreased sample size (P=0.14).

Our study showed an increased point estimate for the risk of ischemic stroke in a risk interval of 1–42 days only among those aged younger than 65 years who received their Pfizer-BioNTech bivalent vaccine and influenza vaccine on the same day, although not statistically significant. This finding is unique and may be attributed to differences in the study design compared to previous studies. First, our study used a calendar-based observation period spanning from September 1, 2022, to March 31, 2023. This extended timeframe enabled us to use a longer risk window of 1–42 days following vaccination in addition to the risk interval of 1–21 days in previous studies. Second, we did not exclude individuals with a history of ischemic stroke, but we did apply criteria to increase the likelihood that ischemic stroke events during the study period represented a new ischemic stroke episode. Nevertheless, it is possible that there was interaction between bivalent vaccination and a history of ischemic stroke. Furthermore, in subgroup analyses, we considered the influence of the history of SARS-CoV-2 infection. SARS-CoV-2 infection is associated with an increased risk of ischemic stroke [28,29] and risk factors for SARS-CoV-2 infection may overlap with risk factors for ischemic stroke. There is a potential interaction between bivalent vaccination and a history of SARS-CoV-2 infection. The finding that the point estimate for the risk of ischemic stroke was elevated among individuals aged younger than 65 years but not among individuals aged 65 years or older is also biologically plausible. This may be due to the relatively heightened immune response and subsequent inflammation in the younger age group versus the older age group, and the fact that inflammation has been shown to be associated with an increased risk of ischemic stroke [30,31]. Moreover, a smaller proportion of younger adults opted for bivalent vaccination [32], and those who did might have had a higher prevalence of comorbidities or poorer overall health status.

Limitations and Strengths

This study had several limitations. First, the study took place in a single health care system. Additionally, the number of ischemic stroke events in individuals with a documented SARS-CoV-2 infection history who received coadministration of Pfizer-BioNTech bivalent vaccine and influenza vaccine was very small. This raises concerns about the validity of the asymptotic large sample assumptions that underlie both the 95% CIs and P values. Second, we did not conduct a chart review of ischemic stroke events among NBR; these events contributed to establishing baseline rates of ischemic stroke events during the study period. In addressing this issue, when analyzing chart-confirmed ischemic stroke events among recipients of bivalent vaccine, we applied the confirmation rate of ischemic stroke events among recipients of bivalent vaccine to ischemic stroke events among NBR. Moreover, we also did not undertake a chart review of ischemic stroke events from those analyses when safety signals were absent. Third, while we excluded ischemic stroke events occurring within 30 days of SARS-CoV-2 infection, it is possible that some ischemic stroke events included in the analyses involved individuals with asymptomatic or mild COVID-19 disease who did not have a documented SARS-CoV-2 infection. Fourth, the elevated point estimate for the risk of ischemic stroke, while not statistically significant, was observed within the 1–42–day period following the coadministration of the Pfizer-BioNTech bivalent vaccine and influenza vaccine. This risk interval was longer than the 1–21 days or 1–28 days investigated in earlier research [10,13,33]. However, the biological plausibility for the occurrence of a vaccine-related ischemic stroke beyond 28 days remains uncertain. Fifth, unaccounted time-varying confounders could have also influenced the findings. Finally, our analysis did not adjust for multiple subgroup analyses by age, coadministration of bivalent COVID-19 vaccine and influenza vaccine, and history of SARS-CoV-2 infection. These specific subgroup analyses were prespecified due to their potential safety concerns. The decision not to make multiple comparison adjustments was deliberate, aimed at ensuring that any potential vaccine safety concern could be detected.

The study also had several strengths. First, we addressed the impact of previous ischemic stroke events on bivalent COVID-19 vaccination by using an event-dependent modified SCCS design. Second, we explored effect heterogeneity by conducting subgroup analyses based on factors such as age, documented history of SARS-CoV-2 infection, and coadministration of influenza vaccine. Third, to adjust for temporal trends, we included ischemic stroke events among NBR. This strategy not only enhanced the accuracy of estimating the baseline rate but also improved the statistical power for identifying potential safety signals. Finally, we reanalyzed ischemic stroke events that were confirmed through chart review for those analyses where safety signals were detected.
Future research should include several key aspects to further enhance the validity and robustness of our findings. Collaborative efforts with additional health care systems will enable us to significantly increase our sample size. A larger sample size could provide sufficient statistical power to conduct sensitivity analyses such as the exclusion of transient ischemic attacks and exclusion of those who had a history of ischemic stroke.

In light of the findings of this study on the risk of ischemic stroke after bivalent COVID-19 vaccination, it is necessary to assess the safety of the monovalent COVID-19 vaccination during 2023-2024 for several reasons. First, while the bivalent COVID-19 vaccines included 2 components (BA.4 and BA.5), the monovalent COVID-19 vaccines during the 2023-2024 season included 1 component (XBB.1.5). This change in vaccine composition warrants continued surveillance to assess any differential safety profiles. Second, during the 2023-2024 season, coadministration of monovalent COVID-19 vaccine and influenza vaccine on the same day was possible given the timing of availability of both products and the recommendation by the CDC.

Conclusions
We found no evidence to suggest that the Pfizer-BioNTech bivalent vaccine increased the risk of ischemic stroke among individuals aged older than 65 years, consistent with the attenuated signal from the VSD surveillance that motivated this study. We found an elevated point estimate for the risk of ischemic stroke within 1-42 days (but not within 1-21 days) after the coadministration of the Pfizer-BioNTech bivalent vaccine and influenza vaccine among individuals younger than 65 years old that did not reach statistical significance, although the sample size was limited. Future studies with a larger sample size are needed to evaluate the association between bivalent COVID-19 vaccination and ischemic stroke, as well as contributing factors such as the history of SARS-CoV-2 infection. Any potential risks of ischemic stroke associated with bivalent COVID-19 vaccination must be balanced against the potential benefits of bivalent COVID-19 vaccination in preventing COVID-19–associated ischemic stroke and severe COVID-19 disease.

Acknowledgments
Funding for this study was provided by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (R01 AI168209). The National Institutes of Health had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the paper; and decision to submit the paper for publication. The findings and conclusions in this report are those of the authors and do not necessarily represent the official views of the National Institutes of Health. Kaiser Permanente Southern California stroke chart abstractions from the Vaccine Safety Datalink were used in this work. Kristin Goddard and Pat Ross from Kaiser Permanente Northern California and Kayla Hanson from Marshfield Clinic Research Institute developed the stroke chart abstraction forms. The authors thank the Kaiser Permanente Southern California chart abstractors Jose Pio and Julliane Bacerdo.

Data Availability
The data sets generated and analyzed during this study are not publicly available due to potentially identifying or sensitive patient information but are available from the corresponding author on reasonable request.

Conflicts of Interest
LSS reports research support from Moderna for a COVID-19 vaccine effectiveness study and GlaxoSmithKline, Dynavax, and Moderna for unrelated studies. LQ reports research support from Moderna, GlaxoSmithKline, and Dynavax for unrelated studies. KJB reports research support from Moderna, Pfizer, GlaxoSmithKline, and Dynavax for unrelated studies. NPK reports research support from Pfizer for COVID-19 vaccine clinical trials and unrelated research support from Pfizer, GlaxoSmithKline, Merck, and Sanofi Pasteur.

Multimedia Appendix 1
Definition of electronically identified ischemic strokes.
[DOCX File, 56 KB - publichealth_v10i1e53807_app1.docx ]

Multimedia Appendix 2
Descriptive statistics of Charlson Comorbidity Index by vaccine type and age (<65 years old and ≥65 years old) among members of Kaiser Permanente Southern California who had ischemic stroke events during the period from September 1, 2022, to March 31, 2023.
[DOCX File, 47 KB - publichealth_v10i1e53807_app2.docx ]

Multimedia Appendix 3
Numbers of electronically identified ischemic stroke events and relative incidences in the 21 days after Pfizer-BioNTech bivalent COVID-19 vaccination among members of Kaiser Permanente Southern California during the period from September 1, 2022, to March 31, 2023.

[DOCX File, 51 KB - publichealth_v10i1e53807_app3.docx]

Multimedia Appendix 4
Numbers of electronically identified ischemic stroke events and relative incidences in the 21 days after Moderna bivalent COVID-19 vaccination among members of Kaiser Permanente Southern California during the period from September 1, 2022, to March 31, 2023.

[DOCX File, 50 KB - publichealth_v10i1e53807_app4.docx]

Multimedia Appendix 5
Stratified analyses by sex for those analyses that signaled using electronically identified ischemic stroke events: numbers of confirmed ischemic stroke events and relative incidences in the 42 days after bivalent vaccination among members of Kaiser Permanente Southern California aged <65 years during the period from September 1, 2022, to March 31, 2023.

[DOCX File, 51 KB - publichealth_v10i1e53807_app5.docx]

Multimedia Appendix 6
SAS and R codes for preparing data and fitting the dependent SCCS models.

[DOCX File, 51 KB - publichealth_v10i1e53807_app6.docx]

References


Abbreviations

AR: attributable risk

CDC: US Centers for Disease Control and Prevention

FDA: US Food and Drug Administration

ICD-10: International Classification of Diseases, Tenth Re却

KPC: Kaiser Permanente Southern California

mRNA: messenger ribonucleic acid
NBR: nonbivalent recipient
NNH: number needed to harm
RI: relative incidence
SCCS: self-controlled case series
VAERS: Vaccine Adverse Event Reporting System
VSD: Vaccine Safety Datalink

Edited by A Mavragani, T Sanchez; submitted 19.10.23; peer-reviewed by M Jabagi, T Tong, T Cerqueira-Silva; comments to author 09.04.24; accepted 21.05.24; published 25.06.24.

Please cite as:
Ischemic Stroke After Bivalent COVID-19 Vaccination: Self-Controlled Case Series Study
JMIR Public Health Surveill 2024;10:e53807
URL: https://publichealth.jmir.org/2024/1/e53807
doi:10.2196/53807
PMID:38916940

©Stanley Xu, Lina S Sy, Vennis Hong, Kimberly J Holmquist, Lei Qian, Paddy Farrington, Katia J Bruxvoort, Nicola P Klein, Bruce Fireman, Bing Han, Bruno J Lewin. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 25.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Adverse Events of COVID-19 Vaccines in the United States: Temporal and Spatial Analysis

Yiming Li¹, MS; Jianfu Li², PhD; Yifang Dang¹, MS; Yong Chen³, PhD; Cui Tao², PhD

¹McWilliams School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, United States
²Department of Artificial Intelligence and Informatics, Mayo Clinic, Jacksonville, FL, United States
³Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Cui Tao, PhD
Department of Artificial Intelligence and Informatics
Mayo Clinic
4500 San Pablo Road
Jacksonville, FL, 32224
United States
Phone: 1 904 953 0255
Email: Tao.Cui@mayo.edu

Abstract

Background: The COVID-19 pandemic, caused by SARS-CoV-2, has had a profound impact worldwide, leading to widespread morbidity and mortality. Vaccination against COVID-19 is a critical tool in controlling the spread of the virus and reducing the severity of the disease. However, the rapid development and deployment of COVID-19 vaccines have raised concerns about potential adverse events following immunization (AEFIs). Understanding the temporal and spatial patterns of these AEFIs is crucial for an effective public health response and vaccine safety monitoring.

Objective: This study aimed to analyze the temporal and spatial characteristics of AEFIs associated with COVID-19 vaccines in the United States reported to the Vaccine Adverse Event Reporting System (VAERS), thereby providing insights into the patterns and distributions of the AEFIs, the safety profile of COVID-19 vaccines, and potential risk factors associated with the AEFIs.

Methods: We conducted a retrospective analysis of administration data from the Centers for Disease Control and Prevention (n=663,822,575) and reports from the surveillance system VAERS (n=900,522) between 2020 and 2022. To gain a broader understanding of postvaccination AEFIs reported, we categorized them into system organ classes (SOCs) according to the Medical Dictionary for Regulatory Activities. Additionally, we performed temporal analysis to examine the trends of AEFIs in all VAERS reports, those related to Pfizer-BioNTech and Moderna, and the top 10 AEFI trends in serious reports. We also compared the similarity of symptoms across various regions within the United States.

Results: Our findings revealed that the most frequently reported symptoms following COVID-19 vaccination were headache (n=141,186, 15.68%), pyrexia (n=122,120, 13.56%), and fatigue (n=121,910, 13.54%). The most common symptom combination was chills and pyrexia (n=56,954, 6.32%). Initially, general disorders and administration site conditions (SOC 22) were the most prevalent class reported. Moderna exhibited a higher reporting rate of AEFIs compared to Pfizer-BioNTech. Over time, we observed a decreasing reporting rate of AEFIs associated with COVID-19 vaccines. In addition, the overall rates of AEFIs between the Pfizer-BioNTech and Moderna vaccines were comparable. In terms of spatial analysis, the middle and north regions of the United States displayed a higher reporting rate of AEFIs associated with COVID-19 vaccines, while the southeast and south-central regions showed notable similarity in symptoms reported.

Conclusions: This study provides valuable insights into the temporal and spatial patterns of AEFIs associated with COVID-19 vaccines in the United States. The findings underscore the critical need for increasing vaccination coverage, as well as ongoing surveillance and monitoring of AEFIs. Implementing targeted monitoring programs can facilitate the effective and efficient management of AEFIs, enhancing public confidence in future COVID-19 vaccine campaigns.

(JMIR Public Health Surveill 2024;10:e51007) doi:10.2196/51007
KEYWORDS
COVID-19; vaccine; COVID-19 vaccine; adverse drug event; ADE; Vaccine Adverse Event Reporting System; VAERS; adverse event following immunization; AEFI

Introduction
The COVID-19 pandemic, caused by SARS-CoV-2, was first identified in China in 2019 and quickly became an uncontrollable outbreak worldwide [1-4]. As of March 2023, the World Health Organization (WHO) had reported 761,402,282 confirmed cases and 6,887,000 deaths due to COVID-19 worldwide [5]. COVID-19 primarily spreads through respiratory droplets, and infection can bring about mild-to-severe symptoms, ranging from general fatigue, cough, fever, loss of taste and smell, diarrhea, and severe pneumonia to even death [3-7]. Elderly individuals and those with underlying conditions, such as obesity, diabetes, and hypertension, are at higher risk [8]. Long-term effects of COVID-19 include fatigue, muscle weakness, sleep difficulties, anxiety, and depression [9]. So far, vaccines are considered the primary method to control the virus, with over 674,375,000 doses administered in the United States by April 2023 [10-14]. Coccia [15] demonstrated that nations enforcing stringent societal restrictions and obligations achieved a high rate of full COVID-19 vaccination, reaching 77.17% (average stringency index of 62.97) by February 2022. Magazzino et al [16] and Aldila et al [17] maintained that achieving higher levels of vaccination could lead to the eradication of COVID-19 in the population by the development of herd immunity, thereby protecting vulnerable individuals. Coccia [18] also revealed that administering an average of about 80 doses of vaccines per 100 inhabitants between countries can sustain a reduction in confirmed cases and deaths. The growth of the pandemic wave in May 2021 increased the optimal level of vaccines to about 90 doses for reducing the numbers of COVID-19–related infections [18]. Although a widespread vaccination campaign is essential to fight against infectious diseases, it alone is not sufficient as a public policy to mitigate the adverse effects of the COVID-19 pandemic crisis [13,19]. Cases have shown that COVID-19 vaccines can trigger adverse events in multiple systems, including oral, digestive, hematological, immune, and nervous systems [20-26]. Common side effects include tenderness at the injection site, fever, fatigue, body ache, and headache [27-30]. To make matters worse, there have been reports of serious adverse events following immunization (AEFIs), such as acute kidney injury, respiratory distress syndrome, coagulation disorders, and cardiac injuries, associated with COVID-19 vaccines [31]. Therefore, new vaccination strategies for nations must be highly responsive, flexible, resilient, scalable, and effective in reducing the negative impact of coronavirus [18,32]. Temporal and spatial factors are critical in the spread of COVID-19, as evidenced by recent research. Coccia’s [33] systematic review highlighted that high air and environmental pollution, as well as unsustainable environments, can facilitate the emergence and rapid spread of pandemics. Coccia [33] also emphasized the importance of an effective contact-tracing system and timely isolation in reducing the transmission dynamics of infectious diseases within and between different outbreak areas, particularly for diseases with a latent presymptomatic phase. Moreover, Coccia’s [34] analysis of seasonality in COVID-19 transmission revealed a correlation between lower temperatures and higher transmission rates, especially in colder regions. Additionally, a spatial analysis demonstrated regional disparities in COVID-19 diffusion, with urban areas showing higher transmission rates than rural areas [35]. These findings underscore the significance of considering temporal and spatial factors in comprehending the COVID-19 spread and suggest that AEFIs related to COVID-19 vaccination may also demonstrate temporal and spatial trends, necessitating further exploration.

Temporal monitoring allows for the identification of trends, potential causal relationships, and patterns of AEFIs over time [36]. Neglecting the heterogeneity or temporal trend of reporting rates across different years can result in missing significant signals, especially given the evolving nature of COVID-19 caused by prevailing strains of SARS-CoV-2 [37-39]. Additionally, spatial analysis examines the similarity of AEFIs across different regions, providing insights into spatial variations, vaccine brands, and populations [40]. The majority of AEFIs are preventable [41]; thus, analyzing these AEFIs reported enables public health researchers and officials to understand their spatial patterns, potential causal factors, and overall impact, supporting evidence-based decision-making and targeted interventions. The Vaccine Adverse Event Reporting System (VAERS), a comprehensive database that collects reports of AEFIs across different states and periods in the United States, proves instrumental in conducting temporal and spatial monitoring of COVID-19 vaccine–related AEFIs [42].

Despite significant research on AEFIs associated with COVID-19 vaccines using VAERS data, previous research has primarily focused on short-term data, neglecting comprehensive temporal and spatial analyses within the United States. Huang et al [43] developed a composite likelihood-based variance component model to analyze the temporal variation of AEFIs reporting using VAERS data. The method accounted for underreporting and zero inflation in passive surveillance systems and identified 14 AEFIs with significantly heterogeneous reporting rates over the years, including 2 events showing an increasing trend [43]. Cai et al [37] proposed a random effects model to test the heterogeneity of reporting rates for vaccine-event combinations across multiple years in the VAERS database. The method demonstrated high statistical power in detecting variations in reporting rates, highlighting potential safety issues associated with changes in influenza vaccines [37]. Askar and Zülfü [40] conducted a study on the similarity of adverse effects of COVID-19 vaccines across different states in the United States using VAERS data. They applied a topic modeling approach to extract latent topics from the AEFIs reported and identified spatial clusters of states exhibiting similar AEFIs [40]. These findings underscore the variation in AEFIs across states and emphasize the importance of further research to understand underlying causes, enhance the
comprehension of adverse effects, and address vaccine hesitancy [40]. However, the random effects model shows limitations in capturing the complex relationships among symptoms, whereas latent Dirichlet allocation (LDA) may overlook semantic similarity between symptom compositions in different regions, although it is innovative to identify spatial clusters of states with similar AEFIs.

In this study, we aimed to conduct a comprehensive temporal and spatial analysis of AEFIs associated with COVID-19 vaccines reported to VAERS. We used the Medical Dictionary for Regulatory Activities (MedDRA) as our medical terminology reference [23]. The rest of the paper is organized as follows. We begin with descriptive analyses of vaccine administration and VAERS data, examining factors such as gender, age, and manufacturer, with a specific focus on serious reports. Subsequently, we discuss how we investigated the temporal variation of AEFIs and symptoms within different system organ classes (SOCs) throughout the study period, enabling a holistic analysis. By modeling weekly reported symptoms relative to administration, we accurately assessed temporal variation, identifying associations between time and AEFIs. In addition to the temporal analysis, we also conducted a spatial analysis using the BioWordVec_PubMed_MIMICIII embedding model. This approach allowed us to construct meaningful vectors that capture the nuances of symptom compositions, enabling us to examine the reporting rates of AEFIs in different regions across the United States. Overall, our approach combines advanced embeddings, semantic similarity, and temporal modeling, providing comprehensive insights into AEFIs reported for COVID-19 vaccines.

Methods

Ethical Considerations

The original data collection was approved by the Institutional Review Board (IRB). The analysis did not receive approval/exemption from the IRB. The secondary analysis did not need a review from the IRB, as we used a publicly available data set [11,44]. The authors had permission to use the data.

Sample and Data

We collected VAERS reports of AEFIs associated with COVID-19 vaccines from December 13, 2020, to December 28, 2022. The reports consist of 3 comma-separated-value (CSV) files grouped by year: VAERSDATA.CSV, VAERSVAX.CSV, and VAERSSYMPTOMS.CSV. VAERSDATA contains demographic information, vaccination and adverse event timing, symptom descriptions, allergy history, and serious outcomes. VAERSVAX provides details on vaccine type and manufacturer for each adverse event, while VAERSSYMPTOMS lists symptoms associated with each adverse event, as mapped from the preferred term (PT) in the MedDRA terminology. The 3 tables are linked by the primary key VAERS_ID.

In addition, we curated COVID-19 vaccine administration data from the Centers for Disease Control and Prevention’s (CDC) COVID Data Tracker during the corresponding period. The COVID Data Tracker is a centralized database maintained by the CDC that provides up-to-date information about COVID-19 vaccine administration across the United States [11]. The COVID Data Tracker contains data on the number of vaccine doses distributed and administered, as well as breakdowns by state, demographic group, and vaccine type [11].

MedDRA is a standardized vocabulary for adverse event reporting, allowing for the consistent classification and analysis of adverse events across different pharmaceutical products and clinical studies [45]. The MedDRA terminology comprises a structural hierarchy of 5 levels: the SOC, high-level group term (HLGT), high-level term (HLT), PT, and lowest-level term (LLT). According to the MedDRA website, the current version of MedDRA (version 24.1 as of September 2021) contains over 84,000 PTs, which is used by the VAERS code to classify adverse events reported to the system [46].

Measures of Variables

We performed several key analyses to examine the characteristics of AEFIs associated with COVID-19 vaccines. An AEFI is defined as any untoward occurrence following immunization [47]. In this study, we focused solely on AEFIs reported to VAERS. We did not investigate whether these AEFIs were caused by other factors occurring during the same study period (eg, COVID-19 infection, symptoms arising from other diseases or interventions). CDC data on COVID-19 vaccine administration were used to determine the number of vaccinations. This information provides a context and allows for comparison with the number of VAERS reports. We also summarized the number of VAERS reports and the symptoms reported, including unique symptoms. The number of COVID-19 vaccines administered and the AEFIs reported were stratified by sex, age, and vaccine manufacturer. Additionally, we analyzed the occurrence of individual symptoms and the co-occurrence of symptom pairs.

We categorized the VAERS reports based on sex, age, and vaccine manufacturer in the following ways. Sex was classified into male, female, and unknown. The reports were divided into 6 age groups based on CDC-recommended cutoff thresholds (5, 12, 18, 65 years), and “unknown” was used for age data that were unavailable. The main COVID-19 vaccine manufacturers in the market included Pfizer-BioNTech, Moderna, Janssen, Novavax, and unknown. We filtered out vaccinations that included a mixture of Pfizer-BioNTech and Moderna, because it was not possible to determine from the reported data whether the AEFIs resulted from Pfizer-BioNTech or Moderna vaccination. We also excluded 3 subjects who did not report any symptoms, resulting in a sample size of 5493 subjects in total.

A report was classified as serious if it contained any of the following outcomes: death; a threat to life at the time of the event; emergency room visit; inpatient hospitalization or prolongation of existing hospitalization; or persistent or significant disability/incapacity, a congenital anomaly/birth defect, or a medically important event based on medical judgment [31], with the corresponding fields indicating “DIED” (death), “L_THREAT” (life threatening), “ER_VISIT” (emergency room visit), “X_STAY” (hospitalization or extended hospital stay), or “DISABLE” (persistent or significant disability/incapacity). To gain a deeper understanding of AEFIs
associated with COVID-19 vaccines in serious reports, we conducted additional analyses of these reports, focusing on the composition of cases by sex, age group, and vaccine manufacturer.

**Data Analysis**

To categorize symptoms reported, we mapped them to the SOC level. The SOC is the top-level hierarchical structure used for broad categorization of medical concepts based on etiology, manifestation site, or purpose [48]. In our study, we suggested a straightforward approach for categorizing AEFIs into SOCs. This method involves using the internationally agreed-upon order of SOCs (refer to Table S1 in Multimedia Appendix 1), which is determined by the relative importance of each SOC [49]. In VAERS, each symptom reported could be mapped to either 1 LLT or 1 PT in MedDRA. First, we matched each symptom to the corresponding LLT or PT, resulting in 66 symptoms matched to LLTs and 12,050 symptoms matched to PTs. Symptoms matched to LLTs were further mapped to the corresponding PTs, and relevant SOC terms were identified. If a report contained multiple symptoms that fell under a specific SOC, we recorded the occurrence of that SOC for each corresponding symptom in the analysis.

In our study, we conducted a thorough temporal analysis to monitor the reporting rate of AEFIs (Equation 1) associated with COVID-19 vaccines and individual SOCs on a weekly basis. This approach allowed us to closely track any temporal trends and evaluate the potential risks associated with COVID-19 vaccination, providing valuable insights for well-informed decisions in public health policy.

Furthermore, we ranked and analyzed the top 10 symptoms reported in serious case reports over time. By examining the reporting rate of these symptoms, we gained a comprehensive understanding of their prevalence and impact. This analysis provided valuable insights into the severity and frequency of AEFIs associated with COVID-19 vaccines.

Additionally, we conducted a separate temporal analysis specifically focusing on AEFIs related to Pfizer-BioNTech and Moderna vaccines. By analyzing these vaccines individually, we were able to gain a more in-depth understanding of the AEFIs reported and identify any unique patterns or differences between vaccine manufacturers.

We conducted a spatial analysis of the reporting rates of cases and serious cases for each state in the United States. To organize the states into standardized regions, we used the Standard Federal Regions, as defined in “Circular A-105” released by the Office of Management and Budget in April 1974 (refer to Table S2 in Multimedia Appendix 1) [50].

Furthermore, we performed statistical analysis on the distribution of the AEFIs within each region and generated vectors that considered both AEFIs and their frequencies. To assess the similarity of the AEFI composition across different regions, we used the BioWordVec_PubMed_MIMICIII embedding model. This model is based on the word2vec algorithm and is specifically designed to enhance biomedical word embeddings. It leverages subword information and incorporates the vast PubMed and MIMIC-III data sets to create embeddings that capture the rich semantics of biomedical terms [51,52]. By using the BioWordVec_PubMed_MIMICIII embedding model, we measured the semantic similarity between AEFI compositions in different regions. This analysis provided insights into the shared patterns and characteristics of AEFIs across various geographic areas.

**Results**

**Descriptive Statistics for COVID-19 Vaccination**

Tables 1 and 2 present descriptive statistics for COVID-19 vaccination, with data divided into administration data provided by the CDC and AEFIs reported in VAERS, respectively. Vaccination rates peaked in 2021 in both administration and VAERS data. Notably, VAERS reports for the age groups of 0-5 years (excluding 5) and 5-12 years (including 5 but excluding 12) had the highest number of AEFIs, followed by elderly individuals (aged ≥65 years). The Pfizer-BioNTech vaccine was the most commonly administered one during the study period, but in 2021, there were more VAERS reports associated with the Moderna vaccine.

Figure S1 in Multimedia Appendix 1 presents a comprehensive breakdown of serious reports analyzed in this study. The total number of reports included was 900,522, with 42,366 (4.7%) classified as serious. Of these serious reports, 21,153 (49.93%) were reported by female patients and 20,275 (47.86%) by male patients. The highest proportion of serious VAERS reports was in the age group of 18-65 years, accounting for 48.82% (n=20,684) of the total serious reports, followed by individuals aged ≥65 years, who submitted 18,681 (44.09%) of the AEFI reports. Regarding vaccine manufacturers, Pfizer-BioNTech had the highest number of VAERS reports, with 20,623 (48.68%) cases, followed by Moderna with 16,936 (39.98%) cases. Serious VAERS reports resulting from Janssen, Novavax, and unknown manufacturers constituted 11.34% (n=4807) of the total reports.

In the serious reports, a total of 314,777 nonunique AEFIs and 7945 unique AEFIs were identified. The most frequent AEFIs reported were death (n=13,323, 31.45%), COVID-19 (n=6431, 15.18%), dyspnea (n=6199, 14.63%), SARS-CoV-2 test positivity (n=5036, 11.89%), and fatigue (n=3610, 8.52%). The most frequent SOCs reported were investigations (n=93,716, 22.11%); general disorders and administration site conditions (n=48,164, 11.29%); nervous system disorders (n=31,689, 74.8%); respiratory, thoracic, and mediastinal disorders (n=22,997, 54.28%); and surgical and medical procedures (n=14,982, 35.36%).
Table 1. COVID-19 vaccine administration data according to the CDC\(^a\) in the United States (2020-2022).

<table>
<thead>
<tr>
<th>Administration data</th>
<th>Year</th>
<th>2020 (n=3,738,130), n (%)</th>
<th>2021 (n=505,569,659(^b)), n (%)</th>
<th>2022 (n=154,514,786), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-12</td>
<td></td>
<td>11,230,026 (2.22)</td>
<td>11,939,125 (7.73)</td>
<td></td>
</tr>
<tr>
<td>12-18</td>
<td></td>
<td>30,175,230 (5.97)</td>
<td>9,754,269 (6.31)</td>
<td></td>
</tr>
<tr>
<td>18-65</td>
<td></td>
<td>336,002,922 (66.46)</td>
<td>80,914,834 (52.37)</td>
<td></td>
</tr>
<tr>
<td>≥65</td>
<td></td>
<td>131,850,714 (26.08)</td>
<td>48,311,353 (31.27)</td>
<td></td>
</tr>
<tr>
<td>0-5+unknown</td>
<td></td>
<td></td>
<td>3,595,205 (2.33)</td>
<td></td>
</tr>
<tr>
<td>Manufacturer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pfizer-BioNTech</td>
<td>2,630,115 (70.36)</td>
<td>294,240,716 (58.20)</td>
<td>98,581,346 (63.80)</td>
<td></td>
</tr>
<tr>
<td>Moderna</td>
<td>1,107,143 (29.62)</td>
<td>193,153,251 (38.21)</td>
<td>54,306,971 (35.15)</td>
<td></td>
</tr>
<tr>
<td>Janssen</td>
<td>0</td>
<td>17,640,334 (3.49)</td>
<td>1,312,374 (0.85)</td>
<td></td>
</tr>
<tr>
<td>Novavax</td>
<td>0</td>
<td>0</td>
<td>69,062 (0.04)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>872 (0.02)</td>
<td>535,358 (0.11)</td>
<td>245,033 (0.16)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)CDC: Centers for Disease Control and Prevention.
\(^b\)The data do not align with the total number of vaccines administered by age group in 2021.
\(^c\)Data not available.

Table 2. COVID-19 vaccine data according to VAERS\(^a\) reports following COVID-19 vaccination in the United States (2020-2022).

<table>
<thead>
<tr>
<th>VAERS data</th>
<th>Year</th>
<th>2020 (n=10,380), n (%)</th>
<th>2021 (n=698,505), n (%)</th>
<th>2022 (n=191,637), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1942 (18.71)</td>
<td>207,399 (29.69)</td>
<td>70,916 (37.01)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8266 (79.63)</td>
<td>465,475 (66.64)</td>
<td>109,583 (57.18)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>172 (1.66)</td>
<td>25,631 (3.67)</td>
<td>11,138 (5.81)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-5</td>
<td>5 (0.05)</td>
<td>320 (0.05)</td>
<td>2676 (1.40)</td>
<td></td>
</tr>
<tr>
<td>5-12</td>
<td>0</td>
<td>6002 (0.86)</td>
<td>9245 (4.82)</td>
<td></td>
</tr>
<tr>
<td>12-18</td>
<td>25 (0.24)</td>
<td>25,737 (3.68)</td>
<td>8117 (4.24)</td>
<td></td>
</tr>
<tr>
<td>18-65</td>
<td>9383 (90.39)</td>
<td>441,617 (63.22)</td>
<td>93,105 (48.58)</td>
<td></td>
</tr>
<tr>
<td>≥65</td>
<td>514 (4.95)</td>
<td>157,014 (22.48)</td>
<td>54,846 (28.62)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>453 (4.36)</td>
<td>67,815 (9.71)</td>
<td>23,648 (12.34)</td>
<td></td>
</tr>
<tr>
<td>Manufacturer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pfizer-BioNTech</td>
<td>7328 (70.6)</td>
<td>308,256 (44.13)</td>
<td>99,381 (51.86)</td>
<td></td>
</tr>
<tr>
<td>Moderna</td>
<td>3029 (29.18)</td>
<td>326,157 (46.69)</td>
<td>82,727 (43.17)</td>
<td></td>
</tr>
<tr>
<td>Janssen</td>
<td>0</td>
<td>62,570 (8.96)</td>
<td>8635 (4.51)</td>
<td></td>
</tr>
<tr>
<td>Novavax</td>
<td>0</td>
<td>0</td>
<td>199 (0.1)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>23 (0.22)</td>
<td>1522 (0.22)</td>
<td>695 (0.36)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)VAERS: Vaccine Adverse Event Reporting System.

Next, we computed the number of occurrences of individual symptoms and symptom co-occurrences for all VAERS reports during the period of 2020-2022. The most frequently reported symptoms following COVID-19 vaccination were headache (n=141,186, 15.68%), pyrexia (n=122,120, 13.56%), and fatigue (n=121,910, 13.54%). The results are shown in Figure S2 in Multimedia Appendix 1 (detailed results are available in “Frequency and reporting rate of adverse events” in Multimedia Appendix 2). The most frequent co-occurrence pair was chills+pyrexia (n=56,954, 6.32%).
In “Covid-19_vaccine_2020_2022” in Multimedia Appendix 2, we can find the top 5 symptoms categorized by gender, vaccine manufacturer, and age group for the years 2020-2022. Furthermore, Multimedia Appendix 2 also provides the same information for each specific year (2020, 2021, and 2022). For the Janssen vaccine, the top 5 symptoms reported were headache, pyrexia, chills, fatigue, and pain. The most common symptoms and AEFIs for the Novavax vaccine were dizziness, followed by headache, fatigue, incorrect product formulation, and pain. Those receiving an unknown vaccine reported COVID-19, headache, pyrexia, pain, and chills. For Moderna and Pfizer-BioNTech, the most common symptoms were headache, pyrexia, fatigue, pain, and chills. Adults aged 18-65 years reported these same symptoms, while the elderly also reported SARS-CoV-2 test positivity. Teenagers reported product errors, whereas infants and toddlers reported fever and dosage issues. Children aged 5-12 years reported no adverse events and product errors.

**Temporal Analysis**

Figure S3 in Multimedia Appendix 1 shows the serious case–reporting rate associated with each COVID-19 vaccine manufacturer between 2020 and 2022. Among the vaccine manufacturers, Janssen had the highest reporting rate, followed by the unknown manufacturer, Novavax, Moderna, and Pfizer-BioNTech in that order. Notably, in 2022, the reporting rate for Janssen (0.0009) was significantly higher compared to any other COVID-19 vaccine manufacturer from 2020 to 2022.

Figure S4 in Multimedia Appendix 1 provides insight into the reporting rate of serious cases associated with COVID-19 vaccines across different age groups between 2021 and 2022. The reporting rate was higher in older-age groups and increased with age. Additionally, the reporting rate for each age group peaked in 2022.

Figure 1 illustrates trends in the proportion of reported SOCs and AEFIs reported in VAERS associated with COVID-19 vaccines during the study period using weekly reported data. The vertical red bars represent the proportion of VAERS reports among weekly administrations, while the lines represent the occurrences of AEFIs in SOCs out of the corresponding weekly administrations. Notably, the reporting symptoms of SOC 22 (“General disorders and administration site conditions”) were more prevalent compared to other symptoms. It is important to note that the figure may not reflect the actual rate of AEFIs and corresponding SOCs due to reporting bias. However, it still provides valuable temporal insights into the development of AEFIs related to COVID-19 vaccination. Interestingly, the local maxima observed in both the bar graph and the line graph align with the 2 peaks of the pandemic in the summer of 2021 and 2022. This suggests a potential correlation between the prevalence of AEFIs and the intensity of the pandemic. Furthermore, the reporting rate of VAERS and SOC symptoms decreased over time, demonstrating that COVID-19–related AEFIs gradually improved.

Figure 2 presents trends in the serious case–reporting rate proportion and the top 10 AEFIs reported related to serious reports in VAERS. The vertical red bars represent the proportion of serious VAERS reports among weekly administrations, while the lines represent the rate of the top 10 AEFIs in serious reports out of the corresponding weekly administrations. Our analysis indicated that the top 10 AEFIs reported in the serious cases included death, COVID-19, dyspnea, SARS-CoV-2 test positivity, fatigue, pyrexia, pain, headache, blood test, and asthenia. Overall, a decreasing trend was observed, particularly after the initial period from December 2020 to April 2021. Notably, 2 points of local peaks occurred in October 2021 and August 2022. However, starting from December 2021, the reporting rate for the other AEFIs decreased significantly over time, while the reports of death remained relatively consistent. Please note that the information provided is based on the analysis available until December 2022, and subsequent updates to the data may reveal different trends or findings.

Figure 3 illustrates trends in the case-reporting rate and the individual SOC-reporting rate, specifically for Pfizer-BioNTech. The vertical red bars represent the proportion of VAERS reports associated with Pfizer-BioNTech immunization among its weekly administrations, while the lines represent the occurrences of AEFIs in SOCs out of the corresponding weekly Pfizer-BioNTech administrations. In general, both the case-reporting rate and the individual SOC-reporting rate demonstrated a decreasing trend. However, there were instances in September 2021, March 2022, July 2022, and September 2022 when the case-reporting rate and all SOC-reporting rates reached a local maximum. Before June 2022, the reporting rate for SOC 22 (“General disorders and administration site conditions”) was the highest among all SOC-reporting rates. However, it was later surpassed by SOC 24 (“Injury, poisoning, and procedural complications”).

Figure 4 displays the case-reporting rate and the individual SOC-reporting rate, specifically for Moderna. The vertical red bars represent the proportion of VAERS reports associated with Moderna immunization among its weekly administrations, while the lines represent the occurrences of AEFIs in SOCs out of the corresponding weekly Moderna administrations. Initially, both the case-reporting rate and the individual SOC-reporting rate for Moderna were relatively high. Although there was a sharp decrease until June 2021, the rate experienced a rebound and reached a local peak in August 2021. Subsequently, there was a gradual decreasing trend observed. Prior to November 2021, the reporting rate for SOC 22 (“General disorders and administration site conditions”) far exceeded that of the other categories. However, from that point onward, it became comparable to that of SOC 24 (“Injury, poisoning, and procedural complications”). In comparison to Figure 3, the peaks representing Moderna in the periods of the SARS-CoV-2 Delta and SARS-CoV-2 Omicron variants were higher than those representing Pfizer-BioNTech.
**Figure 1.** Case-reporting rate and individual SOC-reporting rate associated with COVID-19 vaccines in the United States (2020-2022). SOC: system organ class.

**Figure 2.** Reporting rates for serious cases and top 10 adverse events in serious cases for COVID-19 vaccines in the United States (2020-2022).
Figure 3. Case-reporting rate and individual SOC-reporting rate for the Pfizer-BioNTech vaccine in the United States (2020-2022). SOC: system organ class.

Figure 4. Case-reporting rate and individual SOC-reporting rate for the Moderna vaccine in the United States (2020-2022). SOC: system organ class.

Spatial Analysis

Figure 5 presents the VAERS reporting rate by state. It shows that Montana, Minnesota, Michigan, Colorado, Indiana, Alaska, and Kentucky were the states with reporting rates exceeding 0.0015. Among these, Indiana recorded the highest reporting rate, at 0.0025.

Figure 6 shows the VAERS serious case–reporting rate. Montana, South Dakota, Kentucky, and Tennessee were the states with serious case–reporting rates surpassing 100 μ. Among them, South Dakota had the highest reporting rate, reaching...
205.26 μ. In addition to the health care quality and higher confirmed cases relative to the population (especially for Tennessee and Kentucky), this can be partially attributed to their lower vaccination levels (see "geography_serious_reporting_rate_by_state_vaccination" in Multimedia Appendix 2) compared to other states [53,54]. Specifically, these 3 states ranked among states with the least vaccination levels, with all 3 falling far below the average vaccination level nationwide. This lower vaccination rate may contribute to a higher proportion of AEFIs being reported in these states, as individuals who chose to get vaccinated might be more likely to report any side effects they experienced.

Figure 7 presents a heatmap showing the similarity of symptoms between regions within the United States. The similarity between all regions exceeded 0.99, indicating a high level of similarity in symptoms. Notably, regions IV and VI demonstrated the highest similarity, reaching an impressive value of 0.9999.

Figure 8 illustrates a heatmap depicting the similarity of symptoms in serious reports among different regions within the United States. The analysis revealed that the similarity between all regions exceeded 0.99, indicating a remarkable degree of similarity in the symptoms reported. Notably, regions II and III demonstrated the highest level of similarity, with an exceptional value of 0.9998, suggesting a strong correlation in the symptoms reported between these regions. In a broader sense, based on the dark boxes shown in Figure 8, it appears that regions I, II, III, IX, and X were clustered together with higher similarity within this group, whereas regions IV, V, VI, VII, and VIII indicated another clustering pattern.
**Discussion**

**Principal Findings**

Our research has produced findings that underscore the clinical significance of our study (summarized in Table 3). Headache emerged as the predominantly reported symptom associated with the vaccines made by Pfizer-BioNTech, Moderna, and Janssen. Notably, the top 5 symptoms reported for all 3 vaccines are headache, fatigue, pyrexia, pain, and chills, with only slight differences in the order of rankings. The occurrence of headache can be attributed to the body’s immune response to the vaccine, which triggers the generation of humoral and cellular immunity through a diverse range of mechanisms [55]. Some of these mechanisms may lead to inflammation and a subsequent headache [55].
Table 3. Main results in this study and contributions compared to other studies.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Main results</th>
<th>Contributions compared to other studies</th>
</tr>
</thead>
</table>
| COVID-19 administration data    | • Adults aged 18-65 years were vaccinated the most.  
• Pfizer-BioNTech was the most administered vaccine. | • Provides insights into the timing of peak vaccination rates, aiding in understanding vaccination trends during the study period  
• Highlights the demographic group that received the highest vaccination coverage, aiding in targeting future vaccination campaigns |
| VAERS\(^a\) reports            | • VAERS reports indicated higher reporting rates among adults aged 18-65 years.  
• Reporting rates for AEFIs\(^b\) associated with COVID-19 vaccines were higher in female and older-age groups.  
• Pfizer-BioNTech had the highest number of VAERS reports, followed by Moderna.  
• The reporting rate for Janssen was significantly higher in 2022 compared to other vaccine manufacturers.  
• The most frequently reported AEFIs were headache, pyrexia, and fatigue.  
• The most frequent co-occurrence pair was chills+pyrexia.  
• The most frequent SOCs in serious reports were investigations (SOC 23); general disorders and administration site conditions (SOC 22); nervous system disorders (SOC 8); respiratory, thoracic, and mediastinal disorders (SOC 13); and surgical and medical procedures (SOC 25). | • Provides age- and sex-specific reporting rates, highlighting vulnerable populations and aiding in targeted interventions and vaccine recommendations  
• Features differences in reporting rates among different age/sex groups and vaccine manufacturers, contributing to a better understanding of AEFI patterns  
• Identifies key symptoms and SOCs associated with COVID-19 vaccination, offering insights into common AEFIs and areas of focus for vaccine epidemiological surveillance  
• Highlights common symptom patterns postvaccination, aiding in the recognition of symptom clusters and potential treatment strategies |
| Temporal analysis               | • The trend in the proportion of SOCs and AEFIs reported showed a decrease over time.  
• The reporting rate for the top 10 symptoms and AEFIs reported related to serious reports showed an overall decreasing trend.  
• The AEFI-reporting rate and the individual SOC-reporting rate for Pfizer-BioNTech and Moderna showed a decreasing trend over time.  
• For Pfizer-BioNTech, SOC 22 (“General disorders and administration site conditions”) had the highest reporting rate among all SOCs before being surpassed by SOC 24 (“Injury, poisoning, and procedural complications”) starting June 2022.  
• For Moderna, the reporting rate for SOC 22 far exceeded that of the other categories initially, and it became comparable to that of SOC 24 from November 2021 onward. | • Indicates a potential decrease in AEFIs over time, suggesting the effectiveness of monitoring and intervention strategies  
• Highlights manufacturer-specific trends in AEFI reporting, offering valuable insights for vaccine epidemiological surveillance |
| Spatial analysis                | • States with reporting rates exceeding 0.0015 were Montana, Minnesota, Michigan, Colorado, Indiana, Alaska, and Kentucky.  
• States with serious case–reporting rates surpassing 100 \(\mu\) were Montana, South Dakota, Kentucky, and Tennessee.  
• The similarity between all regions exceeded 0.99, indicating a high level of similarity in symptoms between different regions in the United States. | • Identifies geographic variations in AEFI-reporting rates, suggesting the need for targeted vaccine surveillance in high-reporting states  
• Underscores the consistency of AEFI-reporting patterns across different regions, supporting the generalizability of findings and the reliability of VAERS data |

\(^a\)VAERS: Vaccine Adverse Event Reporting System.  
\(^b\)AEFI: adverse event following immunization.  
\(^c\)A report was classified as serious if it contained any of the following outcomes: death; a threat to life at the time of the event; emergency room visit; inpatient hospitalization or prolongation of existing hospitalization; or persistent or significant disability/incapacity, a congenital anomaly/birth defect, or a medically important event based on medical judgment [31].  
\(^d\)SOC: system organ class.

Conversely, the most commonly reported symptom associated with the Novavax vaccine is dizziness. Research suggests that postvaccination vertigo and dizziness are common in patients with Meniere disease (MD) and vertebrobasilar artery insufficiency (VBI) [56]. MD is a disorder with immunological factors that exacerbate endolymphatic hydrops [56,57]. Moreover, heightened osmolality levels in the inner ear can elevate proinflammatory cytokines and immune cell activation.
which may lead to a possible systemic immune response and an increase in disease-specific immunoglobulin G (IgG) levels, thereby intensifying disease activity [56,58,59]. There have been instances where patients with MD who were stable experienced vertigo following vaccination [56,59]. VBI can induce vertigo through dysregulation of blood flow due to altered plasma viscosity, platelet aggregation, red blood cell deformability, and endothelial function [56]. In rare cases, vaccine-related immunization anxiety can trigger vertigo in patients with autoimmune encephalitis [56,60-63].

In addition to common AEFIs, COVID-19 vaccination has been associated with severe or rare AEFIs, including autoimmune encephalitis (AIE), cerebral venous sinus thrombosis, Guillain-Barre syndrome (GBS), optic neuritis, and polyomyositis. These complications, with reporting rates ranging from 1.89e-5 to 0.001, are often of autoimmune nature [23]. Many conditions observed in temporal association with vaccination in this study were previously reported as potential autoimmune sequelae of SARS-CoV-2 infection, sharing similar clinical and laboratory characteristics [64]. Vaccines containing SARS-CoV-2 antigens may enhance autoimmunity through mechanisms such as polyclonal or bystander activation, epitope spreading, or molecular mimicry [64,65]. Alternatively, the inflammatory response induced by vaccination may enhance autoimmunity in predisposed patients, possibly by activating preexisting autoimmune pathways similar to the pathogenesis of immune-related adverse events following administration of immune checkpoint inhibitors [64-67]. Vaccination could also unmask previously asymptomatic autoimmunity in patients with new-onset autoimmune diseases [64]. Recent population-based studies have linked SARS-CoV-2 vaccination to an increased incidence of GBS, especially following Ad26.COV2.S administration [64]. However, the possibility that new onset or flares of other neurological autoimmune conditions merely coincide with vaccination against SARS-CoV-2 cannot be fully excluded [64]. Several pathogenic mechanisms have been proposed to explain how COVID-19 vaccines can lead to AIE, including molecular mimicry, neuroinflammation, and the role of vaccine adjuvants, such as BNT162 adjuvant polyethylene glycol (PEG), which has been implicated in the autoimmune syndrome induced by adjuvants (ASIA-syndrome) [23]. Moreover, vaccine-induced immune thrombotic thrombocytopenia (VITT) (reporting rate: 2.33e-5), although rare, is a consequential complication associated with vaccination [26]. In VITT, the ChAdOx1/PF4 complex may induce the production of anti-PF4 autoantibodies [26]. Trace amounts of ChAdOx1 may enter the bloodstream following intramuscular vaccine administration, due to slight capillary damage, leading to the formation of the ChAdOx1/PF4 complex and triggering the production of autoantibodies [26,68].

Our temporal analysis revealed a declining trend in symptoms across all SOCs, which may help alleviate vaccine hesitancy. The decrease in the reported incidence of AEFIs associated with COVID-19 vaccines can be attributed to various factors, including the rise in vaccination rates, the improvement of vaccine epidemiological surveillance, the growing experience and knowledge of health care providers and vaccine administrators, modifications implemented by vaccine manufacturers, and the decreasing number of COVID-19 cases due to public health measures and vaccination efforts [69,70]. Interestingly, when compared to Pfizer-BioNTech, Moderna has a relatively higher reporting rate of AEFIs during the initial vaccination stage, the SARS-CoV-2 Delta variant period, and the SARS-CoV-2 Omicron variant period. This disparity in the AEFIs reported between Moderna and Pfizer-BioNTech, despite both vaccines using the messenger RNA (mRNA) platform, can be attributed to multiple factors [71]. One possible reason is the higher dose of mRNA administered in each shot of the Moderna vaccine (100 μg) compared to the Pfizer-BioNTech vaccine (30 μg) [72]. The higher dose may trigger a stronger immune response in some individuals, increasing the likelihood of AEFIs. Additionally, there is a difference in the dosing intervals between the 2 vaccines. Pfizer-BioNTech doses are administered 3 weeks apart and Moderna doses administered 1 month apart [73]. The longer interval for Moderna could potentially allow for a more pronounced immune response, potentially contributing to a higher rate of AEFIs reported. However, as more data were collected and larger populations were vaccinated, it became evident that the overall rates of AEFIs between the 2 vaccines were comparable. This could be due to several factors, such as increased familiarity and experience with the vaccines, improved reporting systems, and a better understanding of potential side effects.

Our spatial analysis revealed that the north and middle regions of the United States exhibit higher case- and serious case–reporting rates compared to the southeast and southwest regions. The elevated reporting rate of AEFIs associated with COVID-19 vaccines in the middle and north regions could be attributed to several factors. One possible explanation is the disparity in vaccination coverage. The middle and north regions may have a lower proportion of vaccinated individuals, increasing the likelihood of AEFIs being reported. Furthermore, variations in health care access across different regions can also contribute to differences in reporting rates.

The COVID-19 symptoms reported show a notable similarity between the southeast region (region IV) and the south-central region (region VI). One possible explanation is the geographical proximity and shared demographics within these regions. When people live in close geographic proximity, they often share similar environmental exposures, lifestyles, and genetic backgrounds. These factors can contribute to a higher likelihood of experiencing similar symptoms when infected with COVID-19.

Interestingly, despite the geographical distance, the northwest (region X) and northeast (region I, II, and III) regions of the United States also exhibit similarity in the COVID-19 symptoms reported. The similarity is particularly pronounced in serious cases. This can be attributed to shared population characteristics, such as age distributions, cultural practices, or socioeconomic factors, which influence the prevalence and reporting of specific symptoms. Moreover, the presence of specific COVID-19 variants or strains within these regions could also contribute to the similarity in the symptoms reported. Variants of the virus may exhibit unique characteristics, including symptom profiles, which can result in similarities in the symptoms reported within specific regions.
Population migration and travel patterns can also play a role in the similarity of the symptoms reported. Individuals residing in close proximity or frequently traveling between regions can contribute to the transmission and dissemination of specific COVID-19 strains, leading to similarities in symptom reporting. Additionally, similarities in health care infrastructure, medical practices, and access to testing facilities may contribute to the observed similarities in the symptoms reported. Consistent testing protocols and diagnostic criteria across these regions can lead to more uniform reporting of symptoms.

A comparative analysis with other countries can provide insights into the generalizability of our findings and the impact of different health care systems and vaccination strategies. For example, Abukhalil et al [30] conducted a questionnaire-based retrospective cross-sectional study to monitor AEFIs associated with COVID-19 vaccines in Palestine. They found that fever, chills, headache, fatigue, and pain are the most commonly reported AEFIs [30]. Similarly, Bannister et al [25] analyzed questionnaires completed by participants in the Danish National Cohort Study of Effectiveness and Safety of SARS-CoV-2 Vaccines (ENFORCE) and revealed that fatigue, muscle pain, and headache are the most commonly reported AEFIs. Additionally, Nawaz et al [22] conducted a survey-based cross-sectional study in Pakistan and found that injection site pain, fatigue, and muscle ache are the most commonly reported AEFIs associated with COVID-19 vaccines. Through comparing data and trends across nations, we can identify similarities in AEFI reporting and population responses. Understanding these patterns can inform the development of more effective vaccination strategies to address AEFIs associated with COVID-19 vaccines. Furthermore, comparing AEFI profiles across countries can contribute to the identification of rare or unusual patterns, prompting further investigation into potential vaccine-related risks or benefits.

**Strengths and Limitations**

Our analysis of AEFIs associated with COVID-19 vaccines is comprehensive compared to existing research. First, unlike previous studies, we used both vaccine administration data and VAERS data. This approach allowed us to evaluate the reporting incidence of AEFIs, rather than relying solely on absolute values. Using both data sets has the advantage of providing a more accurate representation of AEFIs associated with COVID-19 vaccines. Second, we analyzed 3 years of data to provide a more complete and convincing conclusion. Most studies have been conducted over a shorter period, which limits their scope and reliability. In contrast, our study period of 3 years allowed us to assess the long-term AEFIs associated with these vaccines. This approach also provides a more dynamic and objective understanding of the risks associated with COVID-19 vaccines. Our approach allowed us to identify the most common AEFIs associated with COVID-19 vaccines and their underlying mechanisms. This information can help health care providers better manage and treat AEFIs associated with COVID-19 vaccines. Finally, we harnessed the power of an embedding model to examine the similarity of symptoms across diverse regions. The model can handle complex word structures and improve the representation of rare or unseen terms by considering subword units. This rationale helps extract more precise and meaningful insights from biomedical texts and also facilitates various applications, such as biomedical information retrieval, named entity recognition, and text classification.

Overall, the embedding provides a valuable resource for advancing biomedical text analysis and accelerating biomedical research.

However, it is important to acknowledge that our study has several limitations. First, the quality of the data was inadequate as we were unable to access certain administration data due to their unavailability in the CDC database, which could have added more depth to our analysis. Moreover, VAERS is a passive reporting system in which AEFIs are not automatically collected, and anyone can submit VAERS reports, which sometimes lack details or contain errors [74]. As this study reported data collected by a surveillance system, it did not determine the safety of the vaccines but rather was prone to report the most frequently monitored and reported AEFIs. Furthermore, it is crucial to note that VAERS does not validate the causation between COVID-19 vaccines and the AEFIs reported [74]. Second, we failed to include cases where patients received a combination of Pfizer-BioNTech and Moderna vaccines, resulting in selection bias. Unfortunately, this also meant that we were unable to filter out such doses in the administration data. Nonetheless, it is important to note that the number of individuals who received mixed doses per week was minimal compared to the overall number of vaccinations, which mitigated the potential impact of this bias. Additionally, reporting bias cannot be entirely ruled out, as some individuals may choose not to report adverse events to VAERS due to various reasons, including reluctance, lack of awareness, or difficulties in accessing the reporting system. This could result in underreporting of certain AEFIs and potentially affect the accuracy and completeness of our analysis. Lastly, there may be instances of history bias in our data, where the symptoms observed may not be attributed to the COVID-19 vaccine itself but rather to historical events or interventions that occurred during the same period. These external factors may confound the interpretation of the AEFIs reported. Despite these limitations, we believe that our study provides a comprehensive analysis of AEFIs associated with COVID-19 vaccines.

**Conclusion**

The study introduced a potentially valuable approach to monitoring AEFIs, particularly for serious cases, which could bolster research in regions experiencing unusual or severe adverse reactions. These findings imply that higher vaccination coverage may decrease the AEFIs reported, leading to increased confidence in vaccines. Our study highlights the importance of postlicensure monitoring in understanding AEFIs associated with COVID-19 vaccines. Although our analysis provides valuable insights into the temporal and spatial patterns of the symptoms reported, it is crucial to acknowledge the limitations in data quality, including reporting and selection biases. Moving forward, efforts should focus on improving surveillance methods to enhance the accuracy and representativeness of AEFI reporting. This study underscores the need for continuous monitoring, supporting the development of informed public health policies. Looking ahead, nations should prioritize implementing effective measures, such as stringent lockdowns,
widespread testing and contact tracing, robust health care infrastructure, and clear communication strategies. Additionally, ensuring equitable access to vaccines and promoting vaccine confidence are crucial for achieving optimal vaccination coverage. By implementing these strategies, nations can enhance their preparedness and response capabilities, reducing the impact of future pandemics.

Acknowledgments
This paper was partially supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (award numbers R01AI130460 and U24AI171008).

Data Availability
The data sets generated and analyzed during this study are available in the Centers for Disease Control and Prevention repository [11] and the Vaccine Adverse Event Reporting System repository [44].

Authors’ Contributions
CT and YC were responsible for leading the experimental design, while YL contributed to the experimental design, conducted data analysis, and drafted the manuscript. CT also provided support and manuscript editing. YL and JL performed data visualization, and YL and YD conducted data collection. All authors have reviewed and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File, 323 KB - publichealth_v10i1e51007_app1.docx ]

Multimedia Appendix 2
Supplementary files.
[ZIP File (Zip Archive), 357 KB - publichealth_v10i1e51007_app2.zip ]

References


38. Stokel-Walker C. How are vaccines being adapted to meet the changing face of SARS-CoV-2? BMJ 2022 Jun 01;377:o1257. [doi: 10.1136/bmj.o1257] [Medline: 35649582]


Abbreviations

AEFI: adverse event following immunization
AIE: autoimmune encephalitis
CDC: Centers for Disease Control and Prevention
CSV: comma-separated value
GBS: Guillain-Barre syndrome
HLGT: high-level group term
HLT: high-level term
LLT: lowest-level term
MD: Meniere disease
MedDRA: Medical Dictionary for Regulatory Activities
Integrated Network Analysis of Symptom Clusters Across Monkeypox Epidemics From 1970 to 2023: Systematic Review and Meta-Analysis

Shu Su¹*, PhD; Meng Jia¹*, MSc; Yingni Yu¹, MSc; Hu Li¹, MD, PhD; Wenwei Yin¹, PhD; Yi Lu¹, PhD; Rongzhong Huang¹, PhD; Rong Xiang², PhD; Huizhe Huang¹*, PhD; Peng Hu¹*, MD, PhD

Key Laboratory of Molecular Biology for Infectious Diseases, Department of Infectious Diseases, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China

¹these authors contributed equally

Corresponding Author:
Peng Hu, MD, PhD
Key Laboratory of Molecular Biology for Infectious Diseases
Department of Infectious Diseases
The Second Affiliated Hospital of Chongqing Medical University
Tianwen Road 288
Chongqing, 400060
China
Phone: 86 13608338064
Email: hupengcq@hospital.cqmu.edu.cn

Abstract

Background: The worldwide spread of monkeypox (mpox) has witnessed a significant increase, particularly in nonendemic countries.

Objective: We aimed to investigate the changing clinical symptoms associated with mpox from 1970 to 2023 and explore their interrelations.

Methods: In this systematic review and meta-analysis, 3 electronic databases were searched for English peer-reviewed studies conducted from January 1970 to April 2023 that reported any symptoms among confirmed mpox cases. We categorized the mpox epidemics into 3 periods: 1970-2002 (period 1, within the African region), 2003-2021 (period 2, epidemics outside Africa), and 2022-2023 (period 3, worldwide outbreak). Following PRISMA guidelines, a meta-analysis was performed to estimate the pooled prevalence for each symptom. The correlation among symptoms was analyzed and visualized using network analysis.

Results: The meta-analysis included 61 studies that reported 21 symptoms in 720 patients from period 1, 39 symptoms in 1756 patients from period 2, and 37 symptoms in 12,277 patients from period 3. The most common symptom among patients from all 3 periods was rash (period 1: 92.6%, 95% CI 78.2%-100%; period 2: 100%, 95% CI 99.9%-100%; and period 3: 94.8%, 95% CI 90.9%-98.8%), followed by lymphadenopathy (period 1: 59.8%, 95% CI 50.3%-69.2%; period 2: 74.1%, 95% CI 64.2%-84.1%; and period 3: 61.1%, 95% CI 54.2%-68.1%). Fever (99%, 95% CI 97%-100%), enlarged lymph nodes (80.5%, 95% CI 75.4%-85.0%), and headache (69.1%, 95% CI 4%-100%) were the main symptoms in period 1, with a significant decrease in period 3: 37.9%, 31.2%, and 28.7%, respectively. Chills/rigors (73.3%, 95% CI 60.9%-85.7%), fatigue (68.2%, 95% CI 51.6%-84.8%), and dysphagia/swallowing difficulty (61.2%, 95% CI 10.5%-100%) emerged as primary new symptoms in period 2 and decreased significantly in period 3. Most other symptoms remained unchanged or decreased in period 3 compared to the former 2 periods. Nausea/vomiting had the highest degree of correlation (with 13 symptoms) and was highly positively correlated with lymphadenopathy (r=0.908) and conjunctivitis (r=0.900) in period 2. In contrast, rash and headache were 2 symptoms with the highest degree of correlation (with 21 and 21 symptoms, respectively) in period 3 and were highly positively correlated with fever (r=0.918 and 0.789, respectively).

Conclusions: The manifestation of symptoms in patients with mpox has become more diverse, leading to an increase in their correlation. Although the prevalence of rash remains steady, other symptoms have decreased. It is necessary to surveil the evolving nature of mpox and the consequential changes in clinical characteristics. Epidemic countries may shift their focus on the potential association among symptoms and the high synergy risk.

Trial Registration: PROSPERO Registration: CRD42023403282; http://tinyurl.com/yruuas5n
Introduction

Background

Monkeypox (mpox), caused by the mpox virus within the genus Orthopoxvirus (family Poxviridae, subfamily Chordopoxvirinae), is an emerging zoonotic disease that results in a smallpox-like illness in humans [1,2]. The first case of mpox was reported in Zaire (current Democratic Republic of the Congo, DRC) in 1970, after which it spread to other countries in Central and West Africa [1,3]. Although mpox has predominantly affected African populations, in the spring of 2003, the disease was reported outside Africa for the first time in the Midwest region of the United States due to infection from imported animals, and then sporadic reports of infections emerged from other countries caused by imported cases from West Africa until the pandemic [4]. Since May 2022, the disease has emerged as a major worldwide public health crisis, with 87,042 cases reported worldwide by April 30, 2023, mainly among men who have sex with men (MSM) [5]. The potential worldwide impact of mpox necessitates an urgent need for an improved understanding of the disease’s epidemiology, symptomatology, and management.

Prior to the 2022 worldwide pandemic, mpox was considered a rare disease with a low case fatality rate (CFR, <10%) in comparison to its closely related virus, variola major (CFR=30%) [6]. Despite this, mpox remained a serious health threat in the DRC and other countries in Western and Central Africa [7], particularly among the pediatric population and patients with immunodeficiency [8]. The incubation period of mpox ranges from 5 to 21 days and is characterized by nonspecific symptoms, such as fever, headache, and lymphadenopathy [9]. Skin lesions were observed in 95% of cases, with rashes being the most common symptom, typically appearing on the face and extremities within 1-3 days after the onset of fever [10,11]. The diverse range and frequency of the clinical symptoms of mpox are influenced by various risk factors, including modes of transmission and demographics of affected populations [12]. For example, since 2022, the majority of mpox cases have occurred among MSM, resulting in comorbidity with other infectious diseases, such as HIV and syphilis [13]. An improved understanding of the clinical profile of mpox can aid in the development of effective strategies for managing and responding to outbreaks of the disease.

Before the onset of the pandemic in 2022, limited and outdated data existed regarding the epidemiology and clinical characteristics of mpox from an epidemiological perspective [14-16]. However, since the emergence of the worldwide mpox pandemic in 2022, numerous reviews and meta-analyses have begun to investigate the symptoms of mpox, although most focus only on 2022 [17-20]. Only 2 papers have compared mpox symptom prevalence before and after 2022, with limitations including sparse literature, especially prior to 2010 [21,22], and a lack of correlation analyses between symptoms. Therefore, there is a need for a comprehensive longitudinal analysis to investigate the evolution of mpox symptoms and their changes over time.

Objective

To address this gap, we conducted a review and meta-analysis to comparatively analyze mpox symptom prevalence and correlations during 3 periods: 1970-2002 in Africa; 2003-2021 in Africa, Europe, and North America; and 2022-2023 worldwide.

Methods

Protocol Registration

This research adheres to the methodological recommendations of the Cochrane systematic review methodology guidance [23] and conforms to the guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (see the PRISMA checklist in Table S9 in Multimedia Appendix 1) [24]. This study has been registered with PROSPERO (International Prospective Register of Systematic Reviews; registration number CRD42023403282).

Search Strategy

Two independent investigators (authors SS and MJ) conducted a comprehensive literature search of peer-reviewed research papers published from January 1970 to April 2023 in the following English literature databases: PubMed, Web of Science, and ScienceDirect. We also searched all mpox-related public databases. Keywords pattern in the database search were used through combinations of monkeypox (eg, “monkeypox” OR “monkeypox virus” OR “human monkeypox”), symptoms (eg, “symptoms” OR “characteristics” OR “clinical characteristics” OR “clinical symptoms”), or cases (eg, “cases” OR “cases report”) and excluded reviews. Full details of the search strategy are listed in Table S1 in Multimedia Appendix 1. A manual search of the reference lists of the published papers was also performed.

We restricted our search to research conducted on humans with observational studies, discarding veterinary investigations. Papers were included if they contained clinical or epidemiological information relevant to the current monkeypox virus outbreak or historical information to establish context. The following types of publications were excluded: news reports, conference abstracts, mathematical modeling studies, dissertations, and studies with mpox self-reported prevalence. Two independent investigators (SS and MJ) confirmed the eligibility of each study; if necessary, a third senior reviewer (author PH) resolved differences. The PRISMA flow diagram was applied to report the paper search process.

The following information was collected from each of the studies: authors, publication year, title, study design, study period, study area, sample size, study population, age, and
To assess the impact of bias on our findings, we conducted a Sensitivity Analysis using ROBINS-I tool for nonrandomized controlled trial assessment. At the protocol stage, we specified the review question of participants as confirmed patients with mpox and outcomes as any symptom report. Additionally, we listed the confounding domains relevant to all studies, including time-varying confounding and baseline confounding, and the cointerventions that could be different between intervention groups and that could impact outcomes, such as any treatment. We assessed the overall risk of bias according to 7 items: bias due to confounding, bias in the selection of participants for the study, bias in the classification of interventions, bias due to deviations from intended interventions, bias due to missing data, bias in the measurement of outcomes, and bias in the selection of the reported result. Each item was classified as low risk, moderate risk, serious risk, or critical risk of bias. The overall risk of bias was assessed according to the results of the 7 items [25]. The detailed results are shown in Table S2 in Multimedia Appendix 1.

Study Period
In this review, we classified the mpox epidemics into 3 distinct periods. Period 1, spanning from 1970 to 2002, was characterized by the limited transmission of the virus within the African region. Period 2, spanning from 2003 to 2021, marked the first outbreak of mpox outside Africa, with confirmed cases reported in the United States, the United Kingdom, Singapore, and Israel. Period 3, from 2022 to 2023, witnessed an unprecedented worldwide explosion of mpox cases.

Meta-Analysis
To estimate the prevalence of mpox symptoms, we conducted meta-analyses by combining data from various studies using random effects models that fitted study-specific effects. We calculated corresponding 95% CIs. Heterogeneity analysis was performed to quantify the inconsistency among results, and it was expressed using $I^2$, a quantity representing the degree of heterogeneity from 0% to 100%. Therefore, $I^2$ with experience values of 25%, 50%, and 75% indicated low, moderate, and high heterogeneity, respectively [26]. The meta-analyses were conducted using R Studio software (version 2023.06.2, build 561), which is a specialized integrated development environment (IDE) designed for R programming (version 4.2.1, R Foundation for Statistical Computing).

Sensitivity Analysis
To assess the impact of bias on our findings, we conducted a sensitivity analysis by excluding studies with a moderate and higher risk of bias, thereby including only low-risk studies in a subgroup meta-analysis.

Subgroup Analysis
To assess the impact of geographical differences on our findings, we conducted a subgroup analysis by grouping patients based on the regional distribution of the World Health Organization (WHO). $P \leq 0.05$ was considered a statistically significant difference in prevalence among regions.

Network Analysis
We also conducted a network analysis to compare changes in mpox symptoms across the 3 periods and created nodes (symptoms) and edges (symptom correlation). Within each period, the edge color reflected the correlation properties, where pink color represented positive correlation; thickness indicated the correlation strength between symptoms, where correlations were calculated through Spearman intragroup correlation analyses using R software v4.2.1 and filtered significantly correlated ($P \leq 0.05$) symptom pairs. Similarly, the size and color depth of the nodes represented degrees, which were acquired through clustering analyses of a subset of nodes and visualized using Cytoscape software (v3.7.2).

Results
Overview
The initial search strategy included 2916 unique titles. After excluding 825 (28.3%) duplicate records and 516 (17.7%) unrelated subjects, 1575 (54%) studies remained. Of these, 142 (9%) full-text papers were retrieved. The majority were excluded for reasons such as being drug or biological studies, comments, reviews, or conference papers. After reviewing the 142 papers, 19 (13.4%) were excluded due to duplicate data, while 24 (16.9%) and 38 (26.8%) contained unclear or incomplete information, respectively. Additionally, 19 (13.4%) papers presented the same outbreak with duplicate data. Ultimately, our systematic review and meta-analysis included 61 (43%) papers that provided information about cases and symptoms (Figure 1). Of these, 29 (47.5%) papers were assessed as having low risk of bias, 23 (37.7%) were assessed as having moderate risk of bias, 4 (6.6%) exhibited a serious risk of bias, while the remaining 5 (8.2%) were categorized as having a critical risk of bias (Table S2 in Multimedia Appendix 1).

From 1970 to 2002 (period 1), all 720 patients from Africa were included in our meta-analysis (n=4, 6.6%, studies). The median age was 5.4 years (IQR 2.7-9.8; from n=3, 75%, of 4 studies). The proportion of male patients was 34.9% (n=251, range 15.9-56.8%; from n=4, 6.6%, studies). From 2003 to 2021 (period 2), a total of 12 (19.7%) studies were included. Most of the 1756 patients were also from Africa (n=1259, 71.7%, patients from n=5, 41.7%, studies), followed by the Americas (n=457, 26%, patients from n=3, 25%, studies), Europe (n=5, 0.3%, patients from n=2, 16.7%, studies), and the Western Pacific (n=2, 0.1%, patients from n=2, 16.7%, studies). The median age was 21.7 years (IQR 13.6-32.1; from n=8, 66.7%, studies), and the proportion of male patients was 53.3% (n=934; from n=11, 91.7%, studies).
From 2022 to 2023 (period 3), 45 (73.7%) studies involving 12,277 patients were included in our study. A total of 9587 (78.1%) patients were reported in the Americas (n=13, 28.9%, studies), followed by Europe (n=1680, 13.7%, patients from n=24, 5.3%, studies). The median age was 34.5 years (IQR 28.5-42.0; from n=45, 100%, studies), and the proportion of male patients was 86.7% (n=10,640; from n=45, 100%, studies).

Among the male patients, the proportion of MSM was 88.8% (n=2340; from n=27, 60%, studies), with HIV (n=2116, 37.8%; from n=30, 66.7%, studies), syphilis (n=125, 18.4%; from n=9, 20%, studies), and other sexually transmitted infections (STIs; n=221, 23.6%; from n=6, 13.3%, studies). The study details are given in Table 1. The period prevalence was calculated for each period (Figure S1 in Multimedia Appendix 1).

Figure 1. Study selection flowchart.
### Table 1. Characteristics of the included studies (N=61).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Studies, n (%)</td>
<td>Patients</td>
<td>Studies, n (%)</td>
</tr>
<tr>
<td>Overall, n (%)</td>
<td>4 (6.6)</td>
<td>720 (4.9)</td>
<td>12 (19.7)</td>
</tr>
<tr>
<td><strong>Study type, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational</td>
<td>2 (50.0)</td>
<td>370 (51.4)</td>
<td>7 (58.3)</td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>0</td>
<td>0</td>
<td>4 (33.3)</td>
</tr>
<tr>
<td>Cohort</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Prospective cohort</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Retrospective case control</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Retrospective observational</td>
<td>2 (50.0)</td>
<td>370 (51.4)</td>
<td>3 (25.0)</td>
</tr>
<tr>
<td>Descriptive</td>
<td>2 (50.0)</td>
<td>350 (48.6)</td>
<td>5 (41.7)</td>
</tr>
<tr>
<td>Case report</td>
<td>1 (25.0)</td>
<td>6 (0.8)</td>
<td>5 (41.7)</td>
</tr>
<tr>
<td>Case series</td>
<td>1 (25.0)</td>
<td>344 (47.8)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Regions, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>4 (100.0)</td>
<td>720 (100.0)</td>
<td>5 (41.7)</td>
</tr>
<tr>
<td>Americas</td>
<td>0</td>
<td>0</td>
<td>3 (25.0)</td>
</tr>
<tr>
<td>Eastern Mediterranean</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Europe</td>
<td>0</td>
<td>0</td>
<td>2 (16.7)</td>
</tr>
<tr>
<td>Western Pacific</td>
<td>0</td>
<td>0</td>
<td>2 (16.7)</td>
</tr>
<tr>
<td><strong>Patients’ characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>3 (75.0)</td>
<td>5.4 (2.7-9.8); 376 (52.2)a</td>
<td>8 (66.7)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>4 (100.0)</td>
<td>251 (34.9)b</td>
<td>11 (91.7)</td>
</tr>
<tr>
<td>MSMc, n (%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HIV, n (%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Syphilis, n (%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other STIsf, n (%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*a*Patients in the included studies.

bMale patients in the included studies.

cMSM: men who have sex with men.

dMSM patients in the included studies.

ePatients with infection/disease in the included studies.

fSTI: sexually transmitted infection.

### Clinical Symptom Prevalence

We conducted a meta-analysis of 21 symptoms reported from 1970 to 2002 (Figure 2). The most commonly reported symptoms were fever (99%, 95% CI 97%-100%), rash (92.6%, 95% CI 78.2%-100%), and enlarged lymph nodes (80.5%, 95% CI 75.4%-85.0%). These were followed by headache (69.1%, 95% CI 4%-100%), lymphadenopathy (59.8%, 95% CI 50.3%-69.2%), malaise/bedridden status/asthenia (52.4%, 95% CI 44.8%-60.0%), tonsillitis (51.4%, 95% CI 45.4%-57.4%), swelling (51.3%, 95% CI 39.7%-62.8%), mouth ulcers (49.1%, 95% CI 43.7%-54.5%), sore throat (47.5%, 95% CI 25.2%-69.7%), cough (38%, 95% CI 33.7%-42.3%), genital ulcers (25.2%, 95% CI 20.2%-30.7%), sweating/joint pain (16.7%, 95% CI 0.4%-64.1%), stiff neck (16.7%, 95% CI 0.4%-64.1%), conjunctivitis (16.7%, 95% CI 12.5%-21.5%), and hepatomegaly (10.3%, 95% CI 7.0%-14.4%). The prevalence of proctalgia/diarrhea, nausea/vomiting, dehydration, tenesmus, corneal opacity/photophobia, and alopecia was relatively low (<10%).

In our meta-analysis covering the period of 2003-2021, we identified 38 symptoms associated with mpox (Figure 3). Rash was the most prevalent symptom (100%, 95% CI 99.9%-100%), followed by fever (88%, 95% CI 82.1%-94.0%) and...
lymphadenopathy (74.1%, 95% CI 64.2%-84.1%). The former decreased, while the latter increased. Chills/rigors (73.3%, 95% CI 60.9%-85.7%), adenopathy (70.6%, 95% CI 52.5%-84.9%), fatigue (68.2%, 95% CI 51.6%-84.8%), difficulty breathing (63.2%, 95% CI 38.4%-83.7%), dysphagia/difficulty swallowing (61.2%, 95% CI 10.5%-100%), itching (59.2%, 95% CI 45.2%-73.2%), myalgia (52.7%, 95% CI 35.6%-69.8%), and diarrhea (50.3%, 95% CI 0%-100%) were newly reported but remained highly prevalent. The prevalence of genital ulcers (57.1%, 95% CI 43.0%-71.1%) increased compared to the period of 1970-2002, while sweating/joint pain (63.9%, 95% CI 49.8%-78.1%), mouth ulcers (53.6%, 95% CI 33.0%-74.3%), and cough (43%, 95% CI 28.0%-57.9%) remained steady. The prevalence of body pain (43.4%, 95% CI 25.4%-61.4%), shortness of breath (39.5%, 95% CI 0%-84.1%), anorexia (30%, 95% CI 0.0%-72.2%), corneal opacity/photophobia (27.1%, 95% CI 18.9%-35.3%), arthralgia (22.1%, 95% CI 16.0%-29.2%), nasal congestion/rhinorrhea (20%, 95% CI 3.5%-36.4%), and hemorrhagic skin lesions (12.5%, 95% CI 4.2%-26.8%) was newly reported but was <50%. The prevalence of malaise/bedridden status/asthenia (32.3%, 95% CI 4.8%-59.8%), conjunctivitis (23.2%, 95% CI 13.4%-33.1%), and stiff neck (11.8%, 95% CI 3.3%-27.5%) remained steady compared to the former period, while nausea/vomiting (20%, 95% CI 12.4%-27.5%) increased. The prevalence of 10 (26.3%) other symptoms was <10%, of which 7 (70%) were newly reported symptoms: tongue sores, dehydration, ear pain, wheezing, confusion, scrotal edema, and seizures. Swelling, which was reported from 1970 to 2002, had a decreased prevalence of 7.7% (95% CI 0%-14.7%), while hepatomegaly (9.5%, 95% CI 0%-22.7%), swelling, and proctalgia/diarrhea (5.9%, 95% CI 0.7%-19.7%) remained steady in this period.

Since 2022, the included literature reports have listed 37 clinic characteristics observed in patients during the mpox pandemic (Figure 4). Rash was the most prevalent symptom among patients, and its prevalence (94.8%, 95% CI 90.9%-98.8%) was higher than that of other symptoms, similar to the 2003-2021 period, followed by fever (61.1%, 95% CI 56.7%-66.6%), enlarged lymph nodes (51.8%, 95% CI 40.8%-62.7%), headache (37.9%, 95% CI 32.0%-43.8%), and chills/rigors (36.6%, 95% CI 25.5%-47.7%), which were lower than the periods from 2003 to 2021. Furthermore, flu-like symptoms (49.2%, 95% CI 41.6%-56.7%), perianal ulcer (40.8%, 95% CI 33.6%-48.0%), dysuria (33.1%, 95% CI 57.2%-60.5%), perioral lesion (22.9%, 95% CI 9.8%-36.0%), rectal bleeding (15.1%, 95% CI 1.6%-28.6%), tonsillitis (11.7%, 95% CI 8.1%-15.3%), difficulty breathing (11.7%, 95% CI 8.1%-16.3%), pus or blood in stools (11.5%, 95% CI 10.7%-12.2%), tenesmus/constipation (10.7%, 95% CI 10.0%-11.5%), perianal ulcer (9.9%, 95% CI 4.3%-15.5%), and tenesmus/constipation (9.4%, 95% CI 6.4%-12.4%) were newly reported and had a lower prevalence. Other past reported symptoms, including dysphagia/difficulty swallowing (27.9%, 95% CI 21.1%-34.7%), scrotal edema (25.9%, 95% CI 0%-60.8%), body pain (22.6%, 95% CI 10.4%-34.7%), arthralgia (21%, 95% CI 0.8%-41.2%), proctalgia/diarrhea (16.7%, 95% CI 1.8%-25.6%), nausea/vomiting (13.3%, 95% CI 2.2%-24.5%), corneal opacity/photophobia (10.4%, 95% CI 0%-25.5%), nasal congestion/rhinorrhea (7.1%, 95% CI 0%-16.2%), and seizures (0.7%, 95% CI 0%-4.1%) remained steady, while the prevalence of sore throat (25.1%, 95% CI 20.4%-29.9%), mouth ulcers (13.7%, 95% CI 8.2%-19.2%), cough (8.1%, 95% CI 3.9%-12.3%), and conjunctivitis (5.4%, 95% CI 4.6%-6.2%) was lower than that in former periods.

The prevalence of symptoms from 1970 to 2023 was included in the meta-analysis as well, involving 51 reported symptoms from 13,196 patients (Figure 5). Generally, rash (95.1%, 95% CI 92.0%-98.3%) was the prevalent symptom, followed by adenopathy (70.6%, 95% CI 52.5%-84.9%), fever (69.3%, 95% CI 63.9%-74.1%), and lymphadenopathy (61.0%, 95% CI 55.7%-66.3%), enlarged lymph nodes (55.4%, 95% CI 43.7%-67.1%), itching (52.8%, 95% CI 34.2%-71.3%), fatigue (52.6%, 95% CI 41.9%-63.2%), and diarrhea (50.3%, 95% CI 0%-100%). The prevalence of the other 40 (78.4%) symptoms was <50%. The prevalence of 14 (27.5%) symptoms ranged from 30% to 50%, including flu-like symptoms (49.2%, 95% CI 41.6%-46.7%), sweating/joint pain (47.9%, 95% CI 29.1%-66.8%), and chills/rigors (47.2%, 95% CI 33.5%-60.9%). For 17 (33.3%) symptoms, the prevalence ranged from 10% to 30%, including body pain (28.6%, 95% CI 17.4%-39.8%), mouth ulcers (27.9%, 95% CI 15.1%-40.7%), and cough (26%, 95% CI 15.3%-33.6%). Another 9 (17.6%) symptoms, such as tongue sores (9.5%, 95% CI 1.2%-30.4%), ear pain (8.8%, 95% CI 1.9%-23.7%), and tenesmus (8.5%, 95% CI 5.4%-11.6%), had <10% prevalence.
Figure 2. Meta-analysis of the prevalence of clinical symptoms among mpox patients from 1970 to 2002. mpox: monkeypox.

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Events</th>
<th>Total</th>
<th>Prevalence and 95% CI</th>
<th>I² (%)</th>
<th>Egger's test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever</td>
<td>623</td>
<td>632</td>
<td>0.990 (0.970-1.000)</td>
<td>78.8</td>
<td>0.22</td>
</tr>
<tr>
<td>Rash</td>
<td>611</td>
<td>707</td>
<td>0.926 (0.782-1.000)</td>
<td>97.7</td>
<td>0.41</td>
</tr>
<tr>
<td>Enlarged lymph nodes</td>
<td>227</td>
<td>282</td>
<td>0.805 (0.754-0.850)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Headache</td>
<td>284</td>
<td>288</td>
<td>0.691 (0.040-1.000)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Lymphadenopathy</td>
<td>436</td>
<td>714</td>
<td>0.598 (0.503-0.692)</td>
<td></td>
<td>82.9 0.45</td>
</tr>
<tr>
<td>Malaise/bedridden status/asthena</td>
<td>185</td>
<td>330</td>
<td>0.524 (0.448-0.600)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Tonsillitis</td>
<td>145</td>
<td>282</td>
<td>0.514 (0.454-0.574)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Swelling</td>
<td>40</td>
<td>84</td>
<td>0.513 (0.397-0.628)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Mouth ulcers</td>
<td>169</td>
<td>344</td>
<td>0.491 (0.437-0.545)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Sore throat</td>
<td>359</td>
<td>632</td>
<td>0.475 (0.252-0.697)</td>
<td></td>
<td>85.8 0.43</td>
</tr>
<tr>
<td>Cough</td>
<td>242</td>
<td>632</td>
<td>0.380 (0.337-0.423)</td>
<td></td>
<td>37.5 0.37</td>
</tr>
<tr>
<td>Genital ulcers</td>
<td>71</td>
<td>282</td>
<td>0.252 (0.202-0.307)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Sweating/joint pain</td>
<td>1</td>
<td>6</td>
<td>0.167 (0.004-0.641)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Stiff neck</td>
<td>1</td>
<td>6</td>
<td>0.167 (0.004-0.641)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Conjunctivitis</td>
<td>47</td>
<td>282</td>
<td>0.167 (0.125-0.215)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Hepatomegaly</td>
<td>29</td>
<td>282</td>
<td>0.103 (0.070-0.144)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Proctalgia/diarrhea</td>
<td>54</td>
<td>626</td>
<td>0.083 (0.037-0.129)</td>
<td></td>
<td>78.5 /</td>
</tr>
<tr>
<td>Nausea/vomiting</td>
<td>17</td>
<td>282</td>
<td>0.060 (0.036-0.095)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Dehydration</td>
<td>17</td>
<td>282</td>
<td>0.060 (0.036-0.095)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Tenesmus</td>
<td>3</td>
<td>8</td>
<td>0.034 (0.007-0.096)</td>
<td></td>
<td>/</td>
</tr>
<tr>
<td>Corneal opacities/photophobia</td>
<td>14</td>
<td>626</td>
<td>0.022 (0.000-0.058)</td>
<td></td>
<td>88.1 /</td>
</tr>
<tr>
<td>Alopecia</td>
<td>9</td>
<td>432</td>
<td>0.019 (0.006-0.032)</td>
<td></td>
<td>0 /</td>
</tr>
</tbody>
</table>
Figure 3. Meta-analysis of the prevalence of clinical symptoms among mpox patients from 2003 to 2021.
**Figure 4.** Meta-analysis of the prevalence of clinical symptoms among mpox patients from 2022 to 2023. mpox: monkeypox.

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Events</th>
<th>Total</th>
<th>Prevalence and 95% CI</th>
<th>P (%)</th>
<th>Egger’s test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rash</td>
<td>10026</td>
<td>10631</td>
<td>0.948 (0.900-0.988)</td>
<td>94.6</td>
<td>0.01</td>
</tr>
<tr>
<td>Lymphadenopathy</td>
<td>2744</td>
<td>5380</td>
<td>0.611 (0.542-0.681)</td>
<td>93.9</td>
<td>0.06</td>
</tr>
<tr>
<td>Fever</td>
<td>4201</td>
<td>7529</td>
<td>0.611 (0.567-0.656)</td>
<td>85.6</td>
<td>0.02</td>
</tr>
<tr>
<td>Itching</td>
<td>340</td>
<td>1154</td>
<td>0.529 (0.175-0.883)</td>
<td>97.5</td>
<td>0.54</td>
</tr>
<tr>
<td>Enlarged lymph nodes</td>
<td>2448</td>
<td>6515</td>
<td>0.518 (0.408-0.627)</td>
<td>95.7</td>
<td>0.03</td>
</tr>
<tr>
<td>Flu-like symptoms</td>
<td>88</td>
<td>179</td>
<td>0.492 (0.416-0.567)</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Malaise/bedridden status/asthenia</td>
<td>3718</td>
<td>8963</td>
<td>0.487 (0.367-0.566)</td>
<td>97.7</td>
<td>0.58</td>
</tr>
<tr>
<td>Genital ulcers</td>
<td>3590</td>
<td>7924</td>
<td>0.466 (0.375-0.557)</td>
<td>97.8</td>
<td>0.91</td>
</tr>
<tr>
<td>Sweating/joint pain</td>
<td>344</td>
<td>1152</td>
<td>0.449 (0.141-0.757)</td>
<td>96.8</td>
<td>0.56</td>
</tr>
<tr>
<td>Fatigue</td>
<td>545</td>
<td>1254</td>
<td>0.436 (0.376-0.496)</td>
<td>69.4</td>
<td>0.7</td>
</tr>
<tr>
<td>Perianal ulcer</td>
<td>2348</td>
<td>7131</td>
<td>0.408 (0.336-0.480)</td>
<td>96.7</td>
<td>0.09</td>
</tr>
<tr>
<td>Myalgia</td>
<td>1580</td>
<td>3826</td>
<td>0.405 (0.336-0.474)</td>
<td>93</td>
<td>0.82</td>
</tr>
<tr>
<td>Headache</td>
<td>3642</td>
<td>9795</td>
<td>0.370 (0.320-0.438)</td>
<td>98.1</td>
<td>0.73</td>
</tr>
<tr>
<td>Chills/rigors</td>
<td>2893</td>
<td>7280</td>
<td>0.366 (0.255-0.477)</td>
<td>95.5</td>
<td>0.76</td>
</tr>
<tr>
<td>Dysuria</td>
<td>2074</td>
<td>5904</td>
<td>0.331 (0.057-0.605)</td>
<td>99.7</td>
<td>0.93</td>
</tr>
<tr>
<td>Dysphagia/swallowing difficulty</td>
<td>47</td>
<td>166</td>
<td>0.279 (0.211-0.347)</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>Scrotal edema</td>
<td>12</td>
<td>248</td>
<td>0.259 (0.000-0.608)</td>
<td>88.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Sore throat</td>
<td>728</td>
<td>3069</td>
<td>0.251 (0.204-0.290)</td>
<td>84.5</td>
<td>0.06</td>
</tr>
<tr>
<td>Perianal lesion</td>
<td>943</td>
<td>5568</td>
<td>0.229 (0.098-0.360)</td>
<td>99.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Body pain</td>
<td>1087</td>
<td>8081</td>
<td>0.220 (0.104-0.347)</td>
<td>98.0</td>
<td>0.09</td>
</tr>
<tr>
<td>Arthralgia</td>
<td>69</td>
<td>455</td>
<td>0.210 (0.008-0.412)</td>
<td>90.4</td>
<td>0.06</td>
</tr>
<tr>
<td>Proctalgia/diarrhea</td>
<td>925</td>
<td>8544</td>
<td>0.167 (0.018-0.256)</td>
<td>94.6</td>
<td>0</td>
</tr>
<tr>
<td>Rectal bleeding</td>
<td>930</td>
<td>7644</td>
<td>0.151 (0.016-0.286)</td>
<td>96.3</td>
<td>0.79</td>
</tr>
<tr>
<td>Mouth ulcers</td>
<td>282</td>
<td>1755</td>
<td>0.137 (0.082-0.192)</td>
<td>89.1</td>
<td>0.92</td>
</tr>
<tr>
<td>Nausea/vomiting</td>
<td>650</td>
<td>6955</td>
<td>0.133 (0.022-0.245)</td>
<td>96.7</td>
<td>0.68</td>
</tr>
<tr>
<td>Tonsillitis</td>
<td>36</td>
<td>301</td>
<td>0.117 (0.081-0.153)</td>
<td>0</td>
<td>0.11</td>
</tr>
<tr>
<td>Difficulty breathing</td>
<td>31</td>
<td>264</td>
<td>0.117 (0.081-0.163)</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Conjunctivitis</td>
<td>164</td>
<td>3016</td>
<td>0.054 (0.046-0.062)</td>
<td>0</td>
<td>/</td>
</tr>
<tr>
<td>Seizures</td>
<td>1</td>
<td>134</td>
<td>0.007 (0.000-0.041)</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>
Figure 5. Meta-analysis of the prevalence of clinical symptoms among mpox patients from 1970 to 2023. mpox: monkeypox.

Sensitivity Analysis

We excluded studies with moderate and higher risk of bias, including 3 (4.9%) studies from 1970 to 2002, 6 (9.8%) studies from 2003 to 2021, and 23 (37.7%) studies since 2022 (Table S2 in Multimedia Appendix 1). A total of 29 (47.5%) studies were included in the sensitivity analysis. In the 3 periods, rash (period 1: 100.00%, 95% CI 98.7%-100%; period 2: 99%, 95% CI 97.1%-100.00%; and period 3: 93.1%, 95% CI 87.7%-98.6%) was the most common symptom, followed by fever (period 1: 100.00%, 95% CI 98.7%-100%; period 2: 89.4%, 95% CI 83.1%-95.6%; and period 3: 58.3%, 95% CI 52.8%-63.7%). The other symptoms with a relatively higher prevalence included lymphadenopathy, headache, and fatigue. Significantly, symptom prevalence from 1970 to 2002 changed compared to former meta-analysis results, while that in the other 2 periods did not, and almost all symptom prevalence decreased since 2022 (Figure S2 in Multimedia Appendix 1).
Subgroup Meta-Analysis

To compare the geographical differences in clinical symptoms, we performed a subgroup analysis of the symptom prevalence in 3 periods (2003-2021, 2022-2023, and 1970-2023) based on the WHO regional distribution. We did not conduct a subgroup analysis for the period from 1970 to 2002 since all patients were in the African region. From 2003 to 2021, 30 symptoms were reported in Africa, 26 in the Americas, 6 in Europe, and 10 in the Western Pacific. The overall prevalence of symptoms was highest in Africa, particularly in terms of rash (100%, 95% CI 99.9%-100%), fever (92%, 95% CI 85.7%-98.2%), dysphagia/difficulty swallowing (85.5%, 95% CI 57.4%-100%), chills/rigors (73.9%, 95% CI 55.3%-92.5%), headache (66.3%, 95% CI 52.9%-79.6%), shortness of breath (63.2%, 95% CI 38.4%-83.7%), and conjunctivitis (25.3%, 95% CI 17.3%-33.4%), which were significantly higher than in some areas (P ≤ 0.05). Furthermore, characteristic symptoms in African cases included difficulty breathing (63.2%, 95% CI 38.4%-83.7%), genital ulcers (57.1%, 95% CI 43.0%-71.1%), mouth ulcers (53.6%, 95% CI 33.0%-74.3%), anorexia (30%, 95% CI 0%-72.2%), corneal opacity/photophobia (27.1%, 95% CI 18.9%-35.3%), 78.7%), and hemorrhagic skin lesions (12.5%, 95% CI 4.2%-26.8%), whose prevalence was >10%. However, the prevalence of nausea/vomiting (59.7%, 95% CI 0%-100%) in cases from the Americas was higher than in Africa (P ≤ 0.05), and some characteristic symptoms, including adenopathy (70.6%, 95% CI 52.5%-74.9%) and stiff neck (11.8%, 95% CI 33.0%-27.5%), were reported exclusively in the Americas (Table S3 in Multimedia Appendix 1).

Since 2022, cases have been primarily concentrated in the Americas (27 symptoms) and Europe (25 symptoms), with fewer patients reported in the Western Pacific (12 symptoms) and Eastern Mediterranean (7 symptoms) and no cases reported in Africa. The prevalence of symptoms in the 4 regions had generally no significant difference. However, the risk of enlarged lymph nodes (61.3%, 95% CI 49.3%-73.2%) and rectal bleeding (60%, 95% CI 36.1%-80.9%) in European cases was higher than in the Americas (P ≤ 0.05). Eastern Mediterranean cases exhibited a higher prevalence of headache (100%, 95% CI 15.8%-100%) and fatigue (100%, 95% CI 15.8%-100%), but the number of cases was relatively low. Additionally, the characteristic symptom presentations in the Americas were sweating/joint pain (35.2%, 95% CI 32.4%-38.2%), nausea/vomiting (18.1%, 95% CI 3.5%-32.7%), pus or blood in stools (11.7%, 95% CI 6.2%-17.3%), and tenesmus/constipation (10.6%, 95% CI 8.4%-12.9%) and in Europe were flu-like symptoms (49.2%, 95% CI 41.6%-56.7%), scrotal edema (45.5%, 95% CI 24.7%-66.2%), arthralgia (32.4%, 95% CI 0%-75.6%), corneal opacity/photophobia (18.7%, 95% CI 0%-51.1%), tonsillitis (11.7%, 95% CI 8.1%-15.3%), and difficulty breathing (11.7%, 95% CI 8.1%-16.3%); see Table S4 in Multimedia Appendix 1.

The merged periods showed more differences in symptom prevalence among regions. Generally, rash was a prevalent symptom, with its prevalence reaching close to 100% in all regions. Africa had the highest prevalence of the most symptoms, such as fever (94.5%, 95% CI 89.9%-99.0%), headache (68%, 95% CI 51.7%-84.3%), and chills/rigors (73.9%, 95% CI 55.3%-92.5%; P ≤ 0.05), followed by the Americas and Europe, which showed similar risks for the main symptoms. Nonendemic regions, such as the Western Pacific and Eastern Mediterranean, reported fewer symptoms, and the prevalence of the main symptoms was similar to that in Europe and the Americas (Table S5 in Multimedia Appendix 1).

Association Between Clinical Symptoms

We conducted Pearson correlation analysis on the clinical symptoms, screening out significantly correlated paired symptoms (P ≤ 0.05) and drawing a network diagram of 3 periods: 2003-2021, 2022-2023, and the overall period of 1970-2023 (Figure 6a-c). From 1970 to 2002, only 14 symptoms were found due to insufficient literature data, so we did not draw a network diagram specifically for this period.
Figure 6. Network diagram of clinical symptoms among mpox patients in (A) 2003-2021, (B) 2022-2023, and (C) 1970-2023. mpox: monkeypox.

A total of 37 symptoms showed a significantly positive correlation from 2003 to 2021 (Figure 6a). Among them, nausea/vomiting was variable with the highest degree of correlation and was positive correlated with 13 symptoms: lymphadenopathy ($r=0.908$), conjunctivitis ($r=0.900$), cough ($r=0.878$), chills/rigors ($r=0.791$), headache ($r=0.739$), photophobia ($r=0.720$), fatigue ($r=0.713$), fever ($r=0.693$), corneal opacity ($r=0.676$), rash ($r=0.652$), mouth ulcers ($r=0.648$), malaise/bedridden status/asthenia ($r=0.648$), and myalgia ($r=0.648$). This was followed by lymphadenopathy (correlated with 12 symptoms), conjunctivitis (12 symptoms), nasal congestion/rhinorrhea (11 symptoms), swelling (11 symptoms), headache (11 symptoms), cough (11 symptoms), shortness of breath (11 symptoms), and chills/rigors (10 symptoms). A total of 11 symptoms (proctalgia/diarrhea, mouth ulcers, rash, fever, photophobia, stiff neck, ear pain, wheezing, confusion, seizures, and adenopathy) were correlated with 9 other symptoms. In addition, corneal opacity was correlated with 8 symptoms, myalgia with 7 symptoms, and fatigue and genital ulcers with 6 symptoms. Another 13 symptoms (hepatomegaly, dehydration, tongue sores, dysphagia/difficulty swallowing, malaise/bedridden status/asthenia, anorexia, body pain, scrotal edema, hemorrhagic skin lesions, difficulty breathing, sore throat, sweating/joint pain, and diarrhea) had no more than 5 correlated clinic characteristics.

Since 2022, a more complicated network diagram involving 33 symptoms illustrated a substantial positive correlation (Figure 6b). Generally, rash and headache had the highest degrees of correlation and were correlated with 21 symptoms. The symptom with the highest correlation with rash was fever ($r=0.918$), followed by lymphadenopathy ($r=0.789$), headache ($r=0.787$), body pain ($r=0.610$), malaise/bedridden status/asthenia ($r=0.575$), genital ulcers ($r=0.558$), proctalgia/diarrhea ($r=0.554$), pus or blood in stools ($r=0.549$), tenesmus ($r=0.548$), perianal ulcer ($r=0.543$), and rectal bleeding ($r=0.508$). In addition, fever and rash were also the most highly correlated with headache ($r=0.789$ and 0.788, respectively). Other symptoms with a high degree of correlation included body pain (20 symptoms), fever (19 symptoms), malaise/bedridden status/asthenia (19 symptoms), chills/rigors (19 symptoms), rectal bleeding (19 symptoms), pus or blood in stools (18 symptoms), tenesmus (18 symptoms), enlarged lymph nodes (17 symptoms), perianal ulcer (17 symptoms), and proctalgia/diarrhea (17 symptoms).

We also mapped the network diagram after merging 3 periods. The symptom with the highest number of associated symptoms ($n=26$) was headache (Figure 6c). Among them, the 2 symptoms with the highest correlation were fever ($r=0.780$) and rash ($r=0.754$), followed by lymphadenopathy ($r=0.576$), myalgia ($r=0.507$), and proctalgia/diarrhea ($r=0.500$); the $r$ value for the other 21 symptoms was $<0.500$. Conjunctivitis had the second-highest degree of correlation and was correlated with 25 symptoms: it has high correlation with nausea/vomiting ($r=0.844$), and the correlation with the other 24 symptoms was relatively low ($r<0.05$). Other symptoms with a high degree of correlation included body pain (23 symptoms), nausea/vomiting (23 symptoms), cough (23 symptoms), nasal
congestion/rhinorrhea (23 symptoms), rash (22 symptoms), fever (21 symptoms), chills/rigors (21 symptoms), rectal bleeding (19 symptoms), malaise/bedridden status /asthenia (18 symptoms), proctalgia/diarrhea (17 symptoms), pus or blood in stools (17 symptoms), shortness of breath (17 symptoms), and tenesmus (17 symptoms).

We extensively researched the data related to the mpox death toll reports between 1970 and 2023 from various literature sources and websites. The findings are presented in Table S6, Table S7, and Figure S8 in Multimedia Appendix 1. Initially, high case fatality rates (CFRs) were predominantly observed before 1986: CFR=16.7% in 1970, 60% in 1972, 33.3% in 1973, 16.7% in 1975, 23.1% in 1978, and 9.7%-12.5% from 1981 to 1985. No death cases were reported from 1987 to 1995. Subsequently, CFRs remained below 6% between 1996 and 2022. The highest CFR occurred in 2016 (5.7%), followed by 2002 (4.2%), 2008 (4.2%), and 2001 (3.8%). In 2022, the CFR of mpox decreased to 0.09% but increased to 1.85% in 2023. The countries with the highest CFRs in 2023 were Belgium (33.33%), Peru (7.64%), and the United States (4.45%); see Figure S3c in Multimedia Appendix 1. To investigate the relationship between symptoms and death, we included 6 (9.8%) papers reporting outcome indicators related to death. The network analysis showed that rash (r=0.841), nausea/vomiting (r=0.832), conjunctivitis (r=0.832), and corneal opacity/photophobia (r=0.912) were significantly associated with mortality (P≤0.05); see Figure S4 in Multimedia Appendix 1.

**Discussion**

**Principal Findings**

The clinical symptoms of mpox have shown a decreasing prevalence over time, except for the specific symptom of rash, which has remained consistent. However, there has been an increasing concurrent prevalence of symptoms, resulting in a more complex correlation among them. The prevalence estimates of most symptoms were higher in Africa from 1970 to 2021, with about half of them occurring since 2003. The risks have shifted to Europe and the Americas since the 2022 outbreak, accompanied by increases in age and the proportions of males and MSM among the patients.

**General Evolution Trends**

An important highlight is that the presence of rash remained consistent in confirmed cases of monkeypox across all periods studied. This characteristic rash plays a vital role in the accurate diagnosis of poxvirus infections, emphasizing its significance as a key determinant in distinguishing monkeypox from other similar conditions [27]. A consistent and noteworthy trend observed throughout various periods is the continuous reporting of key symptoms, such as fever, lymphadenopathy, headache, cough, and sore throat, in cases of mpox. Although some of these symptoms have displayed a higher prevalence compared to others, they have remained steady since 2003 and decreased since the onset of the 2022 pandemic, which is roughly consistent with the previous 2 meta-analyses with pre- and post-2022 studies [21,22]. Especially in the pandemic since 2022, the prevalence of coreported symptoms was similar to other meta-analyses in 2022 [17,18]. As all symptoms were included in our study, it was observed for the first time that the number of symptoms has increased since 2003 and has remained steady throughout the worldwide pandemic. The decrease in the prevalence of nonspecific symptoms indicated an increase in asymptomatic cases, while new symptoms emerged, revealing a complex correlation among symptoms due to the evolution of mpox influenced by various factors.

**Geographical Differences**

The main reason for this evolution can be explained by differences in the geographical characteristics, population structure, virus source, transmission routes, and medical environment among the 3 periods. Before 2022, Africa emerged as the primary endemic region, exhibiting higher mpox prevalence and more reported symptoms. The epidemic trend can be divided into 2 periods, based on the characteristic patterns of symptom prevalence: before and after the virus spread to the Americas. In the first period (1970-2002), the patients with mpox all occurred in Midwest Africa, and children had the highest proportion among confirmed cases, which was because most of them had not been vaccinated against smallpox, resulting in a higher case prevalence. It was reported that severe symptoms of mpox tend to manifest more frequently in children rather than in adults [14]. As a consequence, fever, enlarged lymph nodes, and headache were reported with a higher prevalence and were closely associated with the risk of rash. Although mpox has been reported in non-African countries since 2003, the disease remains predominantly prevalent in central African nations. During this transitional period, there has been a gradual increase in the average age of patients and the proportion of male patients affected by monkeypox. Therefore, the polled prevalence of primary symptoms experienced a mild decrease from 2003 to 2021. However, the African patients still remained at higher risk of main symptoms and reported new symptoms, compared to nonendemic regions, such as Europe and the Western Pacific, while most nonendemic regions cases were imported from Africa. As the first outbreak outside Africa, cases in the Americas were mostly confirmed in 2003, with relatively higher reported symptom prevalence. Since the 2022 outbreak, Europe and the Americas emerged as 2 epidemic regions that reported a majority of diagnosed cases, with predominantly adult patients, resulting in a relatively favorable clinical presentation in patients. The prevalence of coreported symptoms exhibited consistent characteristics thorough the epidemic and nonendemic regions. Additionally, a higher number of confirmed cases and more comprehensive treatment modalities in endemic regions led to more than half of the newly reported symptoms, while the prevalence was generally lower.

**Clade Evolution**

The varying clades of the mpox virus could potentially exert a significant influence on symptom severities. Two major clades, the Congo Basin clade (clade I) and the West African clade (clade II), evolved from Africa and exhibited a potential impact on the clinical characteristics of patients with mpox. The Congo Basin clade was reported from 1970 to 2018 and is associated with severe disease and higher mortality [28]. The West African...
clade is associated with mild lower mortality and is divided into clade IIA and clade IIB [15,28,29], the former being reported only in the west of the Dahomey Gap from 1962 to 1971 and the United States in 2003 and the latter being reported in Nigeria from 1971 to 1978 and from 2017 to 2019 and worldwide since the 2022 outbreak [30]. Therefore, although the majority of the literature we included did not report branch data, it can be confirmed that a contributing factor to the significantly reduced severity of symptoms in 2022 was the lower virulence of clade IIB compared with clade I [31]. However, concerning the period before 2022, the impact of clades is rather intricate. For the African cases between 2003 and 2021, the prevalence remained unchanged or increased compared to pre-2003, while the prevalence of symptoms in other regions during the same period was relatively lower. For instance, the mpox outbreak in the United States in 2003 was caused only by clade IIA. Despite having a similar clinical characteristic to Africa, the prevalence of symptoms such as chills/rigors, dysphagia/difficulty swallowing, conjunctivitis, and shortness of breath was significantly lower. In addition, the disease has transitioned from being primarily animal-to-human transmitted to being human-to-human transmitted [10]. Animal-to-human transmission mostly occurred in 1970-2002. In nonendemic countries [13], few cases have been reported that were linked to international travel or the importation of animal-to-human monkeypox infection from 2003 to 2021 [32]. Mpox infection at this stage was not attributed to person-to-person contact [33]. In 2022, the majority of cases, with a few exceptions, were identified following instances of human-to-human sexual contact [34]. Reduction in symptom severities may depend on the viral clade [33] to better adapt to human-to-human transmission. Furthermore, favorable living environments and advanced medical conditions in high-income countries or regions may also explain the decrease in symptom prevalence since the 2022 outbreak.

Correlation Among Symptoms

Although the escalation in the number of comorbidities resulted in more intricate interconnections between symptoms, which mutually reinforced the correlation between them in our network diagram. This can be explained by more detailed symptom reports, which contribute to enhanced monitoring and advanced medical conditions. More importantly, a fact is that all symptoms are positively correlated, indicating that they are mutually reinforcing. Furthermore, the human-to-human transmission route has changed. Although HIV-related complications were noted during the monkeypox outbreak in Nigeria from 2017 to 2018, no substantiated evidence exists to indicate that the patients engaged in MSM behavior [35]. Subsequent to the outbreak in 2022, the studies incorporated in our analysis have consistently revealed a predominance of male patients, a substantial proportion of whom reported engaging in MSM behavior. These individuals also presented with coinfection, including HIV, syphilis, and various other STIs. Therefore, the mpox virus transmitted through MSM behavior has caused a rise in the number of additional symptoms, particularly an increase in the prevalence of symptoms associated with MSM behavior, such as perianal ulcer, pus or blood in stools, rectal bleeding, tenesmus/constipation, and dysuria. This may lead to complex response measures and treatment modalities.

Risks Factors of CFRs

All evidence suggests that the mpox virus is becoming mild, including a decrease in hospitalization rates in the reported literature [36] and the CFR in our collected data. The CFR is a complex outcome and not only is influenced by differences in symptoms caused by mpox clades but also is significantly correlated with patients’ age and local economic conditions [37]. This is one of the reasons why the CFR was higher before 2003. However, the CFR rebounded in 2023. This may be due to the lag in reporting death data compared to confirmed reports. However, the STI coinfection caused by sexual contact may have a synergistic effect with the mpox virus, increasing the severity of the disease.

Implications

Our findings are crucial for a comprehensive understanding of the mpox epidemic and for guiding an adequate policy from an epidemiological perspective. First, surveilling the evolving nature of mpox and the consequential changes in clinical characteristics is advised, especially in the epidemic regions, where the policy makers may shift their focus on the potential association between symptoms and the synergy risks to mortality. Furthermore, mpox can be influenced by a range of environmental and human factors. If appropriate measures are not taken to mitigate these factors, there is a heightened likelihood of a resurgence of the epidemic. Therefore, it is crucial to maintain ongoing vigilance and surveillance efforts, even in areas or periods when the disease appears to have been eliminated, in order to rapidly detect and respond to outbreaks and prevent the spread of the disease. However, targeted vaccination programs can be implemented for high-risk populations, including health care workers, laboratory workers, and those working with animals. Finally, efforts to raise public awareness, particularly within high-risk populations, such as MSM, should also be prioritized to help ensure timely detection and response to any potential outbreaks.

Limitations and Strengths

This systematic review and meta-analysis has several limitations. First, the clinical symptoms were only collected from available data, so the results may fail to represent all patients with mpox, especially those in low- and middle-income countries in early studies. The included studies from 1970 to 2002 were limited and of low quality. Especially after 1986, WHO terminated the surveillance program for mpox, which may have underestimated the severity of the symptoms in the results. Second, the symptoms reported during 3 study periods were likely to have bias due to the varying study designs of each year. In addition, the different strains of mpox and transmission routes may have had a direct relevance for how clinically severe the illness was during the study period, which requires further genome-wide analysis to confirm it. In future analyses, it is important to consider other sources of heterogeneity that were not fully documented before the 2022 outbreak, such as age, case determination, and comorbidities. Finally, we cannot confirm which symptoms pose a higher risk of mortality during the
worldwide outbreak since 2022, as limited data have simultaneously reported both the clinical presentation and the outcome of death; therefore, further evidence is needed to interpret this correlation.

Despite this, our review has carefully considered the changing trend of mpox symptoms since the first diagnosis in 1970 and also collected data on the number of confirmed cases of and deaths due to monkeypox in various countries over the years. It provided the evolution tendency of mpox, which can be used for making a related health policy. For example, we may improve treatment targeting for the common symptoms of mpox; according to demographic characteristics, we could suggest building an effective surveillance system for high-risk populations, especially in high-prevalence regions.

Conclusion
During the worldwide mpox outbreak, the clinical symptoms, including rash, resembled what they had been prior to the outbreak; however, some unique symptoms and changes in transmission route should be noted. The exponential increase in positive cases in Europe and the United States may help build relevant early-warning systems to prevent more mpox cases.

Acknowledgments
SS is supported by the National Natural Science Foundation of China (82304246), the Natural Science Foundation of Chongqing (CSTB2023NSCQ-MSX0198), the joint project of the Chongqing Health Commission and Science and Technology Bureau (2024QNXM056), and the Second Affiliated Hospital of Chongqing Medical University Research (grant 2023ITXG26).

Authors' Contributions
SS, MJ, and PH were responsible for the integrity of data and the accuracy of data analysis. Concept and design were managed by SS, HH, and PH. Acquisition, analysis, and interpretation of data were handled by all authors. SS and MJ drafted the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplementary files for publication.

References


14. Sklenovská N, Van Ranst M. Emergence of monkeypox as the most important orthopoxvirus infection in humans. Front Public Health 2018 Sep 4;6:241 [FREE Full text] [doi: 10.3389/fpubh.2018.00241] [Medline: 30234087]


25. Sklenovskyá N, Van Ranst M. Emergence of monkeypox as the most important orthopoxvirus infection in humans. Front Public Health 2018 Sep 4;6:241 [FREE Full text] [doi: 10.3389/fpubh.2018.00241] [Medline: 30234087]


31. Sklenovská N, Van Ranst M. Emergence of monkeypox as the most important orthopoxvirus infection in humans. Front Public Health 2018 Sep 4;6:241 [FREE Full text] [doi: 10.3389/fpubh.2018.00241] [Medline: 30234087]


**Abbreviations**

CFR: case fatality rate  
DRC: Democratic Republic of the Congo  
mpox: monkeypox  
MSM: men who have sex with men  
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
STI: sexually transmitted infection  
WHO: World Health Organization
Research Letter

U-Shaped Relationship Between Fibrinogen Level and 10-year Mortality in Patients With Acute Coronary Syndrome: Prospective Cohort Study

Yi ming Li¹, MBBS, MD; Yuheng Jia¹, MBBS, MD; Lin Bai¹, MBBS, MM; Bosen Yang¹, MBBS, MM; Mao Chen¹, MBBS, MM, MD; Yong Peng¹, MM, MD

Department of Cardiology, West China Hospital, Sichuan University, Sichuan Province, Chengdu City, China

Corresponding Author:
Yong Peng, MM, MD
Department of Cardiology
West China Hospital
Sichuan University
No.37 Guoxue Street
Sichuan Province, Chengdu City, 610041
China
Phone: 86 85423169
Fax: 86 85423169
Email: pengyong@scu.edu.cn

Abstract

This study demonstrated that fibrinogen is an independent risk factor for 10-year mortality in patients with acute coronary syndrome (ACS), with a U-shaped nonlinear relationship observed between the two. These findings underscore the importance of monitoring fibrinogen levels and the consideration of long-term anti-inflammatory treatment in the clinical management of patients with ACS.

(JMIR Public Health Surveill 2024;10:e54485) doi:10.2196/54485

KEYWORDS
fibrinogen; acute coronary syndrome; 10-year mortality; risk factor; coronary artery disease; myocardial; heart disease; inflammatory factor; retrospective study; Kaplan-Meier analysis; mortality; all-cause mortality; cubic-spline curve; regression model

Introduction

Identification of risk factors is essential in patients with coronary artery disease. As an important inflammatory factor and a key participant in coagulation, fibrinogen has attracted attention in the management of acute coronary syndrome (ACS). Previous research has also demonstrated that as an acute-phase protein, the levels of fibrinogen and its degradation products are higher in patients with COVID-19 than in the healthy population [1]. However, the relationship between fibrinogen levels and the prognosis of patients with ACS remains controversial [2]. Our previous study found that fibrinogen was an independent risk factor for mortality during the follow-up in patients with coronary artery disease [3]. In the ACS subgroup, no correlation was found between fibrinogen levels and mortality. In addition, the quantitative relationship between fibrinogen levels and the risk of mortality across a longer follow-up period needs further clarification.

Methods

Overview
We performed a prospective, large-scale, single-center study (Chinese Clinical Trial Registry ChiCTR 2100049313) to investigate the effect of fibrinogen levels on ACS prognoses. Patients were grouped according to the 5 quantile levels of plasma fibrinogen at admission. We used Kaplan-Meier analysis to estimate the cumulative incidence of all-cause mortality. The predictive value of fibrinogen levels for 10-year mortality was estimated by a Cox proportional hazards regression model. The impact of fibrinogen level on 10-year mortality was assessed by a restricted cubic spline (RCS) curve, which was derived from an adjusted Cox proportional hazards regression model.

Ethical Considerations
The study protocol was approved by the institutional review board of West China Hospital in accordance with the Declaration
of Helsinki (approval number 2012(243)). Written informed consent was obtained from all participants before enrollment in the study. Patients who come to the hospital for follow-up will be reimbursed for their transportation expenses.

**Results**

A total of 2434 consecutive patients with ACS who were admitted to West China Hospital in the Sichuan province of China, between December 10, 2010, and December 31, 2012, were enrolled in this study. After exclusion of 29 patients with in-hospital mortality, 2405 patients were included in the analysis. The mean age of the study population was 64.3 (SD 11.1) years. Plasma fibrinogen levels of these patients were normally distributed (skewness/kurtosis tests: $P < .001$), with a mean level of 3.41 (SD 1.06) g/L. The median follow-up time was 104 (IQR 98-111) months. In total, 387 (16.1%) patients died during the follow-up period. Kaplan-Meier analysis revealed that the cumulative incidence of mortality was significantly different among the 5 groups (log-rank test: $P < .001$; Figure 1A). Group 2 (fibrinogen level: 2.63-3.04 g/L) had the lowest incidence of mortality. In the Cox univariate regression analysis, fibrinogen was identified as a risk factor for 10-year mortality (hazard ratio [HR]: 1.24 [1.15-1.35]; $P < .001$). After adjusting for cardiovascular risk factors, fibrinogen was still found to be an independent risk factor (HR: 1.18 [1.08-1.29]; $P < .001$). The fibrinogen category in the multivariate Cox model, with group 2 serving as the referent, showed a nonlinear trend to mortality risk (Table S1 in Multimedia Appendix 1). Subgroup analysis suggested that in the ST-segment elevation myocardial infarction (STEMI)/non-STEMI (NSTEMI) subgroups, statistical results for fibrinogen showed the same trend (Table S2 in Multimedia Appendix 1).

**Discussion**

Acute-phase proteins are often associated with patient prognosis. Previous epidemiologic studies have suggested that elevated fibrinogen levels may mediate plaque rupture via prothrombotic or proinflammatory processes [2]. However, a lower fibrinogen level may indicate a worse general physical condition or may reveal hidden coagulation disorders in patients with ACS [4]. In this study, the long-term follow-up across nearly a decade and the quantitative RCS analysis suggested that fibrinogen levels at admission have a significant impact on the long-term prognosis of patients with ACS. The mechanisms underlying this phenomenon suggest that fibrinogen acts as both an acute-phase protein and a subclinical inflammatory indicator [5].

The U-shaped relationship between fibrinogen levels and 10-year mortality suggest the need for long-term anti-inflammatory or metabolism treatment, as well as the need to determine fibrinogen levels in patients with ACS. In addition, the nonlinear correlation between fibrinogen levels and the long-term mortality may restrict the effect of fibrinogen in some risk models that were derived from extended linear regression. The predictive value of fibrinogen needs to be estimated in nonlinear functions or by using machine learning classifiers in further studies.
Acknowledgments
MC and YP are co-corresponding authors of this paper. MC can be reached at hmaochen@vip.sina.com. This work was supported by the Sichuan Science and Technology Program (grant 2021YFS0330 awarded to YP), National Natural Science Foundation of China (grant 62306192 awarded to YL), and Natural Science Foundation of Sichuan Province (grant 2023NSFSC1638 awarded to YL).

Data Availability
The anonymized data set of this cohort can be requested by contacting the corresponding author via email.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Multivariate Cox model and subgroup analysis.
[DOCX File, 17 KB - publichealth_v10i1e54485_app1.docx]

References

Abbreviations
ACS: acute coronary syndrome
HR: hazard ratio
RCS: restricted cubic spline
STEMI: ST-segment elevation myocardial infarction

©Yi ming Li, Yuheng Jia, Lin Bai, Bosen Yang, Mao Chen, Yong Peng. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 07.06.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited.
cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Geographic Disparities in Access to Assisted Reproductive Technology Centers in China: Spatial-Statistical Study

Qingqing Zhou¹, MSc; Huatang Zeng², MSc; Liqun Wu², MSc; Kaichuan Diao², MSc; Rongxin He³, PhD; Bin Zhu¹, PhD

¹School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen, China
²Shenzhen Health Development Research and Data Management Center, Shenzhen, China
³Vanke School of Public Health, Tsinghua University, Beijing, China

Corresponding Author:
Bin Zhu, PhD
School of Public Health and Emergency Management
Southern University of Science and Technology
No 1088 Xueyuan Avenue, Nanshan District
Shenzhen, 518055
China
Phone: 86 0755 8801 5610
Email: zhub6@sustech.edu.cn

Abstract

A study on infertility in China found that while 543 health care institutions are approved for assisted reproductive technology (ART), only 10.1% offer all ART services, with a significant skew toward the eastern regions, highlighting the accessibility challenges faced by rural and remote populations; this study recommends government measures including travel subsidies and education initiatives to improve ART access for economically disadvantaged individuals.

(JMIR Public Health Surveill 2024;10:e55418) doi:10.2196/55418

KEYWORDS

assisted reproductive technology; spatial accessibility; travel time; travel cost; China

Introduction

Infertility is a growing, serious public health concern [1]. In 2023, the World Health Organization reported that infertility affects about 17.5% of the adult population globally [2]. Infertility is also prominent in China, affecting nearly 50 million people [3]. However, the travel time and costs are often a significant burden faced by patients with infertility [4,5]. Here we used a web path planning engine [6] to explore the spatial-economic disparities in access to assistive reproductive technology (ART) centers in mainland China.

Methods

Study Design

The National Health Commission of the People’s Republic of China supplied a list of ART centers, while population data were sourced from the 2020 Worldpop data set [7]. Residential point locations were gathered from the Gaode Maps open platform. After retaining one residential point within a 1-km radius, 57,469 residential points were acquired. We opted for a contemporary approach by leveraging real-time traffic data from web map navigation services—a departure from traditional methods. Using the Path Planning 2.0 algorithm, we determined optimal paths among residential points and the nearest ART center, predicting detailed travel times and corresponding costs [8,9]. We averaged the travel times and costs of residential points within the county, thus representing its overall level. Subsequently, isochronous maps depicting 1-hour and 2-hour travel times for ART services were generated.

Ethical Considerations

This study used deidentified publicly available data sets, all aggregated at the county level. Per Article 3 of the Southern University of Science and Technology’s institutional review board guidelines, we did not require peer review as this is a secondary analysis using publicly accessible locations of ART centers from the National Health Commission’s website and publicly accessible locations of residential points. These data sets contain no personally identifiable information and pose no risk of ethical violation.
Results

As of June 2022, in total, 543 health care institutions in Mainland China have received approved to conduct ART. However, only 55 (10.1%) institutions offer all 5 types of ART services. The distribution of ART centers in China predominantly favors the eastern plains and coastal regions (Figure 1A). Among 7 geographical subregions of China, East China boasts the highest number of ART centers (n=162, 29.8%). Specifically, Guangdong province leads with the greatest number of ART centers (n=56, 10.3%), with 83.9% (n=104) of the province’s counties accessible to ART centers within 1.5 hours and an average taxi cost of merely 66.7 CNY (US $9.21) [10]. In contrast, the Northwest region has the lowest number of ART centers (n=27, 5.0%). Tibet, in particular, has only 1 ART center, and a mere 12.2% (n=9) of Tibetan counties have access to it within 1.5 hours, with taxi costs soaring as high as 1485.2 CNY (US $205.08).

Figure 1. Travel time to assisted reproductive technology (ART) centers in China. (A) A map of travel times to ART centers at the county level. (B) The population density of China's counties covering 1-hour and 2-hour travel time isochrones. Counties with higher population densities are shaded in dark gray. Areas shaded in red are located within the 1-hour and 2-hour travel time isochrones of the ART centers.

Travel time and costs to ART centers by province are presented in Table 1. Detailed information on travel time and costs is visualized in Multimedia Appendix 1. In China, a substantial portion of the population faces challenges in accessing ART facilities within a short time frame. Specifically, 76.8% (n=180,330,033) of the total population and 63.5% (n=36,390) of residential points are not reachable to an ART center within an hour.
Table 1. Travel time and travel costs to assistive reproductive technology (ART) centers by province and proportion of population access to ART services at different time thresholds.

<table>
<thead>
<tr>
<th>Province</th>
<th>ART centers, n</th>
<th>Travel time (minutes)</th>
<th>Cost (CNY(^a))</th>
<th>Population (%)</th>
<th>≤30 minutes</th>
<th>≤60 minutes</th>
<th>≤120 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>18</td>
<td>20.8</td>
<td>52.2</td>
<td>80.7</td>
<td>97.7</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Tianjin</td>
<td>12</td>
<td>22.5</td>
<td>38.5</td>
<td>78.7</td>
<td>93.9</td>
<td>99.9</td>
<td></td>
</tr>
<tr>
<td>Hebei</td>
<td>31</td>
<td>43.0</td>
<td>93.1</td>
<td>44.3</td>
<td>75.5</td>
<td>94.1</td>
<td></td>
</tr>
<tr>
<td>Shanxi</td>
<td>12</td>
<td>45.0</td>
<td>112.1</td>
<td>49.2</td>
<td>65.1</td>
<td>96.7</td>
<td></td>
</tr>
<tr>
<td>Inner Mongolia</td>
<td>8</td>
<td>89.3</td>
<td>237.3</td>
<td>44.0</td>
<td>55.9</td>
<td>70.5</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heilongjiang</td>
<td>11</td>
<td>63.0</td>
<td>151.3</td>
<td>52.9</td>
<td>64.0</td>
<td>79.3</td>
<td></td>
</tr>
<tr>
<td>Jilin</td>
<td>9</td>
<td>72.4</td>
<td>153.3</td>
<td>40.0</td>
<td>52.3</td>
<td>77.3</td>
<td></td>
</tr>
<tr>
<td>Liaoning</td>
<td>19</td>
<td>47.5</td>
<td>90.3</td>
<td>44.1</td>
<td>69.4</td>
<td>94.9</td>
<td></td>
</tr>
<tr>
<td>East</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shanghai</td>
<td>20</td>
<td>17.0</td>
<td>49.9</td>
<td>86.9</td>
<td>99.5</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Jiangsu</td>
<td>33</td>
<td>38.0</td>
<td>71.6</td>
<td>42.9</td>
<td>83.8</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Zhejiang</td>
<td>27</td>
<td>40.0</td>
<td>90.1</td>
<td>35.5</td>
<td>81.6</td>
<td>99.8</td>
<td></td>
</tr>
<tr>
<td>Anhui</td>
<td>16</td>
<td>44.8</td>
<td>101.8</td>
<td>46.2</td>
<td>63.2</td>
<td>97.8</td>
<td></td>
</tr>
<tr>
<td>Jiangxi</td>
<td>18</td>
<td>50.4</td>
<td>126.9</td>
<td>41.7</td>
<td>59.0</td>
<td>96.1</td>
<td></td>
</tr>
<tr>
<td>Shandong</td>
<td>32</td>
<td>38.1</td>
<td>90.0</td>
<td>44.1</td>
<td>79.5</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Fujian</td>
<td>16</td>
<td>36.2</td>
<td>108.2</td>
<td>52.2</td>
<td>83.7</td>
<td>97.6</td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Henan</td>
<td>33</td>
<td>38.8</td>
<td>66.6</td>
<td>45.5</td>
<td>97.3</td>
<td>99.5</td>
<td></td>
</tr>
<tr>
<td>Hunan</td>
<td>24</td>
<td>52.0</td>
<td>121.0</td>
<td>39.2</td>
<td>59.8</td>
<td>95.6</td>
<td></td>
</tr>
<tr>
<td>Hubei</td>
<td>32</td>
<td>39.5</td>
<td>68.2</td>
<td>49.5</td>
<td>72.1</td>
<td>98.6</td>
<td></td>
</tr>
<tr>
<td>Southern</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guangdong</td>
<td>56</td>
<td>27.7</td>
<td>66.7</td>
<td>67.6</td>
<td>91.2</td>
<td>99.5</td>
<td></td>
</tr>
<tr>
<td>Guangxi</td>
<td>21</td>
<td>42.3</td>
<td>102.0</td>
<td>53.0</td>
<td>65.1</td>
<td>98.0</td>
<td></td>
</tr>
<tr>
<td>Hainan</td>
<td>10</td>
<td>58.3</td>
<td>246.1</td>
<td>50.9</td>
<td>60.0</td>
<td>75.1</td>
<td></td>
</tr>
<tr>
<td>Southwest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chongqing</td>
<td>12</td>
<td>44.9</td>
<td>153.0</td>
<td>44.9</td>
<td>68.4</td>
<td>95.6</td>
<td></td>
</tr>
<tr>
<td>Guizhou</td>
<td>13</td>
<td>46.2</td>
<td>99.9</td>
<td>49.1</td>
<td>66.1</td>
<td>94.4</td>
<td></td>
</tr>
<tr>
<td>Sichuan</td>
<td>14</td>
<td>50.5</td>
<td>140.2</td>
<td>44.6</td>
<td>70.5</td>
<td>91.7</td>
<td></td>
</tr>
<tr>
<td>Yunnan</td>
<td>18</td>
<td>69.8</td>
<td>199.2</td>
<td>44.4</td>
<td>59.1</td>
<td>77.3</td>
<td></td>
</tr>
<tr>
<td>Tibet</td>
<td>1</td>
<td>267.9</td>
<td>1485.2</td>
<td>47.3</td>
<td>53.0</td>
<td>58.2</td>
<td></td>
</tr>
<tr>
<td>Northwest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ningxia</td>
<td>2</td>
<td>63.2</td>
<td>119.3</td>
<td>41.9</td>
<td>61.1</td>
<td>81.6</td>
<td></td>
</tr>
<tr>
<td>Qinghai</td>
<td>2</td>
<td>82.8</td>
<td>190.8</td>
<td>60.2</td>
<td>77.7</td>
<td>85.9</td>
<td></td>
</tr>
<tr>
<td>Gansu</td>
<td>4</td>
<td>131.9</td>
<td>157.1</td>
<td>29.5</td>
<td>32.6</td>
<td>49.3</td>
<td></td>
</tr>
<tr>
<td>Shaanxi</td>
<td>10</td>
<td>64.6</td>
<td>179.2</td>
<td>38.5</td>
<td>58.4</td>
<td>79.3</td>
<td></td>
</tr>
<tr>
<td>Xinjiang</td>
<td>9</td>
<td>108.4</td>
<td>275.5</td>
<td>56.3</td>
<td>61.5</td>
<td>67.1</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)CNY=US $0.14. Travel cost was determined by the local taxi fare, which varies depending on the location. The first price covers a specified distance, often 2.5-3 km, and every additional kilometer is then calculated at a certain price.
Discussion

The distribution of ART centers in China exhibits significant disparities. A higher concentration of ART centers is observed in urban and eastern regions, while individuals in northwestern and rural areas encounter prolonged travel times and elevated transportation costs when seeking ART treatment. The map serves the dual purpose of estimating the likelihood of individuals seeking ART treatment when needed and providing an evidence-based foundation for efficient allocation of limited ART resources to underserved populations (present and future). The government is suggested to implement a series of measures, including counterpart aid and effective initiatives to educate and recruit ART doctors in disadvantaged units. Residents in rural and remote areas contend with extended travel times and substantial travel costs when accessing ART services, which should be covered fully or partially through travel subsidies or paid leaves. The introduction of telehealth services is a viable solution to surmount these barriers for patients residing in remote areas, effectively reducing in-person office visits. Notably, ensuring privacy is paramount when helping patients seeking ART treatments. This study bears limitations. While we used the best available data, the residential point data sets remain susceptible to omission errors, which confines the algorithm’s usage to only 1 mode of transportation—a limitation not potentially aligning with real-world scenarios. Additionally, individuals may not necessarily receive ART treatment at the nearest facility or may opt for alternative transportation means.

Acknowledgments

This study is funded by the Guangdong Philosophy and Social Sciences Planning Project (GD22YGL10), Shenzhen Philosophy and Social Sciences Planning Project (SZ2023C007), and Shenzhen Science and Technology Program (JCYJ20220530113208019). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Authors’ Contributions

BZ had full access to all of the study data and takes responsibility for the integrity of the data and the accuracy of the data analysis. BZ conceptualized and designed the study; acquired the data; obtained the funding; provided administrative, technical, or material support; and supervised the study. QZ carried out the statistical analysis and drafted the manuscript. QZ and BZ interpreted the data. HZ, LW, KD, RH, and BZ critically revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Travel time and cost analyses.

References


**Abbreviations**

ART: assisted reproductive technology
The Effectiveness of National Expanded Program on Immunization With Hepatitis A Vaccines in the Chinese Mainland: Interrupted Time-Series Analysis

Bo-Wen Ming1*, BSc; Li Li1*, PhD; Hao-Neng Huang1, BSc; Jia-Jun Ma1, BSc; Chen Shi1, BSc; Xiao-Han Xu1, BSc; Zhou Yang1, BSc; Chun-Quan Ou1, PhD

State Key Laboratory of Organ Failure Research, Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China

*these authors contributed equally

Corresponding Author:
Chun-Quan Ou, PhD
State Key Laboratory of Organ Failure Research, Department of Biostatistics
School of Public Health
Southern Medical University
No1023, South Shatai Road, Baiyun District
Guangzhou, 510515
China
Phone: 86 020 61649461
Email: ou Chunquan@hotmail.com

Abstract

Background: The high prevalence of hepatitis A delivered a blow to public health decades ago. The World Health Organization (WHO) set a goal to eliminate viral hepatitis including hepatitis A by 2030. In 2008, hepatitis A vaccines were integrated into the Expanded Program on Immunization (EPI) in China to alleviate the burden of hepatitis A, although the effectiveness of the EPI has not been well investigated.

Objective: We aimed to evaluate the intervention effect at both provincial and national levels on the incidence of hepatitis A in the Chinese mainland from 2005 to 2019.

Methods: Based on the monthly reported number of hepatitis A cases from 2005 to 2019 in each provincial-level administrative division, we adopted generalized additive models with an interrupted time-series design to estimate province-specific effects of the EPI on the incidence of hepatitis A among the target population (children aged 2-9 years) from 2005 to 2019. We then pooled province-specific effect estimates using random-effects meta-analyses. We also assessed the effect among the nontarget population and the whole population.

Results: A total of 98,275 hepatitis A cases among children aged 2-9 years were reported in the Chinese mainland from 2005 to 2019, with an average annual incidence of 5.33 cases per 100,000 persons. Nationally, the EPI decreased the hepatitis A incidence by 80.77% (excess risk [ER] –80.77%, 95% CI –85.86% to –72.92%) during the study period, guarding an annual average of 28.52 (95% empirical CI [eCI] 27.37-29.00) cases per 100,000 persons among the target children against hepatitis A. Western China saw a more significant effect of the EPI on the decrease in the incidence of hepatitis A among the target children. A greater number of target children were protected from onset in Northwest and Southwest China, with an excess incidence rate of –129.72 (95% eCI –135.67 to –117.86) and –66.61 (95% eCI –67.63 to –64.22) cases per 100,000 persons on average, respectively. Intervention effects among nontarget (ER –32.88%, 95% CI –39.76% to –25.21%) and whole populations (ER –31.97%, 95% CI –39.61% to –23.37%) were relatively small.

Conclusions: The EPI has presented a lasting positive effect on the containment of hepatitis A in the target population in China. The EPI’s effect on the target children also provided a degree of indirect protection for unvaccinated individuals. The continuous surveillance of hepatitis A and the maintenance of mass vaccination should shore up the accomplishment in the decline of hepatitis A incidence to ultimately achieve the goal set by the WHO.

(JMIR Public Health Surveill 2024;10:e53982) doi:10.2196/53982
KEYWORDS

hepatitis A; incidence; Expanded Program on Immunization; vaccine; interrupted time series; intervention; China

Introduction

Hepatitis A is an acute infectious disease that is transmitted via a fecal-oral route, primarily through the ingestion of food or water contaminated with the virus [1-3]. Hepatitis A would frequently give rise to sporadic cases, outbreaks, and even epidemics [4,5], which posed threats to public health and laid an excessive economic burden on society. Globally, the burden of hepatitis A has been increasing in recent years [6], accounting for about 1.4 disability-adjusted life years (DALYs) and 2.3 DALYs in 2010 and 2019, respectively [7,8]. Given the current global burden of hepatitis A viruses (HAV), the World Health Organization (WHO) developed a long-term goal of eliminating viral hepatitis, including hepatitis A, as a major public health threat by 2030 [9].

Vaccination is an essential intervention for the prevention and control of hepatitis A. Assessing the effectiveness of vaccination on the incidence of hepatitis A is a critical step to the continuous implementation and improvement of the immunization program. Although some studies have investigated the correlation between vaccination coverage and hepatitis A incidence [10] or seroprevalence [11,12], there are limited data on the intervention effect of hepatitis A mass vaccination. A few studies simply described and compared the average yearly incidence between the pre- and postintervention periods in the Chinese mainland and Israel [4,13-17], which neglected potential confounding factors. Two studies used regression models to evaluate the association of the vaccination program with the incidence of hepatitis A in the Chinese mainland and Brazil [17,18]. However, these studies did not consider the potential regional differences in the effect of the vaccination, which would fail to promote precise and differentiated strategies in the prevention and control of hepatitis A. Another research study assessed the impact of the Expanded Program on Immunization (EPI) on the incidence of 11 childhood vaccine-preventable diseases, including hepatitis A, with annual data in the Chinese mainland through a quasi-Poisson regression and a polynomial regression [19]. However, this study constructed the model based on national data of the whole population and assumed that all changes in the hepatitis A incidence from 2003 to 2012 were fully attributable to vaccination, thus failing to assess the impact of the EPI in different age groups. Furthermore, the potential spatial and time-varying effects of the EPI on hepatitis A remained unclear.

In the Chinese mainland, hepatitis A was endemic in most regions from 1990 to 2007, with an average annual incidence of 21.31 cases per 100,000 persons, in the absence of systematic hepatitis A inoculation [20]. Furthermore, children under the age of 10 years were subjected to the highest incidence among all age groups from 1990 to 2007, and public health emergencies of hepatitis A occurred mostly in elementary schools in Western China [2,20]. The Chinese government integrated hepatitis A vaccines into the Chinese National EPI in 2008 to mitigate hepatitis A. Children aged over 18 months are eligible to either (1) receive 1 dose of the live attenuated hepatitis A vaccine or (2) receive 1 dose of the inactivated vaccine at the 18th month and another dose at the 24th month for free in the Chinese mainland [10]. However, the initiation time and the intensity of the EPI were different in each provincial-level administrative division (PLAD). Hence, it is necessary to explore the effect of the EPI in different PLADs and then pool the effect estimates to obtain substantial evidence. Herein, we aimed to assess the effectiveness of the EPI in China at the provincial and national levels.

Methods

Data Collection

We collected monthly data on hepatitis A incidence across 31 PLADs from 2005 to 2019 in the Chinese mainland from the National Population and Health Science Data Sharing Platform of the Chinese Center for Disease Control and Prevention [21]. A hepatitis A case is defined as a patient with acute gastrointestinal symptoms (eg, jaundice) and elevated serum alanine aminotransferase who tests positive for the immunoglobulin M antibody to HAV [22].

The provincial-level annual data (ie, the age-specific resident population, gross domestic product [GDP] per capita, urbanization rates, the number of hospitalization beds per 1000 persons, the proportion of children younger than 14 years old, and illiteracy rates) from 2005 to 2019 were compiled from China Statistical Yearbooks [23,24]. The monthly resident population was derived from linear interpolation. Monthly data on ambient average temperature were downloaded from the China Meteorological Data Sharing Service System [25].

Data Analysis

The exact implementation time of the EPI and the use of the vaccine type (ie, inactivated hepatitis A vaccines or live attenuated hepatitis A vaccines) were different among PLADs. Accordingly, the postintervention period of each PLAD was specified separately (Table S1 in Multimedia Appendix 1). Children born after 2007 and aged >24 months are eligible to receive hepatitis A vaccines. The postintervention period spanned 11 years. Children younger than 13 years old could directly benefit from the program at the end of the study period. However, the raw data were in the form of 1-year age groupings for those aged <10 years and 5-year age groupings after the age of 10 years (10-14 years, 15-19 years, etc). Hence, we identified children aged 2-9 years whom the EPI had a direct bearing on as the study population.

A 2-stage analytic strategy was used to evaluate the effect of the EPI on the incidence of hepatitis A. In the first stage, a quasi-Poisson regression model with interrupted time series was established to capture the province-specific effectiveness of the EPI after adjusting for covariates in each of the 30 PLADs [26]. Tianjin was not considered because the average annual number of cases was extremely low even before the EPI (3.55 cases per year; Table S2 in Multimedia Appendix 1), which cannot ensure the stability of modeling. The model mainly
contains an elapsed time since the start of the study, a dummy variable indicating the pre- or postintervention period, and an interaction term for the aforementioned 2 variables. Furthermore, we included the natural cubic spline of monthly average temperature with 3 df, which was selected by minimizing the value of quasi–Akaike information criterion (Table S3 in Multimedia Appendix 1) and a categorical variable of calendar months to control for the potential seasonality of hepatitis A incidence. We incorporated the reported public health emergencies of hepatitis A, mainly occurring before the implementation of the EPI, into the model to avoid the overestimation of the effects (Table S4 in Multimedia Appendix 1) [27]. An autoregressive term of residuals at lag 1 was taken into account to adjust for autocorrelation if necessary (Table S5 and Figures S1 and S2 in Multimedia Appendix 1). Detailed information on the provincial-specific model is provided in the Supplemental Methods in Multimedia Appendix 1.

The provincial-specific excess risk (ER) of the hepatitis A incidence and the corresponding 95% CI were estimated. In addition, the excess incidence rate (EIR) in each province and the corresponding 95% empirical CI (eCI) were estimated to investigate the change in the hepatitis A incidence due to the implementation of the EPI. The details on the estimation of ERs and EIRs are demonstrated in the Supplemental Methods in Multimedia Appendix 1.

In the second stage, we pooled the province-specific effect of the EPI using random-effects meta-analyses to derive the regional-level and national-level estimates of the ER. The heterogeneity was quantified by I² statistic and Cochran Q test [28]. In addition, we estimated the national population-weighted EIRs.

Furthermore, we performed subgroup analyses. The PLADs were classified into 7 regions (ie, Northeast, North, Northwest, East, Center, South, and Southwest), considering different geographical and environmental conditions [29]. We also estimated the EIR for each subgroup of PLADS by socioeconomic factors (ie, urbanization rates, GDP per capita, the number of hospitalization beds per 1000 persons, the proportion of children aged 0-14 years, and illiteracy rates), the level of average incidence of hepatitis A before the implementation of EPI, and the types of hepatitis A vaccines (Table S1 in Multimedia Appendix 1). All factors, except for regions and types of hepatitis A vaccines, were classified into 2 levels using the medians as the cutoffs (high: ≥median; low: <median).

In addition, we explored the effect of the EPI on the whole population and the nontarget population (children younger than the age of 2 years or older than the age of 10 years), as previous literature suggested that the vaccination might develop herd immunity against hepatitis A [10]. We did not consider the interaction term between time and the indicator variable of the intervention in models for the nontarget population and the whole population because the pooling effects of the interaction term fell short of statistical significance.

We performed sensitivity analyses to examine the robustness of the main model results (Supplemental Methods in Multimedia Appendix 1). First, we substituted calendar months with a natural cubic spline function with 3 df to consider different ways of controlling potential seasonality. Second, we explored the potentially nonlinear intervention effect over time. Finally, we took the transition period from the implementation of the EPI to 2010 into account and excluded this period from the analyses. A 2-sided P value of <.05 was considered statistically significant. We used the R statistical software (version 4.2.1; R Foundation for Statistical Computing) to perform all analyses.

Ethical Consideration
This study was approved by the Research Ethics Committee of Southern Medical University (NFYKDX-ER2022012). The need for informed consent was waived because the data were deidentified and aggregated.

Results

Temporal and Spatial Distribution of Hepatitis A Incidence
Table 1 presents the summary statistics of hepatitis A in different age groups from 2005 to 2019 in the Chinese mainland. During the study period, there was a total of 98,275 reported hepatitis A cases among children aged 2-9 years with an average annual incidence of 5.33 cases per 100,000 persons. Overall, there was a gradual decline in the annual incidence of hepatitis A. The incidence of hepatitis A increased from 10.98 cases per 100,000 persons in 2005 to 16.80 cases per 100,000 persons in 2007, but the incidence had considerably dropped since 2008 and was only 0.56 cases per 100,000 persons in 2019. Among the whole population, 556,737 cases were reported from 2005 to 2019, but the incidence had considerably dropped since 2008 and was only 0.56 cases per 100,000 persons in 2019. Among the whole population, 556,737 cases were reported from 2005 to 2019, with an average annual 2.79 cases per 100,000 persons. The incidence among the whole population and the nontarget population also appeared to be edging downward, with the latter leveling off at 1.50 cases per 100,000 persons after 2011.
Table 1. The annual number of cases and hepatitis A incidence in the Chinese mainland.

<table>
<thead>
<tr>
<th>Year</th>
<th>Target population&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Nontarget population</th>
<th>Whole population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases, n</td>
<td>Incidence (cases per 100,000 persons)</td>
<td>Cases, n</td>
</tr>
<tr>
<td>2005&lt;sup&gt;b&lt;/sup&gt;</td>
<td>13,375</td>
<td>10.98</td>
<td>59,966</td>
</tr>
<tr>
<td>2006</td>
<td>13,602</td>
<td>11.29</td>
<td>55,065</td>
</tr>
<tr>
<td>2007</td>
<td>20,325</td>
<td>16.80</td>
<td>56,810</td>
</tr>
<tr>
<td>2008</td>
<td>11,186</td>
<td>9.15</td>
<td>44,866</td>
</tr>
<tr>
<td>2009</td>
<td>7958</td>
<td>6.41</td>
<td>35,883</td>
</tr>
<tr>
<td>2010</td>
<td>7317</td>
<td>5.66</td>
<td>27,960</td>
</tr>
<tr>
<td>2011</td>
<td>4969</td>
<td>4.22</td>
<td>26,487</td>
</tr>
<tr>
<td>2012</td>
<td>3598</td>
<td>2.95</td>
<td>20,855</td>
</tr>
<tr>
<td>2013</td>
<td>2852</td>
<td>2.33</td>
<td>19,392</td>
</tr>
<tr>
<td>2014</td>
<td>4604</td>
<td>3.68</td>
<td>21,365</td>
</tr>
<tr>
<td>2015</td>
<td>2763</td>
<td>2.17</td>
<td>19,904</td>
</tr>
<tr>
<td>2016</td>
<td>2449</td>
<td>1.91</td>
<td>18,836</td>
</tr>
<tr>
<td>2017</td>
<td>1554</td>
<td>1.20</td>
<td>17,321</td>
</tr>
<tr>
<td>2018</td>
<td>1017</td>
<td>0.77</td>
<td>15,179</td>
</tr>
<tr>
<td>2019</td>
<td>706</td>
<td>0.56</td>
<td>18,565</td>
</tr>
<tr>
<td>Total</td>
<td>98,275</td>
<td>5.33</td>
<td>458,454</td>
</tr>
</tbody>
</table>

<sup>a</sup>The target population is children aged 2-9 years.

<sup>b</sup>There were 8 cases whose ages were missing.

Figure 1 demonstrates the spatial distribution of the average hepatitis A incidence among children aged 2-9 years and the nontarget population. Generally, Western China witnessed a higher incidence of hepatitis A than that of other regions during both the pre- and postintervention periods. However, this regional difference has remarkably dwindled since the implementation of the EPI (Table S6 in Multimedia Appendix 1). The hepatitis A incidence among children aged 2-9 years showed significant seasonal variation, which peaked in the autumn and winter. However, the seasonality was gradually ambiguous, and the peak slumped greatly in the postintervention stage (Figures S3 and S4 in Multimedia Appendix 1). The seasonality pattern of Western China was similar, but the seasonality was not apparent in other regions.
Effect of EPI on Hepatitis A Incidence

Overall, the implementation of the EPI led to a significant decrease in the incidence of hepatitis A for the target population, with an average ER of 80.77\% (95\% CI 72.92\%–85.86\%) among children aged 2-9 years (Figure 2 and Figure S5 in Multimedia Appendix 1). The magnitude of the EPI intervention enhanced with each passing year from the initiation of the EPI. Specifically, the intervention impact was relatively weak in the first year after the implementation of the EPI, associated with a 51.92\% (95\% CI 42.44\%–59.84\%) reduction in the incidence of hepatitis A. The reduction dramatically rose to 97.25\% (95\% CI 92.77\%–98.95\%) in the 11th year following the EPI (Table S7 in Multimedia Appendix 1).

Figure 1. Spatial distribution of the average annual hepatitis A incidence. White regions indicate missing data. (A) Preintervention period among children aged 2-9 years; (B) postintervention period among children aged 2-9 years; (C) preintervention period among nontarget individuals; (D) postintervention period among nontarget individuals.
Figure 2. Excess risks and excess incidence rates towards children aged 2-9 years. The line represents the point estimates of excess risks and excess incidence rates of hepatitis A associated with the EPI. The colored regions are the corresponding 95% CIs and 95% empirical CIs, respectively. The excess risks of Northeast China are not statistically significant during the overall period after the implementation of the EPI. (A) Excess risks; (B) excess incidence rates. EPI: Expanded Program on Immunization.

Table 2 and Tables S8 and S9 in Multimedia Appendix 1 show the EPI-associated reduction in cases and the incidence of hepatitis A among children aged 2-9 years. In general, it was estimated that a yearly average of 28.52 (95% eCI 27.37-29.00) hepatitis A cases per 100,000 persons was averted within 11 years of intervention. In other words, on average, about 34,618 cases are prevented each year by the EPI. In detail, the EPI protected an average of 14.31 (95% eCI 13.50-14.78) cases per 100,000 persons from hepatitis A in the first 5 years of the intervention period. In the subsequent 6-11 years, on average, 39.90 (95% eCI 38.34-40.48) cases were prevented per 100,000 individuals.
Table 2. Annual average excess number of cases and excess incidence rate of hepatitis A in target children.

<table>
<thead>
<tr>
<th>Variable and region</th>
<th>1-5 years after the EPI&lt;sup&gt;a&lt;/sup&gt;</th>
<th>6-11 years after the EPI</th>
<th>1-11 years after the EPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess number of case (95% empirical CI; 1000 cases)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese mainland</td>
<td>–16.99 (–17.55 to –16.03)</td>
<td>–49.31 (–50.03 to –47.38)</td>
<td>–34.62 (–35.20 to –33.22)</td>
</tr>
<tr>
<td>Northeast</td>
<td>–0.01 (–0.01 to 0.03)</td>
<td>–0.01 (–0.01 to 0.10)</td>
<td>–0.01 (–0.01 to 0.07)</td>
</tr>
<tr>
<td>North</td>
<td>–0.20 (–0.22 to –0.16)</td>
<td>–0.23 (–0.24 to –0.15)</td>
<td>–0.21 (–0.23 to –0.16)</td>
</tr>
<tr>
<td>Northwest</td>
<td>–6.34 (–6.72 to –5.77)</td>
<td>–18.51 (–19.31 to –16.79)</td>
<td>–12.98 (–13.57 to –11.79)</td>
</tr>
<tr>
<td>East</td>
<td>–1.76 (–1.84 to –1.65)</td>
<td>–7.77 (–7.79 to –7.69)</td>
<td>–5.04 (–5.08 to –4.95)</td>
</tr>
<tr>
<td>Center</td>
<td>–2.35 (–2.62 to –1.96)</td>
<td>–2.95 (–2.99 to –2.88)</td>
<td>–2.68 (–2.82 to –2.46)</td>
</tr>
<tr>
<td>South</td>
<td>–0.20 (–0.24 to –0.13)</td>
<td>–0.57 (–0.60 to –0.48)</td>
<td>–0.40 (–0.43 to –0.32)</td>
</tr>
<tr>
<td>Southwest</td>
<td>–6.14 (–6.46 to –5.55)</td>
<td>–19.27 (–19.41 to –18.79)</td>
<td>–13.30 (–13.50 to –12.82)</td>
</tr>
</tbody>
</table>

| Excess incidence rate (95% empirical CI; cases per 100,000 persons) | | | |
| Northeast | –0.07 (–0.20 to 0.37) | –0.08 (–0.19 to 1.49) | –0.07 (–0.19 to 0.96) |
| North | –1.47 (–1.64 to –1.18) | –1.58 (–1.67 to –1.06) | –1.53 (–1.65 to –1.14) |
| Northwest | –63.25 (–67.06 to –57.55) | –185.28 (–193.32 to –168.03) | –129.72 (–135.67 to –117.86) |
| Center | –11.95 (–13.32 to –9.96) | –13.35 (–13.50 to –13.00) | –12.75 (–13.42 to –11.71) |
| South | –1.23 (–1.47 to –0.78) | –3.53 (–3.68 to –2.92) | –2.49 (–2.68 to –1.97) |
| Southwest | –30.01 (–31.61 to –27.12) | –98.50 (–99.25 to –96.05) | –66.61 (–67.63 to –64.22) |

<sup>a</sup>EPI: Expanded Program on Immunization.

There were significant differences among provinces in the effect of the EPI intervention on the incidence of hepatitis A among children aged 2-9 years ($I^2=92.80\%; P<.001$). We observed significant effects in Northwest (ER –99.02%, 95% CI –99.69% to –96.87%) and Eastern (ER –99.29%, 95% CI –99.90% to –95.12%) China during the whole intervention period (Figure 2 and Table S10 in Multimedia Appendix 1), while the ER in Northeast China was statistically nonsignificant. Table 2 and Tables S8 and S9 in Multimedia Appendix 1 reflect the moderate effectiveness of the intervention on the target children in Northern, Central, and Southern China. Compared with other regions, a greater number of the target children were protected in Northwest and Southwest China, with an EIR of –129.72 (95% eCI –135.67 to –117.86) and –66.61 (95% eCI –67.63 to –64.22) cases per 100,000 persons on average, respectively. Figure 3 and Table S9 and Figure S6 in Multimedia Appendix 1 provide ERs and EIRs of the EPI for different subgroups based on socioeconomic factors and types of hepatitis A vaccines. The ERs of hepatitis A incidence in the intervention population did not differ greatly at different socioeconomic levels and vaccine types. However, EIRs exhibited greater differences between subgroups than ERs. PLADs with less GDP per capita, lower urbanization rates, insufficient number of hospitalization beds per 1000 persons, higher incidence of hepatitis A before the implementation of the EPI, larger proportion of children younger than 14 years old, higher illiteracy rates, or live attenuated hepatitis A vaccines witnessed a more significant alleviation in the incidence of hepatitis A during the whole intervention period.
In view of all people and nontarget people, the EPI also lessened the burden of hepatitis A (Tables S11 and S12 in Multimedia Appendix 1). Generally, the EPI was likewise linked to the decrease in the incidence in nontarget individuals (ER –32.88%, 95% CI –39.76% to –25.21%) and all people (ER –31.97%, 95% CI –39.61% to –23.37%). The EPI could prevent 0.78 (95% eCI 0.73-0.82) and 0.83 (95% eCI 0.78-0.88) cases per 100,000 persons among nontarget individuals and all people from infection within 11 years of the intervention, respectively.

**Sensitivity Analyses**

We obtained similar results from the sensitivity analyses (Table S7 and Figures S7-9 in Multimedia Appendix 1). First, the effect of the EPI on the incidence did not change substantially when using a natural cubic spline to control for seasonality (ER –80.25%, 95% CI –85.66% to –71.68%). Second, in accordance with the main results, the exposure-response curves revealed that the logarithm transformation of ER in large changed linearly with time. Finally, the results, in consideration of the transition period, were similar to those in the main analysis, with an average ER of 80.45% (95% CI 71.19%-86.48%).

**Discussion**

**Principal Findings**

In this nationwide study, we comprehensively assessed the effectiveness of the EPI on the incidence of hepatitis A among children aged 2-9 years based on province-specific monthly data from 2005 to 2019. The 97% decrease in hepatitis A incidence attributable to the 11-year intervention of the EPI among children aged 2-9 years provides an estimate of the vaccine-preventable HAV cases in childhood, suggesting that the burden of the hepatitis A cyclical epidemic before the EPI has been largely alleviated [10]. By the end of 2019, the incidence of hepatitis A decreased to 0.56 cases per 100,000 persons among target children in the Chinese mainland, which gradually approaches the target of eliminating hepatitis A set by the WHO.

Two previous studies have shown the beneficial effectiveness of the EPI in curbing the incidence of hepatitis A in the Chinese mainland [10,17]. Sun et al [10] described an 82.3% decline in the incidence in 2016 compared to that in 2004. Tang et al [17] indicated that 47% of the decrease in the incidence of hepatitis A among children aged 0-14 years was attributable to the EPI, by simply fitting the annual number of cases using an interrupted linear regression without considering the nonlinear trend and the change in the population size, which might bring about bias in the estimates of the intervention effect. A previous review in China demonstrated that seroconversion rates were well above 80% and vaccine efficacy rates were higher than 90% in vaccinated individuals, whether with live attenuated hepatitis A vaccines or inactivated hepatitis A vaccines [5]. A randomized clinical trial indicated that nearly 100% of vaccinated people developed serum antibodies against HAV after the inoculation [30]. In this population-based interrupted time-series study, we used a robust 2-stage modeling strategy to comprehensively assess the province-specific effect and the pooled national effect. Combined with the previous evidence, this study could corroborate that the EPI did weigh positively on the decrease of hepatitis A incidence among the vaccinated population.

In addition, there were distinct regional differences in the effect of the EPI, with the greatest effect in Western China. The effects on the target children were not statistically significant in several PLADs of Eastern and Northern China where the incidence was low prior to the intervention. A possible explanation may be that a portion of schoolchildren had received self-financed vaccination against hepatitis A before the implementation of

---

**Figure 3.** Average excess incidence rates in different subgroups among children aged 2-9 years. eCI: empirical CI; EIR: excess incidence rate.
the EPI due to better economic conditions and awareness of vaccination in these PLADs compared to Western China [10,31]. The annual number of hepatitis A cases in Tianjin fluctuated between 0 and 6 cases among the target children during the study period from 2005 to 2019. One possible reason was that the local immunization strategy was implemented in 2001, covering 1-year-old infants, first-grade children, and seventh-grade children [31,32]. The hepatitis A vaccine coverage was above 90% in 2001, which facilitated a low-level sporadic state of hepatitis A in Tianjin among children [31,32]. Notably, the postintervention incidence of hepatitis A in Western China, especially in Tibet and Xinjiang, was still significantly higher than that of other regions of China, although the EPI had largely alleviated the local burden of hepatitis A. These discrepancies might be related to impoverishment [24] and the lack of vaccination awareness. For example, the vaccination coverage in children was only about 60% in Tibet in 2004, and about 71% of towns in Xinjiang achieved the goal of 90% coverage of hepatitis A vaccine in 2013 [10,33,34]. A sampling survey found a lower anti-HAV antibody seroprevalence in Western China than in Eastern China (69% vs 72% in 2014) [11]. None of the local health authorities could afford to overlook the gaps in the vaccination coverage among children between the Western region (69.1%) and other regions of China (86% in Eastern China and 74.5% in Middle China) [33]. Further promotion of vaccination and prevention measures against hepatitis A in Western China, including raising public awareness about the prevention of hepatitis A and the efficacy of vaccination, remains a long way off to narrow the gaps. However, there is no denying that the massive disparities in the geographical distribution of the hepatitis A incidence before the intervention, with the vast majority of cases in the Western region, have faded away in most PLADs as the intervention progressed. Overall, focusing on the reduction of HAV exposure and health education on vaccination against hepatitis A among local children in high-prevalence regions will enhance the effectiveness of the intervention and thereby lighten the burden of hepatitis A.

Subgroup analyses indicated that there is no significant difference in reducing the relative risk of hepatitis A onset between regions with different socioeconomic conditions and types of hepatitis A vaccines. However, it is worth noting that the decrease in the incidence of hepatitis A due to the implementation of the EPI varied evidently across PLADs with different socioeconomic levels and types of vaccines. Specifically, the alleviation of the hepatitis A burden in resource-limited regions was much higher than that in resource-rich regions, which are mainly concentrated in Eastern China. Similarly, greater effectiveness was found in PALDs with live attenuated vaccines (ie, all PLADs except Beijing, Tianjin, Shanghai, and Jiangsu). Differences in EIRs could be boiled down to 2 reasons, namely, the level of development and the incidence before the implementation of the EPI. The regions where society and economy have forged ahead during the last decade (ie, high GDP per capita, the fast urbanization process, sufficient medical resources, decreased proportion of children, or advanced education level), accompanied by improvements in sanitary and hygienic conditions (ie, sufficient medical resources), are associated with the less exposure to HAV, which could partly explain the regional difference in EIRs [24,35]. More could be credited to the fact that there were lower incidence of hepatitis A and higher seroprevalence of anti-HAV antibody in regions with thriving economy before the intervention [11].

In terms of the nonintervention population beyond the age group of 2-9 years, the EPI-associated reduction in the incidence of hepatitis A was observed. This finding indicated that routine mass immunization on target children provided a degree of indirect protection for unvaccinated individuals, in line with a previous study in the United States [36]. With fewer hepatitis A cases in target children, the probability of cross-infection from target children to nontarget people reduces greatly. On the other hand, among nonvaccinated individuals, people who were neither covered by the EPI nor had paid for hepatitis A vaccines out of pocket remain vulnerable to hepatitis A. Herd immunity can prevent the spread of hepatitis A in the population to a certain degree but will not adequately protect susceptible individuals upon exposure to HAV [36,37]. Hence, health authorities need to provide education on scientific prevention directives and formulate supervisory policies in light of different infection sources and transmission routes, such as monitoring HAV circulation [38], safeguarding food safety [39], and raising health consciousness, to pave the way for nontarget people to avoid HAV exposure.

Limitations

Several limitations should be acknowledged in this study. First, all hepatitis A cases were diagnosed and reported by hospitals. There is inevitable underreporting. Particularly, children are more likely to become asymptomatic and be underreported. The underreporting decreases with the improvement of the diagnosis process, which could lead to the underestimation of the effectiveness. However, the odds favor a random bias on asymptomatic cases before and after the implementation of the EPI, thus exerting nondifferential impacts on the assessment of the intervention. Second, the lack of the data on hepatitis A vaccination coverage and the seroprevalence blocks us from better understanding the pathway and the regional heterogeneity of the EPI’s effects on the hepatitis A incidence. Finally, our study emphasized the effect of the EPI on the incidence. The economic benefits and decreased DALYs brought by the EPI have not been investigated. Further dynamic research is greatly needed to articulate the substantial positive effect on controlling the HAV endemic in terms of cost-effectiveness after decades of the EPI and to further evaluate the feasibility of achieving WHO 2030 targets.

Conclusion

The implementation of EPI has evidently contributed to the decrease of hepatitis A incidence among children aged 2-9 years in the Chinese mainland, suggesting that continuous children mass vaccination is imperative in the future containment of hepatitis A. In view of the regional differences in the effectiveness of the EPI observed, sustained monitoring of the incidence and strengthening of vaccination in Western China will accelerate efforts toward achieving the target of eliminating viral hepatitis set by the WHO. Furthermore, the findings would serve as a favorable paradigm of mass hepatitis A vaccination for other countries and regions with a heavy burden of hepatitis.
A, which contributes to the eventual elimination of hepatitis A worldwide.

Acknowledgments
We thank the staff of Chinese Center for Disease Control and Prevention who collected the data. This work was supported by the National Natural Science Foundation of China (grants 81903406 and 82003555). The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data Availability
Relevant raw data related to reported cases of hepatitis A in the Chinese mainland can be collected on National Population and Health Science Data Sharing Platform. The data sets generated during and/or analyzed in this study are available from the corresponding authors on reasonable request.

Authors' Contributions
CQO conceptualized and designed the study. CQO and LL supervised the study. BWM, HNH, CS, and ZY collected the data. BWM, JJM, CS, XHX, and ZY analyzed the data. BWM and LL wrote the first draft of the manuscript. All authors contributed to the interpretation of the results and edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Supplemental methods, tables, and figures.

References


Abbreviations

- DALY: disability-adjusted life year
- eCI: empirical confidence interval
- EIR: excess incidence rate
- EPI: Expanded Program on Immunization
- ER: excess risk
- GDP: gross domestic product
- HAV: hepatitis A viruses
- PLAD: provincial-level administrative division
- WHO: World Health Organization

Edited by R Marzo; submitted 26.10.23; peer-reviewed by J Kaewkungwal, Q Wang, P Guo; comments to author 01.01.24; revised version received 08.01.24; accepted 22.01.24; published 28.02.24.

Please cite as:
Ming BW, Li L, Huang HN, Ma JI, Shi C. Xu XH, Yang Z. Ou CQ
The Effectiveness of National Expanded Program on Immunization With Hepatitis A Vaccines in the Chinese Mainland: Interrupted Time-Series Analysis
JMIR Public Health Surveill 2024;10:e53982
URL: https://publichealth.jmir.org/2024/1/e53982
doi:10.2196/53982
PMID:38416563

©Bo-Wen Ming, Li Li, Hao-Neng Huang, Jia-Jun Ma, Chen Shi, Xiao-Han Xu, Zhou Yang, Chun-Quan Ou. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 28.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Review

The Impact of Optimal Glycemic Control on Tuberculosis Treatment Outcomes in Patients With Diabetes Mellitus: Systematic Review and Meta-Analysis

Li Zhao1, BS; Feng Gao2, PhD; Chunlan Zheng1, BS; Xuezhi Sun1, BS

1Department of Tuberculosis III, Wuhan Pulmonary Hospital, Wuhan, China
2Department of Endocrinology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:
Xuezhi Sun, BS
Department of Tuberculosis III
Wuhan Pulmonary Hospital
Hubei Province
Wuhan, 430030
China
Phone: 86 15871745280
Email: sxz16032002@163.com

Abstract

Background: Diabetes mellitus (DM) increases the risk of developing tuberculosis (TB), and optimal glycemic control has been shown to reduce the risk of complications and improve the TB treatment outcomes in patients with DM.

Objective: This study aims to investigate the role of glycemic control in improving TB treatment outcomes among patients with DM.

Methods: MEDLINE, Embase, and the Cochrane Central Register of Controlled Trials databases were searched for randomized controlled trials (RCTs) assessing the impact of oral glycemic control in patients with TB who have DM. Outcomes of interest were radiological findings, treatment success, sputum positivity, and mortality. Evaluations were reported as risk ratios (RRs) with 95% CIs using weighted random-effects models.

Results: The analysis included 6919 patients from 7 observational studies. Our meta-analysis showed significant differences between patients with optimal glycemic control and those with poor glycemic control with regard to improved treatment outcomes (RR 1.13, 95% CI 1.02-1.25; P=.02; P=65%), reduced sputum positivity (RR 0.23, 95% CI 0.09-0.61; P=.003; P=66%), and fewer cavitary lesions (RR 0.59, 95% CI 0.51-0.68; P<.001; P=0%) in radiological findings. There was no significant difference between the 2 groups in terms of mortality (RR 0.57, 95% CI 0.22-1.49; P=.25; P=0%), multilobar involvement (RR 0.57, 95% CI 0.22-1.49; P=.25; P=0%) on radiologic examination, and upper lobe (RR 0.94, 95% CI 0.76-1.17; P=.58; P=0%) and lower lobe (RR 1.05, 95% CI 0.48-2.30; P=.91; P=75%) involvement on radiologic examination.

Conclusions: We concluded that optimal glycemic control is crucial for reducing susceptibility, minimizing complications, and improving treatment outcomes in patients with TB with DM. Emphasizing effective health management and health care strategies are essential in achieving this control. Integrating comprehensive care among patients with TB with DM will enhance patient outcomes and alleviate the burden of disease in this population.

Trial Registration: PROSPERO CRD42023427362; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=427362

(JMIR Public Health Surveill 2024;10:e53948) doi:10.2196/53948

KEYWORDS

optical glycemic control; poor glycemic control; tuberculosis treatment; diabetes mellitus; health management, healthcare; health care; glycemic control; tuberculosis; TB; DM; diabetes; systematic review; meta-analysis; risks; treatment; treatment outcome; mortality; patients; burden; disease burden
Introduction

Tuberculosis (TB) poses an escalating public health threat, particularly in lower- and middle-income countries [1]. The World Health Organization estimates that approximately one-fourth of the world’s population has been infected with TB-causing bacteria [2], with 10.6 million individuals diagnosed with TB in 2021, leading to 1.6 million deaths [1]. Risk factors for TB are divided into 2 main categories, that is, people recently infected with TB and those with an immunocompromised status, including those with HIV, diabetes mellitus (DM), transplants, malnourishment, and tobacco use, and those receiving immunosuppressants [1,2].

DM is also a growing concern, increasing the likelihood of several infections and complications [3]. With 425 million individuals affected in 2017 and an estimated 629 million expected to be impacted by 2045, DM increases the risk of TB incidence by 2-4 folds. Furthermore, it is associated with poor outcomes, doubling the risk of mortality during treatment [4,5]. In the long term, hyperglycemia and poor glycemic control (PGC) impair immunity, leading to immunosuppression and increased susceptibility to TB [1].

Poor treatment outcomes have been associated with patients with TB, including treatment failure, recurrence, delayed culture conversion, and death [6]. Optimal glycemic control (OGC) has been shown to improve TB outcomes by enhancing phagocytic activity and other immunological defense mechanisms [6]. Nevertheless, some studies have found no significant improvement in TB treatment outcomes through glycemic control [7]. As a result, there is a need to examine the current data to establish the relationship between the 2 factors. This paper aims to review the current literature and reach a conclusion regarding the impact of OGC on TB treatment outcomes in patients with DM.

Methods

Overview

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and the risk of bias assessed using AMSTAR (A Measurement Tool to Assess Systematic Reviews) 2 were both used when performing this meta-analysis [8,9]. This study is registered on PROSPERO (The International Prospective Register of Systematic Reviews; ID CRD42023427362).

Ethical Considerations

Since the information was accessible to the general public, institutional review board approval was not necessary.

Data Sources and Search Strategy

MEDLINE, Embase, and the Cochrane Central Register of Controlled Trials were comprehensively searched from inception through May 2023 by 2 independent reviewers (LZ and XS). We extracted studies based on abstracts and titles. A full-text appraisal was sought when required. MeSH (Medical Subject Headings) phrases and keywords were used to formulate search strategies (Table 1).

Table 1. Search strategy used in each database.

<table>
<thead>
<tr>
<th>Database (articles retrieved)</th>
<th>Search strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embase (56 results)</td>
<td>(“tuberculosis”/exp OR “tuberculosis” OR “TB”) AND (“diabetes mellitus”/exp OR “diabetes mellitus” OR “DM”) AND (“glycemic control”/exp OR “glycemic control” OR “blood glucose control” OR “blood sugar control”) AND (“randomized controlled trial”/exp OR “clinical trial”/exp OR “observational study”/exp)</td>
</tr>
<tr>
<td>Cochrane Central Register of Controlled Trials (25 results)</td>
<td>(“Tuberculosis” OR “TB”) AND (“Diabetes Mellitus” OR “DM”) AND (“Glycemic Control” OR “Blood Glucose Control” OR “Blood Sugar Control”) AND (“Randomized Controlled Trial” OR “Clinical Trial” OR “Observational Study”)</td>
</tr>
</tbody>
</table>

Study Selection

We included studies if they (1) were randomized controlled trials (RCTs) or analyses of RCTs that determined the impact of OGC on treatment outcomes of TB in patients with DM in different interventional arms; (2) reported radiological findings including cavity lesions, multifocal involvement, and upper and lower lobe involvement; treatment success; sputum positivity; or mortality; or (3) included patients with a diagnosis of TB and DM. We also included observational studies that reported the aforementioned radiological findings, treatment success, sputum positivity, and mortality. A third investigator (XS) was consulted in case of any disagreement regarding study selection. All articles were then uploaded to EndNote Reference Library (version X7.5; Clarivate Analytics) software to remove any duplicates.

Data Extraction and Assessment of Study Quality

Two reviewers (FG and CZ) independently extracted from the selected studies the characteristics of the studies, patient demographics, summary events, number of events, sample sizes, and treatment type. The quality of the included studies was assessed using the Newcastle-Ottawa Scale across 6 key domains: selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias. This systematic evaluation aimed to enhance the reliability of our findings by critically appraising the internal validity of each study. To enhance the reliability of our quality assessments, 2 independent reviewers (FG and CZ) conducted the evaluations. In instances
of disagreement, a third investigator (XS) was consulted, and consensus was reached through discussion. The process of quality assessment was conducted systematically and transparently, ensuring a rigorous evaluation of each study’s methodological robustness.

**Statistical Analysis**

Radiological findings, consisting of cavitary lesions, multilobar involvement, upper lobe involvement, and lower lobe involvement, were one of the outcomes of interest. Other outcomes were treatment success, sputum positivity, and mortality. RevMan (version 5.4.1; The Cochrane Collaboration) was used to conduct the meta-analysis. The outcomes of interest were provided as risk ratios (RRs), assessing the association between exposure and disease, indicating the risk of developing the disease in the exposed group versus the nonexposed group with 95% CIs and aggregated using an inverse variance–weighted random effects model. Forest plots were used to graphically display the pooled analyses. The Higgins $I^2$ was used to assess heterogeneity between trials; a value of 25%-50% was regarded as low, 50%-75% as moderate, and >75% as serious. In all cases, a $P$ value less than .05 was considered significant.

**Results**

**Search Results**

Our initial search yielded 2760 potentially relevant articles, of which 21 were selected for full-text review. Upon further exclusions, 7 observational studies, with a total of 6919 patients, were shortlisted for data extraction [10-16]. The PRISMA flowchart in Figure 1 shows the literature search in detail.

**Study Characteristics and Quality Assessment**

Study characteristics and baseline demographics are summarized in Table 2. Observational studies were assessed to be of moderate to high quality, achieving scores from 4 to 6 out of a maximum of 9 on the Newcastle-Ottawa scale (Table 3).
Table 2. Baseline characteristics of the included studies.

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Sites</th>
<th>Design</th>
<th>Intervention and exposure</th>
<th>Study duration</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiang et al [10] (2015)</td>
<td>North, south, and east Taiwan</td>
<td>Retrospective cohort study</td>
<td>Criteria for classification of glycemic control at baseline were as follows: HbA1c &lt;7 —</td>
<td>2005-2010</td>
<td>705 patients with DM&lt;sup&gt;b&lt;/sup&gt; with culture-positive pulmonary TB&lt;sup&gt;c&lt;/sup&gt; of both sexes; age not reported irrespective of HIV status; diagnosed with DM within 3 months of initiation of ATT&lt;sup&gt;d&lt;/sup&gt;; included 768 patients with TB without DM.</td>
</tr>
<tr>
<td>Mi et al [11] (2013)</td>
<td>South China</td>
<td>Retrospective cohort study</td>
<td>Criteria for classification of glycemic control (baseline, 2 months, and 6 months): good</td>
<td>2011-2012</td>
<td>189 patients with pulmonary or extrapulmonary TB with DM of all age groups, both sexes, and HIV status not reported; included 1400 patients with TB without DM.</td>
</tr>
<tr>
<td>Magee et al [12] (2013)</td>
<td>Lima, Peru</td>
<td>Retrospective/prospective cohort study</td>
<td>Based on documentation of control in records, an FBG level below median, or an FBG level of &lt;136 mmol/L. Exact criteria are not specified. Details on glucose lowering treatment are available: OHA&lt;sup&gt;f&lt;/sup&gt; only—56 participants; insulin only—16 participants; both—26 participants.</td>
<td>2005-2008</td>
<td>Selected group of patients with TB screened for high risk of developing multidrug-resistant TB (people with presumptive multidrug-resistant TB); 1485 patients with TB without DM and 186 patients with TB and DM receiving new treatment or retreatment, aged ≥15 years, of either sex, regardless of HIV status were also included.</td>
</tr>
<tr>
<td>Nandakumar et al [13] (2013)</td>
<td>Kerala, India</td>
<td>Retrospective cohort study</td>
<td>Criteria for glycemic control: assessed 3 times, at least 1 month apart, and at least in 1 control program. Those fulfilling all 3 criteria were classified as having a “known” diabetic control status. Those with all of the following 3 values less than the cutoff were classified as “controlled”: FBS level of &lt;100 mg, postprandial blood sugar or random blood sugar level of &lt;140 mg.</td>
<td>2010-2011</td>
<td>667 patients with TB and DM, new or retreatment, pulmonary or extrapulmonary TB, aged ≥15 years, belonging to either sex, irrespective of HIV status.</td>
</tr>
<tr>
<td>Park et al [14] (2012)</td>
<td>South Korea</td>
<td>Retrospective cohort study</td>
<td>Criteria for glycemic control assessment at baseline were as follows: glycemic control—HbA1c &lt;7; poor glycemic control—HbA1c ≥7.</td>
<td>2005-2009</td>
<td>New patients with pulmonary TB, 96 with TB and DM, and 148 with TB without DM, aged ≥18 years, of either sex, whose HIV status is not reported.</td>
</tr>
<tr>
<td>Mahishale et al [16] (2017)</td>
<td>India</td>
<td>Prospective cohort study</td>
<td>Glycemic control was defined at baseline: poor glycemic control—HbA1c ≥7%; optimal glycemic control—HbA1c &lt;7%; no mention of NGSP&lt;sup&gt;g&lt;/sup&gt; certification and standardized to the DCCT&lt;sup&gt;h&lt;/sup&gt; assay.</td>
<td>2012-2014</td>
<td>675 new patients with pulmonary TB belonging to either sex, age group unspecified, and excluding known HIV-positive cases.</td>
</tr>
</tbody>
</table>

<sup>a</sup>HbA1c: hemoglobin A1c.<br><sup>b</sup>DM: diabetes mellitus.<br><sup>c</sup>TB: tuberculosis.<br><sup>d</sup>ATT: anti-tuberculosis treatment.<br><sup>e</sup>FBG: fasting blood glucose.<br><sup>f</sup>OHA: oral hypoglycemic agent.<br><sup>g</sup>NGSP: National Glycohemoglobin Standardization Program.<br><sup>h</sup>DCCT: Diabetes Control and Complications Trial.
Table 3. Quality assessment of included observational studies using the Newcastle-Ottawa Scale.

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Selection</th>
<th>Comparability</th>
<th>Outcome</th>
<th>Total, n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
<td>S2</td>
<td>S3</td>
<td>S4</td>
</tr>
<tr>
<td>Chiang et al [10] (2015)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Magee et al [12] (2013)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mahishale et al [16] (2017)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Outcomes

**Treatment Outcome**

Five studies reported treatment outcomes among patients with TB, which included patients who completed the treatment and were completely cured. Our meta-analysis revealed a significant difference in treatment outcomes between patients with OGC and those with PGC (RR 0.86, 95% CI 0.74-1.00; P=0.05; I²=51%; Figure 2).

**Sputum Positivity Following Treatment**

Three studies reported sputum positivity as an outcome among patients with TB. Sputum positivity was more likely among patients with PGC than among those with OGC (RR 0.23, 95% CI 0.09-0.61; P=0.003; I²=66%; Figure 3).

**Mortality**

Two studies reported mortality as an outcome among patients with TB. Mortality was not significantly different between patients with OGC and those with PGC (RR 0.57, 95% CI 0.22-1.49; P=0.25; I²=0%; Figure 4).
Radiological Outcomes

A total of 3 studies reported radiological findings among patients with TB. These findings were further divided into cavitary lesions, multilobar involvement, isolated upper lobe involvement, and lower lobe involvement.

Cavitary Lesions

Three studies reported cavitary lesions as one of their radiological findings. Patients with OGC had a lower risk of cavitary lesions than those with PGC (RR 0.59, 95% CI 0.51-0.68; \( P < .001; \ I^2 = 0\% \); Figure 5).

Multilobar Involvement

Among the 3 studies reporting radiological findings, 2 reported multilobar involvement. There was no significant difference between patients with OGC and those with PGC (RR 0.82, 95% CI 0.58-1.17; \( P = .27; \ I^2 = 0\% \); Figure 6).

Isolated Upper or Lower Lobe Involvement

Two studies reported isolated lower lobe involvement, while 1 study reported isolated upper lobe involvement as their radiological outcomes. There was no significant difference between patients with OGC and those with PGC for both upper (RR 0.94, 95% CI 0.76-1.17; \( P = .58; \ I^2 = 0\% \)) and lower lobe involvement (RR 1.05, 95% CI 0.48-2.30; \( P = .91; \ I^2 = 75\% \); Figures 7 and 8).

Discussion

Principal Findings

This meta-analysis evaluating the impact of OGC among patients with TB reports several key findings. First, patients with OGC demonstrated a decreased risk of the aforementioned treatment outcomes. Second, we observed decreased sputum positivity in patients with OGC compared to that in patients with suboptimal glycemic control. Third, there was a decreased risk of cavitary lesions on radiologic examination among patients with OGC.
The analysis revealed a significant improvement in treatment outcomes among patients with OGC compared to those with PGC. This finding is consistent with those of previous studies highlighting the importance of glycemic control in reducing the risk of complications and improving treatment responses in individuals with DM [17,18]. The ability to achieve OGC may enhance the body's immune response, leading to better control of the TB infection and a more favorable treatment outcome [19]. Moreover, our analysis showed that patients with OGC had decreased sputum positivity. This may be attributable to the quicker clearance of bacteria from the airways, resulting in earlier detection. The decreased sputum positivity following OGC points toward the clinical benefits and importance of achieving OGC among patients with TB.

There was no significant difference in mortality between patients with OGC and those with PGC. This result contradicts the findings of previous studies that have linked DM to an increased risk of mortality among patients with TB [20]. It is important to note that the studies included in the meta-analysis might have varied in terms of follow-up duration and other factors that could influence mortality outcomes. Further research is needed to better understand the relationship among glycemic control, TB treatment outcomes, and mortality.

Patients with OGC had a lower risk of cavitary lesions than those with PGC, according to our analysis. Cavitary lesions are indicative of advanced disease and are associated with an increased risk of TB transmission [21]. This finding suggests that OGC may aid in preventing the progression of disease and decreasing the risk of transmission. However, multilobar involvement, isolated upper lobe involvement, and isolated lower lobe involvement did not differ significantly between patients with OGC and those with PGC. This indicates that OGC may have a limited effect on the distribution of TB lesions in the lungs [22].

Publication bias—the tendency of studies with positive or statistically significant results to be published more readily than those with null or negative results—is a concern in meta-analyses. However, due to the limited number of studies available for each outcome, the statistical power of Egger and Begg tests may have been compromised. As such, the ability to draw definitive conclusions regarding publication bias is constrained. To mitigate this limitation, we attempted to include a broad range of studies by searching multiple databases and imposing minimal restrictions on study design. Additionally, we actively sought unpublished studies, conference abstracts, and gray literature to reduce the impact of publication bias. However, despite these efforts, it is essential to interpret our findings with caution, considering the potential influence of publication bias on the reported results.

Overall, our findings indicate that OGC is essential for enhancing TB treatment outcomes and lowering the risk of advanced disease. In patients with TB, health care providers should consider screening for DM and managing glycemic control [23]. To thoroughly comprehend the relationship among DM, glycemic control, and TB outcomes, additional research is required.

In evaluating the robustness of our findings, we conducted a thorough quality assessment using the Cochrane Risk of Bias Tool, systematically appraising studies across key domains. While our inclusive approach aimed to minimize publication bias by considering a broad range of studies and actively seeking unpublished data, the limitations in conclusively identifying and mitigating publication bias should be acknowledged. Regarding heterogeneity, variations in study design, patient populations, glycemic control thresholds, and outcome measurements were identified as potential sources. These factors introduce complexity and may limit the generalizability of our results. Clinicians should interpret our findings with caution, considering the diverse contexts and populations represented in the included studies. Future research addressing standardized definitions of glycemic control and consistent outcome measures will contribute to a more nuanced understanding of the relationship between glycemic control and TB treatment outcomes.

Our comprehensive findings have clinical implications for both individual patient care and public health strategies in the context of TB management. Notably, OGC not only improves overall TB treatment outcomes but also has emerged as a critical factor in reducing its infectiousness, as evidenced by the observed decreases in sputum positivity and the lower risk of cavitary lesions. The reduction in sputum positivity implies a potential limitation on TB transmission, presenting a dual benefit for both individual patients’ well-being and broader public health goals. Additionally, the lower risk of cavitary lesions, indicative of advanced TB disease, signifies a potential avenue for mitigating the contagiousness of patients with TB. Clinicians should emphasize the importance of achieving and sustaining OGC, recognizing its dual impact on individual health and community-level TB transmission. Integrated health care strategies focusing on glycemic control in patients with TB are vital, providing actionable insights for clinicians and public health practitioners alike and contributing to the overarching goal of TB control and prevention. While our findings provide valuable insights into the association between glycemic control and TB outcomes in patients with DM, generalizing these results to a wider population requires caution. The unique characteristics of patients with TB with DM, the potential variations in glycemic control thresholds, and the diverse health care settings may limit the direct applicability of our findings to those without DM or with different comorbidities. Furthermore, the prevalence of observational studies in our analysis introduces biases that may affect the external validity of our results. Caution is advised in extending these findings to diverse patient populations, and future research should explore the relationship between glycemic control and TB outcomes in broader contexts, considering various comorbidities and health care settings.

However, there are some limitations to consider in interpreting the results. First, the included studies were observational in nature, which may introduce biases, such as selection bias and confounding factors that could influence the results. Moreover, an observational study design has variability in its population, selective and incomplete reporting, and improper randomization, which may contribute to the heterogeneity observed in this
study. Second, heterogeneity was observed in some of the studies, which may be attributed to differences in study design, patient populations, and glycemic control thresholds across the included studies. The limited research carried out on this topic shows that OGC is an important predictor of outcomes in patients with TB; however, there are some discrepancies, which may raise doubt among clinicians. Thus, future investigation on this topic is warranted in order to derive a robust conclusion.

Conclusions
In conclusion, our meta-analysis compiled data from observational studies assessing glycemic control in patients with TB, and our results suggest that OGC may have a significant impact on improving treatment outcomes and reducing sputum positivity in patients with TB. However, no significant difference was found in mortality between patients with OGC and those with PGC. OGC was also associated with a lower risk of cavitary lesions but had no significant effect on multilobar or isolated upper or lower lobe involvement. These findings highlight the importance of early detection of TB in patients with DM so that OGC can be provided promptly to those patients, thus improving their outcomes. This topic warrants further research, especially RCTs focusing on mortality and other outcomes in different severities of TB among patients with DM.

Authors' Contributions
LZ and XS drafted the manuscript. FG and CZ are responsible for all the figures.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA checklist.

References


Abbreviations

AMSTAR: A Measurement Tool to Assess systematic Reviews
DM: diabetes mellitus
MeSH: Medical Subject Headings
OGC: optimal glycemic control
PGC: poor glycemic control
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: The International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RR: risk ratio
TB: tuberculosis

Edited by A Mavragani, R Marzo; submitted 28.10.23; peer-reviewed by MY Essar, S Shrestha; comments to author 18.12.23; revised version received 09.01.24; accepted 10.01.24; published 02.04.24.

Please cite as:
Zhao L, Gao F, Zheng C, Sun X
The Impact of Optimal Glycemic Control on Tuberculosis Treatment Outcomes in Patients With Diabetes Mellitus: Systematic Review and Meta-Analysis
JMIR Public Health Surveill 2024;10:e53948
URL: https://publichealth.jmir.org/2024/1/e53948
doi:10.2196/53948
PMID:38564244
COVID-19 Vaccine Hesitancy: Umbrella Review of Systematic Reviews and Meta-Analysis

Tahani Al Rahbeni1*, PhD; Prakasini Satapathy2, PhD; Ramaiah Itumalla3*, PhD; Roy Rillera Marzo4, MD; Khalid A L Mugeed5, PhD; Mahalaqua Nazli Khatib6, MD; Shilpa Gaidhane6, MD; Quazi Syed Zahiruddin7, MD; Ali A Rabaan8,9,10, PhD; Hayam A Alrasheed11, PhD; Maha F Al-Subaie8,12, PhD; Nawal A Al Kaabili13,14, PhD; Mohammed Alissa15, PhD; Amani Ahmed A L Ibrahim16, PhD; Hussain Abdulkhaliq Alsair17, PhD; Israa Habeeb Naser18, PhD; Sarvesh Rustagi19, PhD; Neelima Kukreti20, PhD; Arkadiusz Dziedzic21*, PhD

1Molecular Toxicology and Genetics, Riyadh Elm University, Riyadh, Saudi Arabia
2Center for Global Health Research, Saveetha Medical College and Hospital, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India
3School of Management, The Apollo University, Chittoor, India
4Faculty of Humanities and Health Sciences, Curtin University, Miri Sarawak, Malaysia
5Division of Evidence Synthesis, Global Consortium of Public Health and Research, Datta Meghe Institute of Higher Education, Wardha, India
6One Health Centre (COHERD), Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education, Wardha, India
7South Asia Infant Feeding Research Network, Division of Evidence Synthesis, School of Epidemiology and Public Health and Research, Jawaharlal Nehru Medical College, Datta Meghe Institute of Higher Education and Research, Wardha, India
8Molecular Diagnostic Laboratory, Johns Hopkins Aramco Healthcare, Dhahran, Saudi Arabia
9College of Medicine, Alfaisal University, Riyadh, Saudi Arabia
10Department of Public Health and Nutrition, The University of Haripur, Haripur, Pakistan
11Department of Pharmacy Practice, College of Pharmacy, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia
12Research Center, Dr Sulaiman Alhabib Medical Group, Riyadh, Saudi Arabia
13College of Medicine and Health Science, Khalifa University, Abu Dhabi, United Arab Emirates
14Sheik Khalifa Medical City, Abu Dhabi Health Services Company (SEHA), Abu Dhabi, United Arab Emirates
15Department of Medical Laboratory, College of Applied Medical Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia
16Department of Pharmacy, Jubail General Hospital, Jubail, Saudi Arabia
17Department of Medicine, Batterjee Medical College, Jeddah, Saudi Arabia
18Medical Laboratories Techniques Department, AL-Mustaqbal University, Babil, Iraq
19School of Applied and Life Sciences, Uttaranchal University, Dehradun, India
20School of Pharmacy, Graphic Era Hill University, Dehradun, India
21Department of Conservative Dentistry with Endodontics, Medical University of Silesia, Katowice, Poland

* these authors contributed equally

Corresponding Author:
Prakasini Satapathy, PhD
Center for Global Health Research
Saveetha Medical College and Hospital
Saveetha Institute of Medical and Technical Sciences, Saveetha University
162, Poonamallee High Road
Poonamallee
Chennai, 600077
India
Phone: 91 8763766300
Email: prakasini.satapathy@gmail.com

Related Article:
This is a corrected version. See correction statement: https://publichealth.jmir.org/2024/1/e64080
Abstract

Background: The unprecedented emergence of the COVID-19 pandemic necessitated the development and global distribution of vaccines, making the understanding of global vaccine acceptance and hesitancy crucial to overcoming barriers to vaccination and achieving widespread immunization.

Objective: This umbrella review synthesizes findings from systematic reviews and meta-analyses to provide insights into global perceptions on COVID-19 vaccine acceptance and hesitancy across diverse populations and regions.

Methods: We conducted a literature search across major databases to identify systematic reviews and meta-analysis that reported COVID-19 vaccine acceptance and hesitancy. The AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews) criteria were used to assess the methodological quality of included systematic reviews. Meta-analysis was performed using STATA 17 with a random effect model. The data synthesis is presented in a table format and via a narrative.

Results: Our inclusion criteria were met by 78 meta-analyses published between 2021 and 2023. Our analysis revealed a moderate vaccine acceptance rate of 63% (95% CI 0.60%-0.67%) in the general population, with significant heterogeneity ($I^2 = 97.59$%). Higher acceptance rates were observed among health care workers and individuals with chronic diseases, at 64% (95% CI 0.57%-0.71%) and 69% (95% CI 0.61%-0.76%), respectively. However, lower acceptance was noted among pregnant women, at 48% (95% CI 0.42%-0.53%), and parents consenting for their children, at 61.29% (95% CI 0.56%-0.67%). The pooled vaccine hesitancy rate was 32% (95% CI 0.25%-0.39%) in the general population. The quality assessment revealed 19 high-quality, 38 moderate-quality, 15 low-quality, and 6 critically low-quality meta-analyses.

Conclusions: This review revealed the presence of vaccine hesitancy globally, emphasizing the necessity for population-specific, culturally sensitive interventions and clear, credible information dissemination to foster vaccine acceptance. The observed disparities accentuate the need for continuous research to understand evolving vaccine perceptions and to address the unique concerns and needs of diverse populations, thereby aiding in the formulation of effective and inclusive vaccination strategies.

Trial Registration: PROSPERO CRD42023468363; https://tinyurl.com/2p9kv9cr

(JMIR Public Health Surveill 2024;10:e54769) doi:10.2196/54769

KEYWORDS

COVID-19; vaccine acceptance; vaccine hesitancy; umbrella review; systematic review; meta-analysis; vaccine; hesitancy; global perceptions; perception; random effect model; synthesis; healthcare workers; patients; patient; chronic disease; pregnant women; parents; child; children

Introduction

The global health landscape has been profoundly altered by the emergence of COVID-19, triggered by SARS-CoV-2. First identified in Wuhan, China, in December 2019, this virulent pathogen swiftly traversed continents, leading the World Health Organization (WHO) to categorize the situation as both a pandemic and public health emergency of international concern. The repercussions of this pandemic have been multifaceted, with a staggering death toll and profound impact on socioeconomic structures worldwide [1]. In the face of this unprecedented challenge, the global community recognized the pressing need for effective countermeasures. Although therapeutic interventions were explored, the primary focus shifted to preventive strategies, with vaccines against COVID-19 emerging as the most promising solution [2]. The efficacy of this approach, however, is contingent not just on the scientific success of vaccine development but equally on the global populace’s acceptance of these vaccines [3].

By the midpoint of 2022, the scientific community had successfully developed, trialed, and secured emergency use authorization for several vaccines [4]. However, the distribution of these vaccines unveiled pronounced disparities [5]. Higher-income nations, with their robust health care infrastructures and financial resources, rapidly initiated vaccination drives. In stark contrast, many resource-limited countries faced challenges ranging from limited vaccine access to infrastructural constraints [6]. A more insidious challenge that emerged globally, irrespective of a country’s economic status, was vaccine hesitancy. Rooted in a complex interplay of factors, including safety apprehensions, distrust toward health advisories, cultural nuances, and the deluge of misinformation, vaccine hesitancy has been observed across diverse geographies, from Africa and Europe to North America [3].

Empirical studies conducted across various regions have painted a mixed picture of vaccine acceptance [3,7]. Although certain demographics exhibited a commendable eagerness to embrace vaccination, others displayed pronounced skepticism [3,8]. These disparities in vaccine acceptance, if unchecked, have the potential to impede global strides toward achieving herd immunity, a critical milestone in the fight against the pandemic. Recognizing the pivotal role of vaccine acceptance in the trajectory of the pandemic, it becomes imperative to understand the nuances of global vaccine perceptions. Numerous systematic reviews and meta-analyses have been published, shedding light on factors that are driving vaccine hesitancy and acceptance [8-17].

In this context, our umbrella review sought to collate and synthesize findings from these diverse studies, aiming to present a holistic understanding of global COVID-19 vaccine acceptance rates and hesitancy rates. This approach offers a comprehensive
overview of existing evidence, highlighting consistencies and discrepancies across studies. By assessing the quality and breadth of current research, umbrella reviews identify knowledge gaps and inform evidence-based decision-making. They serve as a valuable tool for policymakers, clinicians, and researchers, providing a holistic understanding of a topic without the need for sifting through numerous individual studies. Through this study, we aspired to provide valuable insights that can steer future vaccination strategies, ensuring they are both effective and inclusive. Our umbrella review aimed to collate and synthesize findings from these diverse studies to present a comprehensive understanding of global COVID-19 vaccine acceptance and hesitancy rates.

**Methods**

The method for conducting this umbrella review was based on the framework set forth by the Joanna Briggs Institute [18]. This study adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline [19] (Multimedia Appendix 1). This study was registered in PROSPERO.

**Inclusion Criteria**

We specifically targeted meta-analyses of epidemiological studies that investigated either the acceptance/willingness or hesitancy toward the COVID-19 vaccine. Our scope was global, encompassing studies from all geographical locations without any specific focus on a particular population. This inclusivity ensured that our review captured a diverse range of perspectives and settings. However, to maintain the rigor and specificity of our review, we excluded certain types of publications. Specifically, narrative or systematic reviews that did not include a meta-analysis, conference abstracts, and letters to the editors were not considered. In essence, our inclusion criteria were centered on meta-analyses of prospective, retrospective, or cross-sectional studies that evaluated rates of vaccine acceptance or hesitancy (Table S1 in Multimedia Appendix 2).

**Literature Search**

We conducted a comprehensive literature search across 4 major databases: PubMed, Scopus, Embase, and the Cochrane Database of Systematic Reviews. The search spanned from the inception of each database until August 20, 2023. To ensure a thorough retrieval of relevant articles, “keyword search” and “textword search” were used, and different search phrases were combined using Boolean and proximity operators. Specifically, we used the terms (“meta-analysis” OR “systematic review”) AND (Acceptance OR willingness OR hesitancy OR intention OR unwillingness) AND (“COVID-19” OR “Sars-cov-2” OR “corona“”). To further enhance the robustness of our search, we also manually screened the reference lists of the identified articles. This step ensured that we did not overlook any pertinent studies that might not have been captured through the database search. For transparency and replicability, the complete search strategy, including all terms and combinations used, is documented in Table S2 in Multimedia Appendix 2. Importantly, we did not impose any filters or restrictions during our search. This means that articles of any type, published in any year, and in any language were considered, ensuring a broad and inclusive search scope.

**Screening**

The screening process for this systematic review was conducted by 2 independent authors, structured into 2 sequential steps to ensure unbiased selection and comprehensive coverage of relevant studies. The first step, primary screening, involved scrutinizing the titles and abstracts of identified articles to shortlist those potentially relevant to topic. Subsequently, in the second step, articles that passed the primary screening were subjected to a thorough full-text review. During this stage, the authors carefully evaluated the complete content of each article, focusing on the removal of duplicates and a more detailed assessment of each study’s relevance and alignment with the review’s scope and objectives. To enhance the precision and efficiency of our screening process, we used the specialized software Nested Knowledge with its AutoLit function, instrumental in streamlining our workflow and improving the accuracy of article selection. In cases of disagreement or uncertainty between the 2 reviewers, a third reviewer with senior expertise was engaged to mediate and provide decisive judgment, ensuring a consensus-based approach to the final selection of studies.

**Data Extraction**

During the data extraction process, 2 independent authors systematically reviewed each study that met our inclusion criteria. From each eligible meta-analysis, they gathered a comprehensive set of details. This included the first author's name, year of publication, type of study design, total number of participants, type of population, and the date when the database search was conducted. Additionally, they extracted the pooled acceptance rate for each subgroup and specific number of participants within these subgroups, accompanied by their 95% CIs. Furthermore, they documented the P values for pooled effects, the results from the Egger or Beggs test (which measures publication bias), and the I² statistics, which offer insights into the heterogeneity of the studies. Any associated P value for significance was also recorded. Given the complexity of the data and the importance of accuracy, any discrepancies or disagreements that arose between the 2 primary authors were diligently addressed. They consulted a third, senior reviewer to ensure consistency and precision in the data extraction process.

**Assessment of Methodological Quality**

To ensure the rigor and reliability of our review, the methodological quality of each included meta-analysis was meticulously evaluated. This assessment was jointly undertaken by 2 authors using the well-established AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews) tool [20], which is recognized for its robustness in appraising systematic reviews. Based on the criteria set by AMSTAR-2, studies were categorized into 1 of 4 methodological quality grades: high, moderate, low, or critically low. A study was deemed to be of high quality if it exhibited no flaws or only a single minor defect. In contrast, a moderate quality designation was given to studies that presented multiple minor defects. The distinction between...
minor and major defects was made based on the guidelines provided by the AMSTAR-2 tool. Any disagreements or uncertainties regarding the quality grading were discussed and resolved collaboratively between the 2 authors to ensure a consistent and objective assessment.

Data Analysis

Data analysis was conducted using STATA version 17. Proportions, along with their 95% CIs, were pooled from all included eligible meta-analyses for each outcome and based on the population [21]. We used a random effects model to compute the combined effect sizes, recognizing the inherent variability among the studies and providing a more conservative estimate of the overall effect. The degree of variability or heterogeneity in outcomes across studies was quantified using the $I^2$ metric. Values for $I^2$ can range from 0% to 100%, with higher values indicating greater heterogeneity. We predetermined specific thresholds to assess the statistical significance of the observed heterogeneity. The 95% prediction interval provides a more comprehensive understanding of the range in which the true effect size lies, considering the observed heterogeneity. A $P$ value <0.05 was considered statistically significant.

Results

Search Results

We identified a total of 662 articles through the primary search, of which 263 duplicates were eliminated, leaving 399 for title and abstract screening. In the primary screening (title and abstract), 214 articles were excluded, leaving 185 articles for full-text screening (Figure 1). We excluded 108 articles for various reasons, including only systematic reviews without a meta-analysis and the incorrect population, outcomes, or study design. As a result, 77 articles fulfilled the eligibility criteria. Additionally, we conducted a citation search to maintain the rigor of the review and found 4 relevant articles, of which 1 was included. This umbrella review ultimately identified 78 meta-analyses that met the inclusion criteria.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram illustrating the screening and selection process.

Characteristics of Meta-Analyses and Quality Assessment

Table S3 in Multimedia Appendix 2 [9-17,22-90] presents an overview of all included meta-analyses published between 2021 and 2023. These studies looked at different groups, such as the general population; health care workers; parents; pregnant women; migrant workers; Black or African communities; Chinese communities; and people with specific health conditions such as cancer, HIV, diabetes, inflammatory bowel disease, and epilepsy. The focus of these meta-analyses was on outcomes such as the rates of vaccine acceptance, hesitance, willingness, uncertainty, and unwillingness, as well as the intention to receive vaccines. Most of the meta-analyses included articles from around the world, while some concentrated on specific countries. The quality of these meta-analyses was evaluated using the AMSTAR-2 criteria. Among the meta-analyses examined, 19 meta-analyses were rated as being of high quality; 38 meta-analyses received a moderate quality rating; 15 meta-analyses received a low quality rating, which points to potential limitations in their methodology; and 6 meta-analyses were classified as having critically low quality, implying significant concerns about their methods and the reliability of their findings (Table S4 in Multimedia Appendix 2 [9-17,22-90]).
Vaccine Acceptance and Hesitancy Among Populations

Table 1 provides a summary of vaccine acceptance and hesitancy rates among different populations. Acceptance rates were studied in 58 systematic reviews and meta-analyses, and 12 distinct reviews reported on hesitancy rates (Table 2).
<table>
<thead>
<tr>
<th>Studies by population</th>
<th>Acceptance rate, % (95% CI)</th>
<th>Heterogeneity ($I^2$), %</th>
<th>Overall acceptance rate, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al, 2021 [22]</td>
<td>0.74 (0.71-0.77)</td>
<td></td>
<td>0.63 (0.60-0.67)</td>
</tr>
<tr>
<td>Alimohamadi et al, 2022 [10]</td>
<td>0.67 (0.62-0.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdelmoneim et al, 2022 [23]</td>
<td>0.81 (0.75-0.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nehal et al, 2021 [15]</td>
<td>0.66 (0.6-0.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khabour, 2022 [24]</td>
<td>0.39 (0.33-0.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sahile et al, 2022 [25]</td>
<td>0.57 (0.47-0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norhayati et al, 2021 [12]</td>
<td>0.61 (0.59-0.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wake, 2021 [26]</td>
<td>0.48 (0.39-0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alarcón-Braga et al, 2022 [27]</td>
<td>0.78 (0.74-0.82)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mekonnen and Mengistu, 2022 [28]</td>
<td>0.56 (0.47-0.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mengistu et al, 2022 [29]</td>
<td>0.64 (0.6-0.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gudayu and Mengistie, 2023 [30]</td>
<td>0.68 (0.67-0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kumar et al, 2023 [16]</td>
<td>0.62 (0.55-0.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alemayehu et al, 2022 [31]</td>
<td>0.60 (0.52-0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kawuki et al, 2023 [32]</td>
<td>0.58 (0.49-0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al, 2021 [22]</td>
<td>0.67 (0.67-0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belay et al, 2022 [33]</td>
<td>0.51 (0.43-0.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robinson et al, 2021 [34]</td>
<td>0.72 (0.66-0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahmud et al, 2022 [35]</td>
<td>0.62 (0.58-0.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azanaw et al, 2022 [36]</td>
<td>0.55 (0.47-0.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terry et al, 2022 [37]</td>
<td>0.73 (0.64-0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yenew et al, 2023 [38]</td>
<td>0.67 (0.60-0.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kukreti et al, 2022 [39]</td>
<td>0.60 (0.51-0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nnaemeka et al, 2023 [40]</td>
<td>0.52 (0.46-0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akem Dimala et al, 2021 [41]</td>
<td>0.71 (0.66-0.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renzi et al, 2022 [42]</td>
<td>0.66 (0.61-0.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kazeminia et al, 2022 [43]</td>
<td>0.63 (0.59-0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mose et al, 2022 [44]</td>
<td>0.51 (0.43-0.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yanto et al, 2022 [9]</td>
<td>0.71 (0.69-0.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chronic disease</strong></td>
<td></td>
<td></td>
<td>87.50</td>
</tr>
<tr>
<td>Wang et al, 2021 [22]</td>
<td>0.85 (0.82-0.88)</td>
<td></td>
<td>0.69 (0.61-0.76)</td>
</tr>
<tr>
<td>Yazdani et al, 2022 [45]</td>
<td>0.76 (0.67-0.85)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhao et al, 2023 [46]</td>
<td>0.65 (0.59-0.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin et al, 2022 [47]</td>
<td>0.58 (0.45-0.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meybodi et al, 2022 [48]</td>
<td>0.59 (0.39-0.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ejamo et al, 2023 [49]</td>
<td>0.62 (0.56-0.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ekpor and Akyirem, 2023 [50]</td>
<td>0.76 (0.66-0.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prabani et al, 2022 [51]</td>
<td>0.59 (0.52-0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Health care workers</strong></td>
<td></td>
<td></td>
<td>91.72</td>
</tr>
<tr>
<td>Wang et al, 2021 [22]</td>
<td>0.65 (0.55-0.75)</td>
<td></td>
<td>0.64 (0.57-0.71)</td>
</tr>
<tr>
<td>Luo et al, 2021 [11]</td>
<td>0.51 (0.41-0.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Studies by population</td>
<td>Acceptance rate, % (95% CI)</td>
<td>Heterogeneity ($I^2$), %</td>
<td>Overall acceptance rate, % (95% CI)</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------------------------</td>
<td>--------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Alimohamadi et al, 2022 [10]</td>
<td>0.55 (0.47-0.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shui et al, 2022 [52]</td>
<td>0.78 (0.73-0.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin et al, 2022 [85]</td>
<td>0.81 (0.72-0.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ackah et al, 2022 [53]</td>
<td>0.46 (0.37-0.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moltot et al, 2023 [54]</td>
<td>0.54 (0.42-0.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ulbrichtova et al, 2022 [55]</td>
<td>0.71 (0.67-0.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Politis et al, 2023 [56]</td>
<td>0.64 (0.55-0.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al, 2022 [13]</td>
<td>0.66 (0.61-0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pregnant women</strong></td>
<td></td>
<td>74.20</td>
<td>0.48 (0.42-0.53)</td>
</tr>
<tr>
<td>Sarantaki et al, 2022 [57]</td>
<td>0.53 (0.44-0.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nikpour et al, 2022 [58]</td>
<td>0.54 (0.45-0.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nassr et al, 2022 [59]</td>
<td>0.47 (0.38-0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halemani et al, 2022 [60]</td>
<td>0.54 (0.46-0.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shamshirsaz et al, 2022 [61]</td>
<td>0.47 (0.38-0.57)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galanis et al, 2022 [62]</td>
<td>0.27 (0.18-0.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhattacharya et al, 2022 [63]</td>
<td>0.49 (0.42-0.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worede et al, 2023 [14]</td>
<td>0.42 (0.28-0.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azami et al, 2022 [64]</td>
<td>0.53 (0.47-0.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parents regarding vaccinating their children</strong></td>
<td>50.29</td>
<td>0.61 (0.56-0.67)</td>
<td></td>
</tr>
<tr>
<td>Wang et al, 2022 [65]</td>
<td>0.58 (0.28-0.98)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galanis et al, 2022 [66]</td>
<td>0.6 (0.517-0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen et al, 2022 [67]</td>
<td>0.61 (0.53-0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ma et al, 2022 [68]</td>
<td>0.7 (0.62-0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alimoradi et al, 2023 [69]</td>
<td>0.57 (0.52-0.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Migrants and refugees</strong></td>
<td></td>
<td>74.04</td>
<td>0.69 (0.56-0.82)</td>
</tr>
<tr>
<td>Alimoradi et al, 2023 [70]</td>
<td>0.7 (0.62-0.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hajissa et al, 2023 [71]</td>
<td>0.56 (0.449-0.685)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chinese community residents</strong></td>
<td>N/A*</td>
<td></td>
<td>0.80 (0.72-0.88)</td>
</tr>
<tr>
<td>Xu and Zhu, 2022 [72]</td>
<td>0.8 (0.71-0.87)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aN/A: not applicable.
Table 2. Summary of vaccine hesitancy rates across different populations.

<table>
<thead>
<tr>
<th>Studies by population</th>
<th>Hesitancy rate, % (95% CI)</th>
<th>Heterogeneity ($I^2$, %)</th>
<th>Overall hesitancy rate, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General population</strong></td>
<td></td>
<td>73.90</td>
<td>0.32 (0.25-0.39)</td>
</tr>
<tr>
<td>Patwary et al, 2022 [73]</td>
<td>0.382 (0.272-0.497)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Islam et al, 2023 [17]</td>
<td>0.265 (0.22-0.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kawuki et al, 2023 [32]</td>
<td>0.29 (0.18-0.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fajar et al, 2022 [74]</td>
<td>0.25 (0.19-0.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cénat et al, 2022 [75]</td>
<td>0.423 (0.337-0.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Older adults</strong></td>
<td>N/A</td>
<td>0.27 (0.16-0.39)</td>
<td></td>
</tr>
<tr>
<td>Veronese et al, 2021 [76]</td>
<td>0.27 (0.15-0.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Black/African American people</strong></td>
<td>N/A</td>
<td>0.35 (0.25-0.44)</td>
<td></td>
</tr>
<tr>
<td>Ripon et al, 2022 [77]</td>
<td>0.35 (0.26-0.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>People with diabetes</strong></td>
<td>N/A</td>
<td>0.27 (0.14-0.40)</td>
<td></td>
</tr>
<tr>
<td>Bianchi et al, 2023 [84]</td>
<td>0.27 (0.156-0.419)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Health care students</strong></td>
<td>N/A</td>
<td>0.26 (0.18-0.33)</td>
<td></td>
</tr>
<tr>
<td>Patwary et al, 2022 [79]</td>
<td>0.258 (0.185-0.338)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pregnant and breastfeeding women</strong></td>
<td>N/A</td>
<td>0.48 (0.43-0.53)</td>
<td></td>
</tr>
<tr>
<td>Bianchi et al, 2022 [80]</td>
<td>0.484 (0.434-0.534)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parents regarding vaccinating their children</strong></td>
<td>95.79</td>
<td>0.39 (0.07-0.70)</td>
<td></td>
</tr>
<tr>
<td>Bianchi et al, 2023 [81]</td>
<td>0.55 (0.43-0.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galanis et al, 2022 [66]</td>
<td>0.229 (0.173-0.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Migrants</strong></td>
<td>N/A</td>
<td>0.31 (0.21-0.41)</td>
<td></td>
</tr>
<tr>
<td>Hajissa et al, 2023 [71]</td>
<td>0.31 (0.215-0.42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Health care workers</strong></td>
<td>97.30</td>
<td>0.29 (0.18-0.33)</td>
<td></td>
</tr>
<tr>
<td>Bianchi et al, 2022 [78]</td>
<td>0.13 (0.069-0.209)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kigongo et al, 2023 [82]</td>
<td>0.46 (0.38-0.54)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Not applicable.

**Vaccine Acceptance**

We synthesized findings from 29 systematic reviews to assess the vaccine acceptance rate in the general population. The pooled acceptance rates ranged from 51% to 81%. Our meta-analysis revealed a consolidated vaccine acceptance rate of 63% (95% CI 0.60%-0.67%). Notably, a high level of heterogeneity was observed, with an $I^2$ of 97.59% (Figure 2). The 8 systematic reviews focused on individuals with chronic diseases reported a pooled acceptance rate of 69% (95% CI 0.61%-0.76%) and an $I^2$ of 87.5% (Multimedia Appendix 3). Health care workers were the focus of 10 systematic reviews and meta-analyses, indicating a 64% acceptance rate (95% CI 0.57%-0.71%) and a heterogeneity $I^2$ of 91.72% (Multimedia Appendix 4). Vaccine acceptance was comparatively lower among pregnant women, as depicted by 9 systematic reviews, showing a rate of 48% (95% CI 0.42%-0.53%) and an $I^2$ of 74.2% (Multimedia Appendix 5). Similarly, 5 systematic reviews presented a 61.29% acceptance rate (95% CI 0.56%-0.67%) with 50% heterogeneity for parents consenting for their children (Multimedia Appendix 6). Vaccine acceptance among migrants and refugees was investigated by 2 reviews, showing a prevalence of 69% (95% CI 0.56%-0.82%; Multimedia Appendix 7).
Vaccine Hesitancy

Figure 3 shows the forest plot of COVID-19 vaccine hesitancy for different populations. Vaccine hesitancy in the general population was reported by 5 systematic reviews, with the observed hesitancy varying between 25% and 42%. Our meta-analysis showed a pooled vaccine hesitancy rate of 32% (95% CI 0.25%-0.39%), with a high level of heterogeneity ($I^2=73.90\%$). In older adults, 1 review reported a hesitancy rate of 27% (95% CI 0.15%-0.38%). For Black people or African Americans, vaccine hesitancy was 35% (95% CI 0.26%-0.45%) in another review. Pregnant or breastfeeding women exhibited a higher hesitancy rate of 48.4% (95% CI 0.43%-0.53%), as reported by another systematic review. Results on the hesitancy rates among parents considering vaccinating their children were provided by 2 reviews, revealing a pooled hesitancy rate of 39% (95% CI 0.07%-0.70%), accompanied by a high level of heterogeneity ($I^2=95.7\%$). Migrant workers exhibited a hesitancy
rate of 31% (95% CI 0.21%-0.41%) according to 1 study. Last, health care workers showed a rate of 29% (95% CI 0.18%-0.33%), as concluded from 2 systematic reviews.

**Figure 3.** Forest plot showing COVID-19 vaccine hesitancy rates for different populations. REML: restricted maximum likelihood.

**Discussion**

Our umbrella review synthesized findings from numerous systematic reviews and meta-analyses, providing insights into global vaccine acceptance and hesitancy rates across diverse populations and geographies. The consolidated acceptance rate of 63% in the general population indicates a moderate level of willingness to receive the vaccine.

The emergence of the COVID-19 pandemic necessitated the prompt development and distribution of vaccines to curb the spread of the virus and mitigate its adverse impacts on global health, economies, and societies [91]. As of the midpoint of
2021, several vaccines had received emergency use authorization, signifying a milestone in the fight against the pandemic. However, the realization of the potential of these vaccines is significantly influenced by the global population's acceptance and willingness to get vaccinated [92,93]. However, the significant heterogeneity observed in this and other populations studied underscores the diverse and complex landscape of vaccine perceptions globally [53]. The vaccine acceptance rates among health care workers and individuals with chronic diseases were relatively higher, possibly reflecting a better understanding of the disease's risk and the vaccine's benefits by these populations [94,95]. However, the observed heterogeneity suggests diverse opinions and possibly varied information dissemination within these groups. The disparities in vaccine acceptance and hesitancy across populations are emblematic of the intricate tapestry of perceptions, beliefs, and information access that characterize the global populace. For instance, the lower acceptance rates observed among pregnant women and parents consenting for their children are likely influenced by concerns about vaccine safety in these vulnerable groups, emphasizing the need for targeted communication strategies addressing these concerns [96].

The pronounced disparities in vaccine acceptance across different populations highlight the urgent need for tailored, population-specific intervention strategies [97]. A one-size-fits-all approach may not address the unique concerns, misconceptions, and information needs of different demographic groups. For example, pregnant women exhibited lower acceptance and higher hesitancy rates, possibly due to concerns regarding the vaccine's impact on pregnancy and the fetus [98,99]. Addressing such specific concerns through targeted awareness campaigns and counseling can enhance vaccine acceptance in this group. Similarly, the lower acceptance rates in parents consenting for their children necessitate interventions addressing parental concerns about vaccine safety and efficacy in children [100]. Engaging pediatricians and child health care providers in vaccine advocacy can potentially alleviate parental apprehensions and foster trust. The high heterogeneity observed across studies denotes the existence of multiple influencing factors, including cultural, socioeconomic, educational, and individual beliefs, which vary extensively within and across populations [101,102]. The variations may also reflect the differences in study designs, populations, and time frames, emphasizing the need for standardization in future research to facilitate comparability and generalizability [103].

The emergence of vaccine hesitancy as a global phenomenon irrespective of a country's economic status underscores the influential role of information dissemination and public perception in shaping vaccine-related behaviors [104,105]. Misinformation and distrust in health advisories have been pivotal in fostering hesitancy, indicating the need for credible, clear, and consistent communication from health authorities and governments [106-110]. Addressing misinformation necessitates a multifaceted approach involving the collaboration of health care providers, public health officials, social media platforms, and community leaders. The propagation of accurate, comprehensible, and transparent information regarding vaccine development, efficacy, and safety can contribute to counteracting misinformation. Health care workers, with an acceptance rate of 64%, play a crucial role in shaping public perceptions and behaviors regarding vaccination [111]. As trusted sources of health information, health care providers can effectively address concerns, clarify misconceptions, and advocate for the benefits of vaccination [63,68]. Their interactions with patients and communities can significantly influence vaccine acceptance, especially in populations with high hesitancy levels, such as pregnant women and parents. However, the hesitancy rate of 29% among health care workers is concerning, as it can potentially impact their vaccine advocacy efforts. Addressing the concerns and information needs of health care workers is imperative to fostering confidence in vaccines and enhancing their role as vaccine advocates.

The acceptance and hesitancy rates in specific communities, such as Black or African American and Chinese communities, underscore the impact of cultural and community nuances on vaccine perceptions [112,113]. Culturally sensitive approaches, community engagement, and addressing systemic barriers are essential to enhancing vaccine acceptance in such communities [114,115]. The 80% acceptance rate in Chinese communities may be indicative of the influence of community norms, government policies, and public health campaigns in shaping vaccine perceptions. Understanding the sociocultural dynamics and leveraging community influences can be instrumental in developing effective strategies to enhance vaccine acceptance in different cultural contexts.

This umbrella review, although offering insights into global vaccine acceptance and hesitancy, does possess limitations inherent to the studies included. The substantial heterogeneity across these studies hinders the ability to draw definitive conclusions and underscores the necessity for a cautious interpretation of the findings. Variations in study designs, targeted populations, time frames, and geographic locations highlight the need for standardization in future research to improve comparability and generalizability. Our review only included articles published in English. The overlap of the same primary studies is inevitable; different systematic reviews might have included the same primary studies.

Future research should focus on exploring the underlying factors influencing vaccine acceptance and hesitancy in diverse populations and contexts. Qualitative studies can provide in-depth insights into individual beliefs, perceptions, and information needs, enabling the development of targeted interventions. Longitudinal studies can assess the temporal variations in vaccine perceptions and the impact of evolving information landscapes on vaccine-related behaviors.
Authors' Contributions
TAR, PS, RI, and AD conceived of or designed the study. AAR, HAA, MFAS, NAAK, MA, AAALI, HAA, IHN, SR, and NK collected the data. KALM, MNK, SG, and QSZ analyzed the data. PS, TAR, RI, and AD drafted the manuscript. KALM, MNK, SG, and QSZ critically revised the manuscript. All authors gave final approval for submission. All authors attest they meet the ICMJE criteria for authorship.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA Checklist.
[PDF File (Adobe PDF File), 134 KB - publichealth_v10i1e54769_app1.pdf ]

Multimedia Appendix 2
Supplementary tables.
[PDF File (Adobe PDF File), 592 KB - publichealth_v10i1e54769_app2.pdf ]

Multimedia Appendix 3
Forest plot illustrating the pooled vaccine acceptance rate for individuals with chronic diseases.
[PNG File, 113 KB - publichealth_v10i1e54769_app3.png ]

Multimedia Appendix 4
Forest plot depicting the vaccine acceptance rate among health care workers.
[PNG File, 127 KB - publichealth_v10i1e54769_app4.png ]

Multimedia Appendix 5
Forest plot depicting the vaccine acceptance rate among pregnant women.
[PNG File, 127 KB - publichealth_v10i1e54769_app5.png ]

Multimedia Appendix 6
Pooled vaccine acceptance rate of parents consenting for their children.
[PNG File, 87 KB - publichealth_v10i1e54769_app6.png ]

Multimedia Appendix 7
Vaccine acceptance rates for Chinese community residents, migrants, and refugees.
[PNG File, 125 KB - publichealth_v10i1e54769_app7.png ]

References


Abbreviations

AMSTAR: A Measurement Tool to Assess Systematic Reviews
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization
Laboratory Data Timeliness and Completeness Improves Following Implementation of an Electronic Laboratory Information System in Côte d’Ivoire: Quasi-Experimental Study on 21 Clinical Laboratories From 2014 to 2020

Yao He¹, MPH, PhD; Yves-Rolland Kouabenan², MPH; Paul Henri Assou³, BSc, MSc; Nancy Puttkammer¹, MPH, PhD; Bradley H Wagenaar³,⁴, MPH, PhD; Hong Xiao⁵, MPH, PhD; Stephen Gloyd³, MPH, MD; Noah G Hoffman⁶, MD, PhD; Pascal Komena³, BSc; N’zi Pierre Fourier Kamelan³, BSc; Casey Iiams-Hauser¹, MIA; Adama Sanogo Pongathie⁷, MD; Alain Kouakou⁷, MS; Jan Flowers¹, MSc; Nadine Abiola², PharmD; Natacha Kohemun⁸, PharmD; Jean-Bernard Amani⁸, MSc; Christiane Adje-Toure⁹, MSc, MIH, PhD; Lucy A Perrone⁴,¹⁰, MSPH, PhD

¹Digital Initiatives Group at International Training and Education Center for Health, Department of Global Health, Schools of Public Health and Medicine, University of Washington, Seattle, WA, United States
²International Training and Education Center for Health - Côte d’Ivoire, Abidjan, Cote D’Ivoire
³Department of Global Health, Schools of Public Health and Medicine, University of Washington, Seattle, WA, United States
⁴Department of Epidemiology, Schools of Public Health and Medicine, University of Washington, Seattle, WA, United States
⁵Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, United States
⁶Department of Pathology and Laboratory Medicine, University of Washington, Seattle, WA, United States
⁷Direction de l’Informatique et de l’Information Sanitaire, Ministry of Health, Public Hygiene and Universal Health Coverage, Abidjan, Cote D’Ivoire
⁸Laboratory Branch, United States Centers for Disease Control and Prevention, Abidjan, Cote D’Ivoire
⁹Retro-CI Laboratory, United States Centers for Disease Control and Prevention, Abidjan, Cote D’Ivoire
¹⁰Department of Pathology and Laboratory Medicine, University of British Columbia (UBC), Vancouver, BC, Canada

Corresponding Author:
Lucy A Perrone, MSPH, PhD
Department of Pathology and Laboratory Medicine
University of British Columbia (UBC)
2211 Wesbrook Mall
UBC Hospital, Koerner Pavilion, GF401
Vancouver, BC, V6T 2B5
Canada
Phone: 1 6048272972
Email: lucy.perrone@ubc.ca

Abstract

Background: The Ministry of Health in Côte d’Ivoire and the International Training and Education Center for Health at the University of Washington, funded by the United States President’s Emergency Plan for AIDS Relief, have been collaborating to develop and implement the Open-Source Enterprise-Level Laboratory Information System (OpenELIS). The system is designed to improve HIV-related laboratory data management and strengthen quality management and capacity at clinical laboratories across the nation.

Objective: This evaluation aimed to quantify the effects of implementing OpenELIS on data quality for laboratory tests related to HIV care and treatment.

Methods: This evaluation used a quasi-experimental design to perform an interrupted time-series analysis to estimate the changes in the level and slope of 3 data quality indicators (timeliness, completeness, and validity) after OpenELIS implementation. We collected paper and electronic records on clusters of differentiation 4 (CD4) testing for 48 weeks before OpenELIS adoption until 72 weeks after. Data collection took place at 21 laboratories in 13 health regions that started using OpenELIS between 2014 and 2020. We analyzed the data at the laboratory level. We estimated odds ratios (ORs) by comparing the observed outcomes with modeled counterfactual ones when the laboratories did not adopt OpenELIS.
**Results:** There was an immediate 5-fold increase in timeliness (OR 5.27, 95% CI 4.33-6.41; \( P < .001 \)) and an immediate 3.6-fold increase in completeness (OR 3.59, 95% CI 2.40-5.37; \( P < .001 \)). These immediate improvements were observed starting after OpenELIS installation and then maintained until 72 weeks after OpenELIS adoption. The weekly improvement in the postimplementation trend of completeness was significant (OR 1.03, 95% CI 1.02-1.05; \( P < .001 \)). The improvement in validity was not statistically significant (OR 1.34, 95% CI 0.69-2.60; \( P = .38 \)), but validity did not fall below pre-OpenELIS levels.

**Conclusions:** These results demonstrate the value of electronic laboratory information systems in improving laboratory data quality and supporting evidence-based decision-making in health care. These findings highlight the importance of OpenELIS in Côte d’Ivoire and the potential for adoption in other low- and middle-income countries with similar health systems.

(JMIR Public Health Surveill 2024;10:e50407) doi:10.2196/50407

**KEYWORDS**
clinical laboratory; electronic laboratory information system; data quality; evaluation; impact evaluation; time series; laboratory; information system; information systems; HIV; sexually transmitted infection; sexually transmitted disease; sexually transmitted; lab; labs; adoption; implementation; effectiveness

**Introduction**
Clinical laboratories provide critically important data for disease investigations and supply timely information for public health. Laboratory results help identify and characterize pathogens, conduct routine surveillance, and respond to communicable and noncommunicable diseases. Effective and timely laboratory services help achieve national disease priorities and global targets including the Sustainable Development Goals and the 95-95-95 targets launched by the Joint United Nations Programme on HIV/AIDS [1-5].

Clusters of differentiation 4 (CD4) testing is critical for monitoring the immune status of people living with HIV and the risk of opportunistic infections. Health care providers depend on laboratory results to decide on differentiated care models and tuberculosis screening [6]. For a national HIV program to be highly successful and reduce HIV-related deaths, the health system must integrate data from an effective clinical laboratory system. Access to quality laboratory services remains a challenge in low- and middle-income countries (LMICs), and many still rely on paper-based records [7]. Throughout the cascade of care, paper documents and records require a careful chain of possession, tracking, and version control, and their use and sometimes loss can impede an already complex process of service delivery. Problems in data quality could occur at every step of this manual process, jeopardizing the quality of care.

Laboratory information systems (LISs) are an important part of the infrastructure for laboratory operation and information management. According to the requirements for quality and competence of medical laboratories issued by the International Organization for Standardization (ISO), most of the information management requirements are for electronic systems [8]. This signals that, to reach ISO standards, laboratories should use electronic rather than paper-based LISs. The World Health Organization established the Stepwise Laboratory Quality Improvement Process Toward Accreditation (SLIPTA) checklist to guide clinical laboratories in Africa through continuous quality improvement toward accreditation to international standards [9]. The SLIPTA checklist highlights computerized LISs as an integral step to ensuring laboratory operation and fulfilling the criteria of laboratory information management on ensuring data quality, data storage and backup, and patient confidentiality [9]. The lack of effective electronic LISs is one of the barriers to quality laboratory service delivery in LMICs [10].

LISs can improve the quality of care, patient safety, and disease surveillance [10-14]. Compared to paper-based information systems, electronic LISs provide more timely and more accurate monitoring and reporting of turnaround time, test failure rates, and other laboratory quality indicators essential to informing clinical care [10]. The features in the Open-Source Enterprise-Level Laboratory Information System (OpenELIS) are intended to ensure data timeliness, completeness, and validity. Control elements and graphical widgets such as drop-down menus and radio buttons make data entry easier and faster, supporting data timeliness and validity. Required data fields must be completed to generate the laboratory result for a patient, assuring data completeness. Valid ranges displayed alongside each test result, the flags for manual review of out-of-range results, and the mandatory validation step before generating results help ensure data validity.

The International Training and Education Center for Health (I-TECH) has been collaborating with the Ministry of Health, Public Hygiene, and Universal Health Coverage (MSHPCMU) in Côte d'Ivoire since 2009 with funding from the United States President’s Emergency Plan for AIDS Relief to build a functional, accredited diagnostic and laboratory network for effective HIV care and treatment. One of the key program interventions is establishing and supporting the use of OpenELIS interfacing with laboratory testing analyzers [15]. From 2009 to 2021, OpenELIS was installed at 108 laboratories in Côte d'Ivoire, including 68 sites that use it primarily for HIV data, 1 for food and drug safety data, 27 for tuberculosis data only, and 12 sites that use OpenELIS for routine laboratory testing data.

This study aims to estimate the effect of OpenELIS on data quality indicators including data completeness, timeliness, and validity, using CD4 testing data. Specifically, we aimed to answer the following questions: (1) Did the use of OpenELIS in Côte d'Ivoire’s clinical laboratories lead to statistically significant improvements in the data quality outcomes? (2) If so, what was the magnitude of the improvements compared to the counterfactual scenario in which the laboratories had not
adopted OpenELIS? and (3) Had the improvements persisted after the initial adoption of OpenELIS?

Beyond a single study of the effect of an LIS on laboratory turnaround time in Peru [16], there is limited evidence on how LIS interventions enhance data quality in LMICs. Findings from our study will provide evidence on the effectiveness of the OpenELIS intervention in improving data quality; inform policy recommendations and guidance for the Côte d'Ivoire MSHPCMU; and inform decision makers elsewhere to consider adopting OpenELIS in countries with a context similar to Côte d'Ivoire.

Methods

Study Design and Data Sources

We conducted an interrupted time-series analysis using weekly time-series data. The interrupted time-series analysis is a quasi-experimental design in the evaluation of health care interventions or policies that are introduced at a specific time [17,18]. We hypothesized that, compared to paper registries, OpenELIS would improve laboratory data quality measured by 3 primary outcomes, that is, timeliness, completeness, and validity (Figure 1). Timeliness was defined as the proportion of test results produced within 1 day upon receiving test samples. Completeness was defined as the proportion of test results having complete data on all 4 required data fields, namely patient identifier, age, sex, and result date. Validity was defined as the proportion of CD4 cell count test results that were within the valid range (0-2000 cells/mm³).

We collected time-series data in January 2021 from 21 clinical laboratories across 13 health regions in Côte d'Ivoire that started using OpenELIS between 2014 and 2020 (Figure 2). The time frame of the time-series data covered up to 48 weeks before a laboratory started using OpenELIS until up to 72 weeks after. This sampling considered adequate representation in terms of geography and service capacity, as well as the feasibility of data collection. A total of 5 sites were regional reference laboratories that provided HIV viral load testing services, 3 were regional laboratories that were not reference sites for HIV viral load testing, and 13 were part of either at general hospitals or urban hospitals that provided HIV-related laboratory services at the district level. We abstracted deidentified, individual-level patient data for all outcomes from laboratory paper registries for the preimplementation period and from local OpenELIS servers for the postimplementation period. Data were aggregated by laboratory and organized by week to model the outcomes. The number of laboratories whose data were incorporated in the analyses for timeliness, completeness, and validity was 8, 21, and 20, respectively. The reasons why the number of laboratories for each outcome was different are the following. To estimate the timeliness of the data in paper registries, we needed to link individual-level records from test sample reception registries with those from test result registries by unique medical record number. A total of 13 of the 21 sampled laboratories did not have any sample reception registries from before they started using OpenELIS, so we could only estimate changes in timeliness at 8 laboratories. Accidents such as fire and flood were the most common reasons why laboratories were missing paper records, and some laboratories destroyed paper records older than 5 years because the national policy only requires 5 years for archiving paper records. To estimate completeness, we had all necessary data from all 21 sampled sites. To estimate validity, we needed data on CD4 cell counts, but 1 laboratory did not record CD4 cell counts in OpenELIS. Therefore, we could only estimate changes in validity for 20 sites.
**Figure 1.** Data flow process for a typical clinical laboratory test in Côte d’Ivoire (A) before and (B) after adopting OpenELIS. * denotes the difference between OpenELIS and a paper-based laboratory information system. CSV: comma-separated values; OpenELIS: Open-Source Enterprise-Level Laboratory Information System.
Figure 2. The 13 health regions in Côte d’Ivoire where data were collected from paper registries and OpenELIS for the evaluation of the effect of OpenELIS on data quality in Côte d’Ivoire from 2014-2020. OpenELIS: Open-Source Enterprise-Level Laboratory Information System. Adapted from Wikimedia [19], which is published under Creative Commons Attribution 4.0 International License [20].

Statistical Analysis

We used a mixed-effects, segmented logistic regression model to examine the trend of the primary outcomes and their changes associated with OpenELIS implementation. We estimated odds ratios (ORs) representing weekly outcomes compared with outcomes that would have occurred if the laboratories had not adopted OpenELIS. We reported both the level change and slope change, specifically investigating the abrupt changes in outcomes observed during the first week of OpenELIS implementation and the weekly changes in the trend starting from the second week of OpenELIS implementation.

We included fixed-effect monthly indicator variables in the models to account for seasonal effects, random intercepts to account for clustering at the laboratory level, and random slopes over time. We also accounted for the weekly number of HIV clients recorded at a given laboratory as weight. The logistic regression model equation we used was:

\[
y_{it} = \alpha_0 + \alpha_T \cdot \text{Time} + \alpha_OE \cdot OE_{i1} + \alpha_OE \cdot OE_{i2} + \alpha_P \cdot \text{Post time} + \alpha_M \cdot \text{Month} + \epsilon_{it}
\]

where \( i \) indicates laboratory and \( t \) indicates week as the time unit. \( y_{it} \) represents the proportion of laboratory records achieving each of the 3 primary outcomes comparing the factual to the counterfactual; \( \alpha_0 \) represents the model intercept with a fixed effect and a random effect at the laboratory level; \( \text{Time} \) counts the weeks from the earliest preimplementation time point until the latest postimplementation time point; \( \text{OE} \) is the dummy variable indicating the weeks when a laboratory \( i \) was using OpenELIS (1 for the first week when a laboratory started using and thereafter and 0 otherwise); \( \text{Post time} \) counts the weeks after the first week of implementation; and \( \text{Month} \) is an individual dummy variable indicating the month of the year with January as the reference.

To develop expected counterfactual forecasts, we reran the logistic models to extend preimplementation trends through 72 weeks after a laboratory started using OpenELIS as a counterfactual to the observed postimplementation trends. The mixed-effects models accounted for data that were missing at certain weeks using standard maximum likelihood estimation. To visualize the factual and counterfactual trends across time, we plotted the predicted proportions of test results achieving the primary outcomes under the 2 scenarios and the 95% prediction intervals for the factual estimates accounting for random variations at the laboratory level. All analyses were conducted using R (version 4.1.0; R Core Team) and a 2-sided \( \alpha \) value of .05.

Ethical Considerations

This study was determined to be nonhuman subject research by the University of Washington institutional review board and the United States Centers for Disease Control and Prevention (CDC). The study was approved by Côte d’Ivoire Comité National d’Ethique des Sciences de la Vie et de la Santé (CNESVS, Ivorian Institutional Review Board; reference 006-21/MSHP/CNESVS-km). The US CDC reviewed this research and paper but had no influential role in the study design, data collection, analysis, and interpretation.

Results

The number of laboratory weeks in the analyses were 613, 1820, and 1629 for timeliness, completeness, and validity, respectively. The analysis included an average of 34 weeks of
preimplementation data and 53 weeks of postimplementation data from each of the 21 laboratories. A total of 24,381 and 40,040 HIV client records from the pre- and postimplementation periods, respectively, were in the analysis (Table 1).

Only 3 of the 21 laboratories had all 48 weeks of complete data on sample reception before the laboratories adopted OpenELIS and when they used paper registries exclusively. A total of 10 laboratories had all 48 weeks of complete data on test results. After adopting OpenELIS, 19 of the 21 laboratories had 72 weeks of complete data for both sample reception and test results.

Before adopting OpenELIS, a typical laboratory had minor improvements in the 3 data quality outcomes that were not statistically significant (timeliness: OR 1.03, 95% CI 0.99-1.07; P=.12; completeness: OR 1.11, 95% CI 0.93-1.32; P=.24; and validity: OR 1.00, 95% CI 0.99-1.02; P=.76; Table 2). Within the first week of OpenELIS implementation, we observed an immediate level change in data timeliness, completeness, and validity (Figure 3). After adopting OpenELIS, the laboratories had 5.27-fold greater odds of timely results (OR 5.27, 95% CI 4.33-6.41; P<.001); 3.59-fold greater odds of complete data for the required fields (OR 3.59, 95% CI 2.40-5.37; P<.001); and 1.34-fold greater odds of valid results (OR 1.34, 95% CI 0.69-2.60; P=.38; Table 2). The immediate improvements in timeliness and completeness were both programmatically substantial and statistically significant.

Starting from the second week of OpenELIS implementation, we observed a slope change in the trend of completeness. The average odds of test results having complete data for the required fields each week thereafter were 1.03 times the odds had the laboratories not started using OpenELIS (OR 1.03, 95% CI 1.02-1.05; P<.001). In fact, the average proportion of test results having complete data stayed at almost 100% after a laboratory started using OpenELIS, while the average proportion declined over time in the counterfactual had there been no OpenELIS (Figure 3B). There was no significant change in the trend of timeliness and validity (timeliness: OR 1.01, 95% CI 0.99-1.02; P=.49 and validity: OR 0.99, 95% CI 0.97-1.01; P=.55).

In terms of heterogeneity at the laboratory level, laboratories at local general hospitals seemed to have more substantial improvements in timeliness immediately after starting to use OpenELIS and have maintained the improvements better over time as compared to regional laboratories (Figure S1 in Multimedia Appendix 1). Since a greater number of regional laboratories had less complete data than laboratories at general hospitals before OpenELIS implementation, immediate improvements in completeness were more obvious at regional laboratories, and the improvements also maintained over time (Figure S2 in Multimedia Appendix 1). Consistent with the result at the aggregate level, no laboratories had changes in validity (Figure S3 in Multimedia Appendix 1).

<table>
<thead>
<tr>
<th>Health region</th>
<th>Clinical laboratories, n</th>
<th>Weeks of data collected, n</th>
<th>HIV client records, n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
<td>Before</td>
</tr>
<tr>
<td>All regions</td>
<td>21</td>
<td>34</td>
<td>53</td>
</tr>
<tr>
<td>Abidjan II</td>
<td>3</td>
<td>35</td>
<td>68</td>
</tr>
<tr>
<td>Agneby-Tassa-Me</td>
<td>1</td>
<td>48</td>
<td>61</td>
</tr>
<tr>
<td>Cavally-Guemon</td>
<td>3</td>
<td>45</td>
<td>39</td>
</tr>
<tr>
<td>Gbeké</td>
<td>1</td>
<td>8</td>
<td>70</td>
</tr>
<tr>
<td>Gibokle-Nawa-San Pedro</td>
<td>3</td>
<td>27</td>
<td>65</td>
</tr>
<tr>
<td>Goh</td>
<td>1</td>
<td>39</td>
<td>23</td>
</tr>
<tr>
<td>Hambof*</td>
<td>1</td>
<td>14</td>
<td>31</td>
</tr>
<tr>
<td>Haut Sassandra</td>
<td>1</td>
<td>32</td>
<td>69</td>
</tr>
<tr>
<td>Indenie-Djuablin</td>
<td>2</td>
<td>32</td>
<td>51</td>
</tr>
<tr>
<td>Loh Djiboua</td>
<td>1</td>
<td>35</td>
<td>64</td>
</tr>
<tr>
<td>N’zi-Iffou</td>
<td>2</td>
<td>46</td>
<td>30</td>
</tr>
<tr>
<td>Poro-Tchologo-Bagoue</td>
<td>1</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>Tonkpi</td>
<td>1</td>
<td>14</td>
<td>69</td>
</tr>
</tbody>
</table>

*The time period for this evaluation was from 2014-2020.

OpenELIS: Open-Source Enterprise-Level Laboratory Information System.

The laboratory sampled in this region did not have data that could be included in the analysis on validity.
Table 2. Estimated ratios comparing odds of achieving the quality outcomes of timeliness, completeness, or validity using specifically CD4\textsuperscript{a} testing data, each week to the counterfactual scenario where the clinical laboratories in Côte d’Ivoire had not adopted OpenELIS\textsuperscript{b,c}.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Preadoption weekly change in the slope of the outcome</th>
<th>Immediate level change during first week of OpenELIS implementation</th>
<th>Postadoption weekly change in slope of the outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR\textsuperscript{d} (95% CI)</td>
<td>P value</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>Data timeliness</td>
<td>1.03 (0.99-1.07)</td>
<td>.12</td>
<td>5.267 (4.33-6.41)</td>
</tr>
<tr>
<td>Data completeness</td>
<td>1.11 (0.93-1.32)</td>
<td>.24</td>
<td>3.594 (2.40-5.37)</td>
</tr>
<tr>
<td>Data validity</td>
<td>1.00 (0.99-1.02)</td>
<td>.76</td>
<td>1.341 (0.69-2.60)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}CD4: clusters of differentiation 4.  
\textsuperscript{b}The time period for this evaluation was 2014-2020.  
\textsuperscript{c}OpenELIS: Open-Source Enterprise-Level Laboratory Information System.  
\textsuperscript{d}OR: odds ratio.

Figure 3. Laboratory test result timeliness, completeness, and validity from 48 weeks before OpenELIS implementation to 72 weeks after at clinical laboratories in Côte d’Ivoire. (A) Timeliness: average proportion of test results reported within 1 day of test sample reception; (B) Completeness: average proportion of test results having complete information for all required data fields; (C) Validity: average proportion of test results having valid results. Vertical dotted line: first week when a laboratory started using OpenELIS. Blue dots: weekly grand means of the observed outcomes at all laboratories. Black solid line: estimated trend around the observed data points. Light gray area: 95% prediction interval around the estimated observed trend. Red dashed line: counterfactual estimates if there were no OpenELIS. OpenELIS: Open-Source Enterprise-Level Laboratory Information System.
Discussion

This study shows that data availability improved after clinical laboratories in Côte d’Ivoire adopted OpenELIS, and there was rapid and substantial improvement in data timeliness and completeness. Following the immediate post-OpenELIS implementation improvements, both timeliness and completeness remained close to 100%. Data validity was very close to 100% both before and after OpenELIS adoption. There was some heterogeneity in the results at the laboratory level, with laboratories at local general hospitals showing more substantial improvements in timeliness while regional laboratories showed more obvious immediate improvements in completeness.

To our knowledge, this is the first study to demonstrate the immediate and maintained effect of a scaled, multisite LIS on data quality in sub-Saharan Africa. There are limited studies from LMICs despite significant financial investments from the international donor community. Previous studies from LMICs took place on other continents or did not include as many laboratories or subnational geographic areas as ours did. The e-Chasqui LIS for tuberculosis testing data at 14 laboratories in Peru was the only one that was implemented and evaluated through a cluster randomized controlled trial [21]. The trial showed that the LIS decreased turnaround time and improved quality of care, but the study did not examine the effect on data quality [21]. Our study contributes to the evidence by quantifying the effect size of data quality improvement and its change over a period of 30 months at 21 clinical laboratories. The findings offer critical information for policy makers and digital health researchers in LMICs when considering the real-world effectiveness of the OpenELIS software or digitalization of a sub-Saharan African laboratory system.

Systematic implementation of a single LIS across an entire national health system is rare in LMICs. A rapid literature review of peer-reviewed journal articles identified 8 different LISs that 6 countries have implemented at more than 1 site nationally, namely Ethiopia, Malawi, Namibia, South Africa, Peru, and Vietnam [16,21-30]. The LIS in South Africa is the only 1 that has the capacity to record data on all laboratory testing services at all clinical laboratories in the country [22]. It is likely that, as a result of the nationwide LIS adoption in South Africa, usage of laboratory data in research has also increased, since almost 20% of the rapid literature review search results from LMICs were epidemiological studies from South Africa using data from the LIS [31-39].

Part of the success of OpenELIS in Côte d’Ivoire can be attributed to the systematic approach taken by I-TECH in establishing and supporting the use of OpenELIS. The approach included the endorsement by the MSHPCMU and locally trained technical staff who facilitated implementation and provided continuous technical support of the system to the laboratories and their users. The relative advantages of OpenELIS compared to paper registries, ease of implementation, and continuous technical support have facilitated the routine use of OpenELIS. In addition, an experienced software development team designed various functional and interface features of the software specifically intended to improve data quality, for example, drop-down menus for fast and accurate selections and automatic data completeness and validity checks. The decade-long development and implementation of OpenELIS in Côte d’Ivoire has been open-source, iterative, and collaborative so that feedback from users, laboratory managers, and governing bodies was directly incorporated into the software to match the needs of the users [15].

Our study provides evidence for improving laboratory data quality, which is important for the improvement of clinical care and treatment. A study from Kenya identified that reduced turnaround time and timely communication of laboratory test results are some of the key factors for improving HIV treatment retention [12].

In addition, in terms of patient safety, LISs reduced the probability of errors in patient identification and subsequent inappropriate treatment by reducing manual clerical work through automation and user-friendly widgets [10]. A survey of clinicians in Malawi revealed that laboratories’ poor documentation of test results was one of the reasons for having little trust in laboratory capability and not frequently using laboratory test results in patient management [13]. Last but not least, LISs can also improve disease surveillance by making it easier to record accurate demographic data that inform disease classification, assessment of population-specific rates, and contact tracing for infectious diseases [11], as well as to report laboratory-confirmed diagnoses to surveillance systems that allow more accurate estimation of disease burdens [14].

This evaluation has many strengths and some limitations. The use of multiple data points before and after OpenELIS helps capture the changes in time-varying variables through the data we used to generate the pre- and posttrends. However, since all sites that have been using OpenELIS, including the ones sampled in this evaluation, were purposefully selected due to their higher capacity, they might also have had a higher probability of adhering to the OpenELIS user instructions. The regression models did not specifically account for potential confounding effects from the COVID-19 pandemic and related control measures that were enforced from March to July 2020 mostly in the Greater Abidjan area in Côte d’Ivoire [40]. However, since the percentage of weekly data points occurring after March 2020 was low (3.3% for timeliness, 7.0% for completeness, and 7.7% for validity), and most of the data points came outside Abidjan, we thought that the influence of the pandemic on our outcomes was low. Furthermore, the 3 outcomes were proportions of the total testing data, so the estimates were less likely to be influenced by fluctuations in total testing volume due to the pandemic. Nonetheless, we still included the total number of tests performed by a laboratory as weight in the regression model. The results may be generalizable to remaining clinical laboratories in Côte d’Ivoire that are at the same tiers in the health system as the sites sampled in this study, but generalizability to laboratories in other LMICs needs further studies to demonstrate.

The implementation of an LIS can have a significant effect on improving data quality in LMICs. This study found that the implementation of OpenELIS led to improved timeliness and...
completeness of laboratory data. The system also facilitated better data availability compared to paper registries, enabling health care providers to make informed decisions based on accurate and up-to-date information.

The effectiveness of an LIS depends on many factors beyond the capabilities of the LIS itself, including systematic planning for adoption, supportive technical and implementation staff, attitudes of policy makers and users, presence of supportive policies, financial resources, infrastructure, and organizational and individual readiness to adopt innovations. Future research is needed to explore the facilitators and barriers to implementing health information systems such as an LIS in LMICs.

Acknowledgments
This work has been supported by United States President’s Emergency Plan for AIDS Relief (PEPFAR) through the US Centers for Disease Control and Prevention under the terms of grant #NU2GGH001968-05-00 awarded to the University of Washington. We thank the Ministry of Health, Public Hygiene, and Universal Health Coverage (M SHPCMU) in Côte d’Ivoire and all implementing laboratories for the continuous collaboration. We thank Misti McDowell and Jennifer Gilvydis at International Training and Education Center for Health (I-TECH) for editorial review; Marisa Van Osdale at I-TECH for program management; Ahoua Koné at the University of Washington for her guidance and insight on the Ivorian context; and the project interns and data consultants at I-TECH CIV. The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the funding agencies.

Data Availability
The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest
CI-H is the product owner of OpenELIS Global, a public property and a global good. He made unique and substantial contributions to the conceptualization of this study and provided important information about the background and features of Open-Source Enterprise-Level Laboratory Information System (OpenELIS) in Côte d’Ivoire, but he did not partake in or influence the data collection, analysis, and interpretation of the evaluation.

Multimedia Appendix 1
Laboratory test result data timeliness, completeness, and validity at individual laboratories from 48 weeks before OpenELIS implementation to 72 weeks after at clinical laboratories in Côte d’Ivoire.

References


19. Wikimedia Commons. URL: https://upload.wikimedia.org/wikipedia/commons/archive/7/71/202304201070846%21Regions _and_Autonomous_Districts_of_C%C3%B4te_d%27Ivoire.svg [accessed 2024-03-12]

20. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-03-12]


Abbreviations

CDC: Centers for Disease Control and Prevention
CD4: clusters of differentiation 4
CNESVS: Côte d'Ivoire Comité National d'Ethique des Sciences de la Vie et de la Santé
ISO: International Organization for Standardization
I-TECH: International Training and Education Center for Health
LIS: laboratory information system
LMIC: low- and middle-income country
MSHPCMU: Ministry of Health, Public Hygiene, and Universal Health Coverage
OpenELIS: Open-Source Enterprise-Level Laboratory Information System
OR: odds ratio
SLIPTA: Stepwise Laboratory Quality Improvement Process Toward Accreditation

Edited by A Mavragani; submitted 29.06.23; peer-reviewed by P Moorehouse, JR Medina, C Zhao; comments to author 14.12.23; revised version received 04.01.24; accepted 23.01.24; published 20.03.24.

Please cite as:
Spatial Access to Continuous Maternal and Perinatal Health Care Services in Low-Resource Settings: Cross-Sectional Study

Qin Li¹,², PhD; Elsa Kanduma³, PhD; Isaías Ramiro³, PhD; Dong (Roman) Xu⁴,⁵, PhD; Rosa Marlene Manjate Cuco⁶, PhD; Eusebio Chaquisse⁶, PhD; Yili Yang¹,², PhD; Xiuli Wang¹,²*, PhD; Jay Pan¹,⁷*, PhD

¹HEOA Group, West China School of Public Health, West China Fourth Hospital, Sichuan University, Chengdu, China
²Institute for Healthy Cities, West China Research Center for Rural Health Development, Sichuan University, Chengdu, China
³Comité para a Saúde de Moçambique, Maputo City, Mozambique
⁴Acacia Lab for Implementation Science, Center for World Health Organization Studies, School of Health Management, Southern Medical University, Guangzhou, China
⁵SMU Institute for Global Health, Dermatology Hospital, Southern Medical University, Guangzhou, China
⁶Mozambique Ministry of Health and Faculty of Medicine, Eduardo Mondlane University, Maputo City, Mozambique
⁷China Center for South Asian Studies, Sichuan University, Chengdu, China
*these authors contributed equally

Corresponding Author:
Jay Pan, PhD
HEOA Group
West China School of Public Health, West China Fourth Hospital
Sichuan University
17 People’s South Road
Chengdu, 610041
China
Phone: 86 1 861 579 9937
Email: panjie.jay@scu.edu.cn

Abstract

Background: Maternal and perinatal health are fundamental to human development. However, in low-resource settings such as sub-Saharan Africa (SSA), significant challenges persist in reducing maternal, newborn, and child mortality. To achieve the targets of the sustainable development goal 3 (SDG3) and universal health coverage (UHC), improving access to continuous maternal and perinatal health care services (CMPHS) has been addressed as a critical strategy.

Objective: This study aims to provide a widely applicable procedure to illuminate the current challenges in ensuring access to CMPHS for women of reproductive age. The findings are intended to inform targeted recommendations for prioritizing resource allocation and policy making in low-resource settings.

Methods: In accordance with the World Health Organization guidelines and existing literature, and taking into account the local context of CMPHS delivery to women of reproductive age in Mozambique, we first proposed the identification of CMPHS as the continuum of 3 independent service packages, namely antenatal care (ANC), institutional delivery (ID), and postnatal care (PNC). Then, we used the nearest-neighbor method (NNM) to assess spatial access to each of the 3 service packages. Lastly, we carried out an overlap analysis to identify 8 types of resource-shortage zones.

Results: The median shortest travel times for women of reproductive age to access ANC, ID, and PNC were 2.38 (IQR 1.38-3.89) hours, 3.69 (IQR 1.87-5.82) hours, and 4.16 (IQR 2.48-6.67) hours, respectively. Spatial barriers for women of reproductive age accessing ANC, ID, and PNC demonstrated large variations both among and within regions. Maputo City showed the shortest travel time and the best equity within the regions (0.46, IQR 0.26-0.69 hours; 0.74, IQR 0.47-1.04 hours; and 1.34, IQR 0.83-1.85 hours, respectively), while the provinces of Niassa (4.07, IQR 2.41-6.63 hours; 18.20, IQR 11.67-24.65 hours; and 7.69, IQR 4.74-13.05 hours, respectively) and Inhambane (2.69, IQR 1.49-3.91 hours; 4.43, IQR 2.37-7.16 hours; and 10.76, IQR 7.73-13.66 hours, respectively) lagged behind significantly in both aspects. In general, more than 51% of the women of reproductive age, residing in 83.25% of Mozambique’s land area, were unable to access any service package of CMPHS in time (within 2 hours), while only about 21%, living in 2.69% of Mozambique’s land area, including Maputo, could access timely CMPHS.
Conclusions: The spatial accessibility and equity of CMPHS in Mozambique present significant challenges in achieving SDG3 and UHC, especially in the Inhambane and Niassa regions. For Inhambane, policy makers should prioritize the implementation of a decentralization allocation strategy to increase coverage and equity through upgrading existing health care facilities. For Niassa, the cultivation of well-trained midwives who can provide door-to-door ANC and PNC at home should be prioritized, with an emphasis on strengthening communities’ engagement. The proposed 2-step procedure should be implemented in other low-resource settings to promote the achievement of SDG3.

(JMIR Public Health Surveill 2024;10:e49367) doi:10.2196/49367

KEYWORDS
continuous maternal and perinatal health care services; sub-Saharan Africa; SSA; spatial access; resource allocation; low-resource settings

Introduction

Access to continuous maternal and perinatal health care services (CMPHS) has been proposed by the World Health Organization (WHO) as an essential strategy to improve maternal and child health [1]. Instead of focusing on single interventions, CMPHS emphasizes packages of interventions delivered throughout the lifecycle at all levels of health care facilities in the health delivery system [2]. The provision of CMPHS is also a priority for achieving universal health coverage (UHC) and is listed as a key indicator for monitoring the progress of maternal, newborn, and child outcomes [3]. It has been widely recognized that CMPHS can be understood as the integration of 3 service delivery packages: antenatal care (ANC), institutional delivery (ID), and postnatal care (PNC) [4]. ANC facilitates the prompt detection and management of complications among pregnant women [5]. Research suggests that even a 7% increase in ANC coverage might save 160,000 newborn lives annually in Africa [6]. ID ensures appropriate equipment and supplies are available on site or through immediate referral to a higher-level facility if necessary [7]. As suggested by the literature, perinatal mortality is 21% lower for facility-based deliveries compared to home deliveries. In the best settings, up to 14 perinatal deaths might be averted per 1000 births if the obstetric deliveries occurred at facilities instead of at homes [8]. PNC provides great opportunities to check for dangerous signs and symptoms associated with poor health outcomes in babies [9]. The importance of PNC has been documented, with an estimation that if PNC rates reach 90% in sub-Saharan Africa (SSA), neonatal deaths could be averted by 10%-27% [10].

Mozambique represents a low-resource setting of the type prevalent in the SSA region, where the world’s highest maternal and neonatal mortality persists as a critical issue (533 maternal deaths per 100,000 live births vs 211 per 100,000 worldwide and 27 neonatal deaths per 1000 live births vs 18 per 1000 worldwide). Timely provision of CMPHS could prevent a significant proportion of these deaths [11-13]. Several recent studies have examined health care accessibility in low-resource settings, including emergency obstetric and newborn care, skilled birth attendance and caesarean delivery services, and primary health care [14-18]. However, efforts to assess the spatial accessibility of CMPHS remain insufficient [19] despite widespread recognition that the effectiveness of the 3 service packages is greatly influenced by health care services achieved in the previous stages. Moreover, there is a lack of targeted suggestions on resource investment and policy making in this area.

Among various barriers preventing women of reproductive age from obtaining timely access to CMPHS, spatial accessibility (geographic proximity) has become one of the primary barriers, particularly in low-resource settings [20]. As a crucial aspect of accessibility, use, and the provision of health care services to the population group in need [21], assessment of spatial accessibility has been carried out worldwide and been incorporated into policy making in many countries. However, there is a paucity of studies assessing the spatial accessibility of CMPHS by integrating multiple indicators into 3 separate service packages in low-resource settings.

As the deadline for sustainable development goal 3 (SDG3) approaches, and the fundamental need for maternal and prenatal health care persists, the efficient allocation of limited health care resources, especially in low-resource settings, is of great importance. Mozambique, one of the world’s poorest and most underdeveloped countries (with a United Nations Human Development Index of 0.446 in 2021), faces significant challenges in implementing SDG3, particularly concerning CMPHS indicators [22]. This study uses Mozambique as a case study, aiming to provide a widely applicable procedure for assessing the spatial access to CMPHS and to offer evidence-based suggestions for policy makers that will guide the prioritization of health resource allocation to improve CMPHS accessibility and maternal, newborn, and child health outcomes in low-resource settings.

Methods

Study Area

Mozambique, located on the southeast coast of Africa, is divided into 11 provinces, including 1 provincial city (Multimedia Appendix 1, Figure S1A). According to data announced by the General Population and Housing Census 2017 in Mozambique [23], the estimated total population was 30.07 million, with the majority living in rural areas and generally sparsely and unevenly distributed (Multimedia Appendix 1, Figure S1B). Nearly 80% of people living in poverty live in rural areas distant from basic public facilities [24]. The country’s overall economy remains extremely underdeveloped (gross domestic product <US $1.00/person/day); the highest average consumption level in Maputo City is 6.9 times higher than the lowest, in Zambezia province (Multimedia Appendix 1, Figure S1C) [25].
According to a report from the Mozambique Ministry of Health, the country’s maternal mortality rate was as high as 451 deaths per 100,000 live births, and the infant mortality rate was 67.4 deaths per 1000 live births in 2020 [26]. Factors such as long travel distances from residential locations to health care facilities, lack of available transportation, and poorly constructed transportation networks contribute to significant service delays for women of reproductive age accessing CMPHS [15,27].

Data Sources

Data on 2 perspectives were needed for identifying indicators reflective of the delivery of the 3 service packages in Mozambique for calculating spatial accessibility. Following the existing literature and considering both representativeness and implementability, 2 categories of data were used for indicator identification and 3 categories of data were used to facilitate the analysis of spatial accessibility, including supply side data (health care facility data), demand side data (distribution data for women of reproductive age), and environmental data (for calculating geospatial barriers) [28-30]. The data sources are shown in detail in Table 1.

Table 1. Description and sources of data use in the 2 steps of the proposed procedure.

<table>
<thead>
<tr>
<th>Perspectives</th>
<th>Data type</th>
<th>Data description</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifying indicators representing 3 service packages of CMPHS&lt;sup&gt;a&lt;/sup&gt; in Mozambique</td>
<td>For representativeness: global guidelines</td>
<td>Guidelines for essential practice in maternal, newborn, and child health and pregnancy, childbirth, postpartum, and newborn care</td>
<td>World Health Organization guidelines [31,32]</td>
</tr>
<tr>
<td>For implementability: Mozambique-specific data</td>
<td>Continuous maternal and perinatal health care services (antenatal care, institutional delivery, and postnatal care)</td>
<td>The questionnaire of the National Survey on Infrastructure, Equipment, Human Resources and Health Services 2018</td>
<td></td>
</tr>
<tr>
<td>Calculating spatial accessibility of CMPHS in Mozambique</td>
<td>Supply side: health care facility data</td>
<td>Geographic coordinates (longitude, latitude) of all health care facilities</td>
<td>Public health facilities in sub-Saharan Africa; Figshare data set [33] and Google Maps</td>
</tr>
<tr>
<td>Demand side: population data</td>
<td>Address of all health care facilities</td>
<td>The questionnaire of the National Survey on Infrastructure, Equipment, Human Resources and Health Services 2018</td>
<td></td>
</tr>
<tr>
<td>Environmental data</td>
<td>Data on number of women of reproductive age at the provincial level (2018)</td>
<td>Statistical Yearbook 2018 of Mozambique [34]</td>
<td></td>
</tr>
<tr>
<td>Environmental data</td>
<td>Raster data on spatial distribution of pregnancies at 1×1 km&lt;sup&gt;2&lt;/sup&gt; resolution (2015)</td>
<td>WorldPop provides open geospatial data on distribution of pregnancies in Mozambique [35]</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>CMPHS: continuous maternal and perinatal health care services.

Data Pretreatment and Analysis

A detailed flowchart describing the 2-step procedure is shown in Figure 1.
Step 1: Identification of Indicators Representing the 3 Service Packages of CMPHS

The selection of indicators was based on 2 main principles. The first principle was accordance with maternal, newborn, and child health (MNCH) guidelines, as well as the pregnancy, childbirth, postpartum, and newborn care (PCPNC) guidelines proposed by the WHO. The second principle was to fully reflect the region-specific characteristics of CMPHS delivery in the study area.

In adhering to the 2 main principles, recommended and available indicators were first filtered to identify a group of separate indicators from each individual service package. The available indicators were extracted from Mozambique’s Survey on Infrastructure, Equipment, Human Resources and Health Services 2018 (Multimedia Appendix 1, Table S1) [37,38]. Then, with consideration of both avoiding fragmented services and ensuring the quality of CMPHS, as well as extensive reference to existing literature, [39], the filtered indicators were further screened and integrated to the 3 service packages. ANC includes 7 integrated indicators, ID includes 8 integrated indicators, and PNC includes 10 integrated indicators (Multimedia Appendix 1, Table S2).

Step 2: Calculation of Spatial Access to CMPHS (Data Pretreatment)

For the supply side data, 3 levels of health care facilities were included, consisting of Mozambique’s health care delivery system: primary (urban and rural health centers and community health posts), secondary (rural, district, and general hospitals), and tertiary (central and provincial, specialized, and military hospitals). Combined with results from step 1, the capability of delivering the 3 service packages of CMPHS was identified for every health care facility.

For demand side data, gridded estimates of pregnancies were obtained at 1×1 km² spatial resolution from the WorldPop database for 2015, which served to weight the spatial distribution of the demand side data. Then, data on the nationally announced provincial-level aggregated number of women aged 15-49 years in Mozambique was used to adjust the spatial distribution of pregnant women at a resolution of 1×1 km² to generate a more accurate and up-to-date demand distribution, as process generally used in related research [40-42].

The following formula was used to calculate the spatial distribution of women of reproductive age at 1×1 km² resolution [40], in which \( P_i \) is the number of women of reproductive age of grid \( i \) in the province or provincial capital \( j \), and \( G_i \) is the corresponding WorldPop value for pregnancy density of grid \( i \). \( G_i \) is the sum of the grid values for WorldPop in province or provincial capital \( j \), and \( T_i \) is the actual number of women of reproductive age in province or provincial capital \( j \) from the Statistical Yearbook 2018 of Mozambique [23]:
Analysis of Spatial Accessibility (Establishment of Geo-Database)

ArcGIS (version 10.5; Esri, Inc) was used to calculate spatial access to CMPHS in this study. A geodatabase incorporating supply side data, demand side data, and administrative boundary data was first generated [43].

Calculation of the Shortest Travel Time Needed to Reach ANC, ID, and PNC

The nearest-neighbor method (NMM) was used to calculate the shortest travel time necessary to access ANC, ID, and PNC [44]. For each grid of women of reproductive age, only the nearest health care facility was considered [41]. The travel mode of walking at a speed of 4 km/h for spatial accessibility assessment and a threshold of 2 hours were used as criteria for timely access (Multimedia Appendix 1, Table S3) [5]. Additionally, the travel time along the straight-line distance was used to compare the travel time criterion [18,45]. The shortest travel time needed to reach ANC, ID, and PNC was calculated separately, and the average travel time was calculated for each province or provincial capital.

Overlap Analysis of Spatial Coverage of CMPHS

An overlap analysis was applied to identify multilevel health care access zones. On the basis of identifying areas able to access ANC, ID, and PNC in a timely manner (ie, within a 2 h threshold), 3 layers representing timely access to ANC, ID, and PNC, respectively, were overlapped to generate 8 types of multilevel health care access zones. Then, percentages for area and women of reproductive age covered within each type of multilevel health care access zone were calculated.

Ethical Considerations

The study was approved by the Swiss Confederation, represented by the Swiss Federal Department of Foreign Affairs, acting through the Swiss Agency for Development and Cooperation Global Cooperation Department and Southern Medical University of the People’s Republic of China (81067392). Core data used in this study were derived from the National Survey on Infrastructure, Equipment, Human Resources and Health Services 2018, which was collected between February and April 2018 through a cross-sectional survey of health care facilities. The survey included a census of all health facilities in the country and collected data through interviews with key informants at health care facilities. The core standard Service Availability and Readiness Assessment (SARA) questionnaire of the WHO was used and adapted to the Mozambique context [37,38]. The SARA questionnaire and linked protocols were granted approval by Mozambique’s Ministry of Health. More information about the primary data collection procedures is available in the Mozambique SARA 2018 report [38]. Permission to use the data was obtained from Mozambique’s Ministry of Health by coauthors from Mozambique. Other data, including population distribution and administrative boundaries, were extracted from publicly available data sets. No human participants or animals were involved in the study.

Results

Identification and Coverage of Indicators Representing CMPHS

The identified indicators for the 3 service packages and the percentage of health care facilities capable of providing each service (indicators) are shown in Table 2. Among all health care facilities (N=1542), 784 were able to provide ANC, 365 were able to provide ID, and 299 were able to provide PNC. Most health care facilities within the study area managed to cover all services listed in each individual service package, with some indicators having fairly good penetration. The provision of intermittent preventive treatment in pregnancy for malaria, a component of the ANC service package, was offered in 1408 (91.31%) health care facilities. Conversely, some indicators showed suboptimal coverage; for example, emergency obstetric care was available in only 504 (32.68%) health care facilities and injectable antibiotics intended for neonatal sepsis were available in just 766 (49.68%) health care facilities (Table 2).
Table 2. Identified indicators of the 3 service packages and the percentage of health care facilities able to provide each service (indicators) based on global guidelines for maternal, newborn, and child health, as well as pregnancy, childbirth, postpartum, and newborn care proposed by the World Health Organization and data extracted from the National Survey on Infrastructure, Equipment, Human Resources and Health Services 2018.

<table>
<thead>
<tr>
<th>Phases and key indicators</th>
<th>Health care facilities providing antenatal care, institutional delivery, and postnatal care (N=1542), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antenatal care (n=7 indicators; n=784 health care facilities)</strong></td>
<td></td>
</tr>
<tr>
<td>Iron supplementation</td>
<td>1070 (69.39)</td>
</tr>
<tr>
<td>Folic acid supplementation</td>
<td>1035 (67.12)</td>
</tr>
<tr>
<td>Tetanus toxoid vaccination</td>
<td>1364 (88.46)</td>
</tr>
<tr>
<td>Monitoring for hypertensive disorder of pregnancy</td>
<td>1198 (77.69)</td>
</tr>
<tr>
<td>IPTp(^a) for malaria</td>
<td>1408 (91.31)</td>
</tr>
<tr>
<td>HIV counseling and testing for HIV-positive pregnant women</td>
<td>1369 (88.78)</td>
</tr>
<tr>
<td>Antiviral treatment for HIV-positive pregnant women</td>
<td>1347 (87.35)</td>
</tr>
<tr>
<td><strong>Institutional delivery (n=8 indicators; n=365 health care facilities)</strong></td>
<td></td>
</tr>
<tr>
<td>Monitoring of labor with partograph</td>
<td>1273 (82.56)</td>
</tr>
<tr>
<td>Parenteral administration of oxytocin</td>
<td>1243 (80.61)</td>
</tr>
<tr>
<td>Assisted vaginal delivery</td>
<td>1125 (72.96)</td>
</tr>
<tr>
<td>Manual removal of placenta</td>
<td>1103 (71.53)</td>
</tr>
<tr>
<td>Antibiotics for preterm</td>
<td>849 (55.06)</td>
</tr>
<tr>
<td>Blank partograph</td>
<td>1204 (78.08)</td>
</tr>
<tr>
<td>Parenteral administration of magnesium sulfate</td>
<td>1079 (69.97)</td>
</tr>
<tr>
<td>Emergency obstetric care</td>
<td>504 (32.68)</td>
</tr>
<tr>
<td><strong>Postnatal care (n=10 indicators; n=299 health care facilities)</strong></td>
<td></td>
</tr>
<tr>
<td>Immediate and exclusive breastfeeding</td>
<td>1313 (85.15)</td>
</tr>
<tr>
<td>Thermal protection</td>
<td>1293 (83.85)</td>
</tr>
<tr>
<td>Hygienic cord care</td>
<td>1307 (84.76)</td>
</tr>
<tr>
<td>Neonatal resuscitation</td>
<td>1122 (72.76)</td>
</tr>
<tr>
<td>Kangaroo mother care</td>
<td>1094 (70.95)</td>
</tr>
<tr>
<td>Injectable antibiotics for neonatal sepsis</td>
<td>766 (49.68)</td>
</tr>
<tr>
<td>HIV counseling and testing for infants born to HIV-positive women</td>
<td>1351 (87.61)</td>
</tr>
<tr>
<td>ARV(^b) prophylaxis to newborns of HIV-positive pregnant women</td>
<td>1345 (87.22)</td>
</tr>
<tr>
<td>HIV-positive infant and young child feeding counseling</td>
<td>1367 (88.65)</td>
</tr>
<tr>
<td>BCG(^c) vaccine</td>
<td>886 (57.46)</td>
</tr>
</tbody>
</table>

\(^a\)IPTp: intermittent preventive treatment in pregnancy.
\(^b\)ARV: AIDS-related virus.
\(^c\)BCG: bacillus Calmette-Guérin.

Distribution of Health Care Facilities Providing CMPHS

There were 1542 health care facilities in total across the study area, among which were 1490 primary-level health care facilities (urban and rural health centers or community health posts), 43 secondary-level health care facilities (rural, district, and general hospitals), and 9 tertiary-level health care facilities (central and provincial hospitals or specialized and military hospitals). Among the primary-level facilities, 51.48% (767/1490), 21.34% (318/1490), and 18.46% (275/1490) were able to provide ANC, ID, and PNC, respectively. For the secondary facilities, the respective percentages were 37% (16/43), 91% (39/43), and 23% (17/43). For the tertiary-level health care facilities, the respective percentages were 11% (1/9), 89% (8/9), and 78% (7/9).

Figure 2 shows the spatial distribution of the 3 levels of health care facilities able to deliver the 3 service packages. The spatial distribution of different levels of health care facilities able to provide ANC, ID, and PNC, respectively, is shown in Multimedia Appendix 1, Figure S2.
Figure 2. Capability of delivering the 3 service packages of continuous maternal and perinatal health care services was identified for every health care facility on the supply side. The WorldPop data set was used to provide spatial distribution of pregnant women at a 1×1 km² resolution on the demand side. (A) The spatial distribution of women of reproductive age (people/0.01 km²). (B) The spatial distribution of the 3 levels of health care facilities in Mozambique in 2018. Maputo and Maputo City are displayed in the figures separately.

Spatial Access to The 3 Service Packages of CMPHS

In Mozambique, the median shortest travel times needed to access ANC, ID, and PNC were 2.38 (IQR 1.38-3.89) hours, 3.69 (IQR 1.87-5.28) hours, and 4.16 (IQR 2.48-6.67) hours, respectively. The shortest travel times required to reach the nearest health care facilities offering these services varied significantly across different regions. For ANC services, the shortest travel time ranged from 0.46 (IQR 0.26-0.69) hours in Maputo City to 4.95 (IQR 3.03-7.21) hours in Manica province. For ID services, it ranged from 0.74 (IQR 0.47-1.04) hours in Maputo City to 18.20 (IQR 11.67-24.65) hours in Niassa province. For PNC services, the shortest travel time ranged from 1.34 (IQR 0.83-1.85) hours in Maputo City to 10.76 (IQR 7.53-13.66) hours in Inhambane province. Except for Manica province, the shortest travel time needed to access ID was greater than for ANC in all other provinces. The shortest travel time spent to access ID was found to be distinctly longer than that for PNC in 3 provinces, namely Gaza, Maputo, and Niassa (Table 3).

The results for coverage of different services in Figure 3 shows that women of reproductive age living in Maputo City were able to obtain timely ANC, ID, and PNC. Conversely, multiple provinces with lower population density were recognized as underserved areas, including the provinces of Niassa, Cabo Delgado, Gaza, and Inhambane, especially for the provision of ID and PNC.
Table 3. The shortest travel time spent to reach antenatal care, institutional delivery, and postnatal care was calculated for different regions using the nearest-neighbor method with a criterion of 2 hours’ walking, which was set under the scenario that demanders walked along a straight line with a speed of 4 km/hour.

<table>
<thead>
<tr>
<th>Region</th>
<th>Time to reach antenatal care (hours)</th>
<th>Time to reach institutional delivery (hours)</th>
<th>Time to reach postnatal care (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (IQR)</td>
<td>Maximum</td>
<td>Median (IQR)</td>
</tr>
<tr>
<td>Mozambique</td>
<td>2.38 (1.38-3.89)</td>
<td>30.38</td>
<td>3.69 (1.87-5.28)</td>
</tr>
<tr>
<td>Cabo Delgado</td>
<td>4.01 (2.39-6.82)</td>
<td>25.71</td>
<td>4.89 (2.89-8.40)</td>
</tr>
<tr>
<td>Gaza</td>
<td>2.90 (1.56-4.25)</td>
<td>20.40</td>
<td>6.09 (3.71-8.64)</td>
</tr>
<tr>
<td>Inhambane</td>
<td>2.69 (1.49-3.91)</td>
<td>16.95</td>
<td>4.43 (2.37-7.16)</td>
</tr>
<tr>
<td>Manica</td>
<td>4.95 (3.03-7.21)</td>
<td>27.48</td>
<td>3.17 (1.94-4.41)</td>
</tr>
<tr>
<td>Maputo</td>
<td>1.94 (1.03-3.02)</td>
<td>18.92</td>
<td>3.07 (1.78-4.79)</td>
</tr>
<tr>
<td>Maputo City</td>
<td>0.46 (0.26-0.69)</td>
<td>2.19</td>
<td>0.74 (0.47-1.04)</td>
</tr>
<tr>
<td>Nampula</td>
<td>2.13 (1.37-2.80)</td>
<td>10.61</td>
<td>2.81 (1.77-4.08)</td>
</tr>
<tr>
<td>Niassa</td>
<td>4.07 (2.41-6.63)</td>
<td>30.38</td>
<td>18.20 (11.67-24.65)</td>
</tr>
<tr>
<td>Sofala</td>
<td>2.11 (1.29-3.10)</td>
<td>18.16</td>
<td>4.90 (2.98-7.22)</td>
</tr>
<tr>
<td>Tete</td>
<td>3.87 (2.41-5.34)</td>
<td>18.06</td>
<td>5.70 (3.58-8.19)</td>
</tr>
<tr>
<td>Zambezia</td>
<td>2.47 (1.52-3.63)</td>
<td>12.83</td>
<td>3.78 (2.35-5.69)</td>
</tr>
</tbody>
</table>

Figure 3. The spatial coverage of antenatal care (ANC), institutional delivery (ID), and postnatal care (PNC) provided by different levels of health care facilities by walking along a straight line with a speed of 4 km/hour and the standard of 2-hour security service range. The coverage area (in terms of 2-hour travel time) for (A) ANC, (B) ID, and (C) PNC. Maputo and Maputo City are displayed in the figures separately.

Classification and Identification of Underserved Areas for CMPHS

For women of reproductive age in Mozambique, more than 21% (living in about 2.69% of Mozambique’s area) lived in type 1 areas (able to obtain timely CMPHS including ANC, ID, and PNC); more than 51% (living in 83.25% of Mozambique’s area) lived in type 8 areas (not able to obtain timely access to any service packages of CMPHS); only 27.5% (living in about 14.07% of Mozambique’s area) lived in type 2-7 areas (able to access 1 or 2 types of services of CMPHS). The second highest percentage of women of reproductive age fell within type 5 areas, indicating that approximately 10% (living in about 7.42% of Mozambique’s area) were able to obtain ANC services in a timely manner. The third highest percentage of women of reproductive age fell within type 2 areas, meaning that more than 9% (living in about 2.19% of Mozambique’s area) were able to reach ANC and PNC in a timely manner (Table 4).
Three layers representing timely access to antenatal care (ANC), institutional delivery (ID), and postnatal care (PNC), respectively, were overlapped to generate 8 types of multilevel health care access zones. The percentages by land area and by women of reproductive age covered within each type of multilevel health care access zone were calculated using an overlap analysis with a criterion of 2-hour security service range.

<table>
<thead>
<tr>
<th>Type</th>
<th>Multilevel health care access zones</th>
<th>Land area, %</th>
<th>Population, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Able to obtain timely CMPHS, including ANC, ID, and PNC</td>
<td>2.69</td>
<td>21.1</td>
</tr>
<tr>
<td>2</td>
<td>Able to obtain timely ANC and ID</td>
<td>2.19</td>
<td>9.7</td>
</tr>
<tr>
<td>3</td>
<td>Able to obtain timely ANC and PNC</td>
<td>1.61</td>
<td>3.79</td>
</tr>
<tr>
<td>4</td>
<td>Able to obtain timely ID and PNC</td>
<td>0.95</td>
<td>1.63</td>
</tr>
<tr>
<td>5</td>
<td>Able to obtain timely ANC</td>
<td>7.42</td>
<td>9.96</td>
</tr>
<tr>
<td>6</td>
<td>Able to obtain timely ID</td>
<td>1.03</td>
<td>1.47</td>
</tr>
<tr>
<td>7</td>
<td>Able to obtain timely PNC</td>
<td>0.87</td>
<td>0.95</td>
</tr>
<tr>
<td>8</td>
<td>Not able to obtain timely CMPHS, including ANC, ID, and PNC</td>
<td>83.25</td>
<td>51.4</td>
</tr>
</tbody>
</table>

aCMPHS: continuous maternal and perinatal health care services.

As can be seen in Figure 4 and Multimedia Appendix 1, Figure S3, the coverage of CMPHS was low in Mozambique, especially in the provinces of Inhambane, Niassa, and Gaza. Maputo City was the only region that was able to provide timely CMPHS for women of reproductive age.
Figure 4. Overlay analysis of spatial access among women of reproductive age to continuous maternal and perinatal health care services (CMPHS) with a criterion of a 2-hour secure service range. Multilevel health care access zones are areas able to access all 3 service packages, 2 of the 3 service packages, only 1 of the 3 service packages, or none of the 3 service packages. Green is used to represent full access to CMPHS (i.e., a good service type, namely type 1), shades of yellow are used to represent access to any 2 types of service (i.e., medium service types, namely type 2, type 3, and type 4), shades of red are used to represent access to any 1 type of service (i.e., poorer service types, namely type 5, type 6, and type 7), and the blank areas represent no access to any type of services (i.e., the poorest service type, namely type 8). Maputo and Maputo City are displayed in the figure separately. ANC: antenatal care; ID: institutional delivery; PNC: postnatal care.
Discussion

Principal Findings

Our study proposed a 2-step procedure to assess the spatial accessibility of CMPHS and verified its applicability through a case study in Mozambique. The results first showed spatial access to different types of CMPHS and disparities among different service packages. By using an overlay analysis, this procedure could identify 7 types of underserved areas and the detailed spatial distribution of existing well-served areas. The results from this procedure have evidence-based implications to inform decision-making processes related to health resource allocation, particularly in low-resource settings confronted with similar challenges. This can ultimately aid in improving maternal and newborn health outcomes. The necessity of enhancing spatial accessibility as a critical solution to improve coverage of CMPHS is also highlighted. In the specific case of Mozambique, only 21% of women of reproductive age could access CMPHS in a timely manner, and these women were predominately clustered in the most-developed region, Maputo City. This highlights that spatial access to CMPHS remains a significant challenge and that improvements in the coverage of ID and PNC are urgently needed.

Large disparities among regions in accessing CMPHS were also a critical issue. Inequities in CMPHS exist at the national, provincial, and district levels and always present fundamental barriers to progress, particularly among the most disadvantaged population groups [46]. This has been validated by our findings that the provinces of both Niassa and Cabo Delgado, which have lower population densities, tended to be more underserved, especially in accessing ID and PNC service packages. Our study also found that Maputo City was the only administrative unit that was capable of providing effective access to all 3 service packages. We assume this is mainly because Maputo City is the capital of Maputo and the economic and social development center of Mozambique, so its development status far exceeds all the other areas [47]. In addition, the frequent occurrence of natural disasters, which greatly hinders health care service delivery, further adds to health inequities between the southern and northern regions in Mozambique, leading to lower rates of both ANC and ID visits in the southern region, corresponding to 74 additional maternal deaths and 726 additional deaths of children under the age of 1 month by the end of 2021 [48].

In the context of Mozambique’s poor accessibility and low equity in health care, the implementation of a decentralized allocation strategy to address the inequity issue has been recognized as an effective measure to promote both accessibility and coverage of CMPHS, which can take China’s successful experiences as an example [49,50]. China has made remarkable achievements and passed milestones in achieving SDGs through implementing a decentralized allocation strategy to improve accessibility and equality of health care services. China clearly indicated that women and children were target groups in the cooperation for global health in the Belt and Road Initiative in 2017 and implemented 100 foreign assistance programs for women and children’s health in low-income countries, so decentralized allocation strategies from China can help low-income countries improve their own CMPHS [51]. For example, strategies for the promotion of primary health care, along with the optimization of a hierarchical health care delivery system intended for maternal and child health care, were implemented. According to these strategies, integrated service packages delivered via the collaborative efforts of tertiary hospitals and primary maternal health care facilities proved to be an effective solution to strengthen the capacity of local health care facilities, as well as their corresponding referral systems, which resulted in increased in-hospital births and significantly reduced maternal mortality [50,52,53]. Furthermore, community-based interventions with components to engage or mobilize communities have been increasingly implemented to improve maternal and newborn health outcomes in low-resource settings [54-56]. For example, training programs should be provided to cultivate highly skilled health care professionals among local health care facilities [55,57].

Comparison With Prior Work

Few studies have assessed the spatial accessibility of health care resources in Mozambique. A previous study reported that 67.3% of Mozambique’s population lived outside a 1-hour travel distance from primary health care centers [58]. However, spatial barriers to accessing health care services are gaining increasing attention. According to the Statistical Yearbook 2018 of Mozambique, the average travel distance needed to access a health care facility for people living in remote areas was 12.3 km [34]. The 2018-2019 Mozambique Humanitarian Response Plan reported that about 50% of the population lived more than 20 km away from the nearest health care facility [59]. Given the national focus on maternal and perinatal health, studies assessing spatial access to CMPHS have been carried out in the context of low-resource settings [14,17,60]. Previous findings suggest that 49.8% of women of reproductive age in Mozambique lived outside a 2-hour travel area from the nearest hospital for emergency services [61]. In southern Mozambique, 46% of pregnant women were able to access the nearest primary health care facility within 1 hour by walking, while 64% of women living in the region could receive life-saving service delivery within 2 hours [46]. While most existing literature has only assessed spatial access to single services, our results are consistent with previous findings that spatial access to maternal and perinatal health care services is poor in Mozambique. Moreover, our study further considered 3 separate service packages, representing important stages in the continuity of maternal and perinatal health care services, revealing spatial access to CMPHS in a straightforward manner.

Public Health Implications

Our findings provide important evidence on the development of tailored, country-specific policies and interventions to address the geographical disparities of CMPHS in low-resource settings, which can support the development of resource allocation strategies based on the priority of CMPHS delivery. Using Mozambique as a case study of a low-resource setting, we recommend that public health policy makers and decision-makers draw from successful lessons such as decentralized allocation strategies. Through a comprehensive understanding of the geographical barriers to access to CMPHS,
these strategies can quickly and efficiently improve maternal and neonatal health outcomes and contribute to the achievement of SDG3 and UHC.

**Limitations**

This study does have several limitations. First, we set a scenario of women of reproductive age walking to health care facilities to reflect the most basic requirement of the vast number of vulnerable people in Mozambique. Other travel modes were not considered due to data limitations and the complexity of factors influencing the choice of travel mode. Specific travel thresholds for different levels of health care facilities were also not set due to a lack of reference data. Second, we used the spatial distribution of women of reproductive age as demand-side data instead of pregnancies due to data availability. Third, the integration of indicators representing the 3 service packages failed to take into account the type of health care facilities providing each service. Consequently, the continuum of services in different health care facilities was not further analyzed.

**Conclusions**

Using Mozambique as a case study, this research used a 2-step procedure to assess the accessibility of CMPHS and proposes targeted policy suggestions for improving health outcomes. Further studies could be conducted, especially in low-resource settings, with reference to local context, and thus inform resource allocation–related decision-making procedures. In turn, this could potentially contribute to the improvement of global maternal and neonatal health, the realization of SDG3, and the attainment of UHC.

**Acknowledgments**

The authors are grateful to staff from the Mozambique National Institute of Health for providing data on health care facilities in Mozambique.

**Data Availability**

Data from the National Survey on Infrastructure, Equipment, Human Resources and Health Services 2018 can be accessed with approval from Mozambique’s Ministry of Health. Other data used in this study come from publicly available data sets with sources listed in the main text.

**Authors’ Contributions**

QL and XW designed the study. QL, XW, and JP designed, had full access to, and verified the underlying data reported in the study. QL performed the data analysis and wrote the first draft with supervision from XW and JP. EK, IR, RMMC and EC provided suggestions on the data processing. DRX and YY helped revise the draft. All authors contributed to interpretation of data, revised the article critically for important intellectual content, and approved the final version of the manuscript.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Supplementary Tables S1-S3 and Figures S1-S3. [DOCX File, 2186 KB - publichealth_v10i1e49367_app1.docx ]

**References**


7. Delivering for women: improving maternal health services to save lives. UNICEF. URL: https://data.unicef.org/resources/delivering-for-women/ [accessed 2024-06-05]


32. Improving the quality of care for maternal, newborn and child health: implementation guide for national, district and facility levels. World Health Organization. URL: https://iris.who.int/handle/10665/353738 [accessed 2024-06-05]
36. OpenStreetMap. URL: https://www.openstreetmap.org/ [accessed 2024-06-05]


Abbreviations

ANC: antenatal care
CMPHS: continuous maternal and perinatal health care services
ID: institutional delivery
MNCH: maternal, newborn, and child health
NNM: nearest-neighbor method
PCPNC: pregnancy, childbirth, postpartum and newborn care
PNC: postnatal care
SARA: Service Availability and Readiness Assessment
SDG3: sustainable development goal 3
SSA: sub-Saharan Africa
UHC: universal health coverage
WHO: World Health Organization

©Qin Li, Elsa Kanduma, Isaías Ramiro, Dong (Roman) Xu, Rosa Marlene Manjate Cuco, Eusebio Chaquisse, Yili Yang, Xiuli Wang, Jay Pan. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 18.07.2024. This is
an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Variations in Unmet Health Care Needs by Perceptions of Social Media Health Mis- and Disinformation, Frequency of Social Media Use, Medical Trust, and Medical Care Discrimination: Cross-Sectional Study

Jim P Stimpson¹, PhD; Sungchul Park², PhD; Fernando A Wilson³, PhD; Alexander N Ortega⁴, PhD

Abstract

Background: Unmet need for health care is defined as choosing to postpone or completely avoid necessary medical treatment despite having a need for it, which can worsen current conditions or contribute to new health problems. The emerging infodemic can be a barrier that prevents people from accessing quality health information, contributing to lower levels of seeking medical care when needed.

Objective: We evaluated the association between perceptions of health mis- and disinformation on social media and unmet need for health care. In addition, we evaluated mechanisms for this relationship, including frequency of social media use, medical trust, and medical care discrimination.

Methods: Data from 3964 active adult social media users responding to the 2022 Health Information National Trends Survey 6 (HINTS 6), a nationally representative survey, were analyzed. The outcome was unmet need for medical care, defined as delaying or not getting the necessary medical care. The predictor variables were perception of social media health mis- and disinformation, frequency of social media use, level of trust in the health care system, and perceived racial and ethnic discrimination when receiving health care.

Results: Multivariable logistic regression models indicated that perception of substantial social media health mis- and disinformation (odds ratio [OR] 1.40, 95% CI 1.07 - 1.82), daily use of social media (OR 1.34, 95% CI 1.01 - 1.79), low medical trust (OR 1.46, 95% CI 1.06 - 2.01), and perceived discrimination (OR 2.24, 95% CI 1.44 - 3.50) were significantly associated with a higher likelihood of unmet need for medical care. Unmet need among adults who did not use social media daily and who did not perceive substantial mis- and disinformation (24%; 95% CI 19% - 30%) was lower compared to daily social media users who perceived substantial mis- and disinformation (38%; 95% CI 32% - 43%). Adults who perceived substantial mis- and disinformation and had low trust in health care had the highest probability of reporting unmet need (43%; 95% CI 38% - 49%) compared to the other three groups. Adults who perceived substantial mis- and disinformation and experienced medical care discrimination had a statistically significant higher probability of reporting unmet need (51%; 95% CI 40% - 62%) compared to adults who did not experience medical care discrimination and did not perceive substantial mis- and disinformation (29%; 95% CI 26% - 32%).

Conclusions: Unmet need for medical care was higher among individuals who perceived a substantial degree of social media mis- and disinformation, especially among those who used social media daily, did not trust the health care system, and experienced racial or ethnic discrimination when receiving health care. To counter the negative effects of social media mis- and disinformation on unmet need for health care, public health messaging must focus on daily social media users as well as improving trust and reducing structural racism in the health care system.

(JMIR Public Health Surveill 2024;10:e56881) doi:10.2196/56881

KEYWORDS

United States; cross-sectional study; trust; consumer health information; misinformation; disinformation; perceived discrimination; social media; unmet need; unmet needs; health care; discrimination; racism; adult; adults; medical care; frequency; multivariable regression; user; users; cross-sectional; survey; surveys; questionnaire; questionnaires; HINTS; Health Information National Trends Survey
Introduction

Unmet health care needs arise when someone with a known health issue skips or delays necessary medical treatments, such as rescheduling appointments, refusing treatments, or avoiding preventive measures [1-3]. This can happen for various reasons, including costs or difficulty accessing care [4]. Unfortunately, unmet health care needs can lead to untreated conditions becoming worse or developing new health problems. Certain populations are more likely to have unmet need for care, such as individuals with impairments, older people, marginalized racial and ethnic groups, and those without health insurance [5-8]. Systematic reviews have revealed areas that require further investigation [9,10]. For example, patient perceptions about the health care system and their own health conditions could be a significant factor contributing to unmet health care demands, underscoring a potential intervention option to improve unmet health care needs and subsequently improve health outcomes [7,9].

Theoretical Framework

The health care utilization model developed by Anderson provides a strong foundation for analyzing the complex problem of unmet need for health care [11]. This model identifies several pathways that influence health care seeking and utilization, including resources that facilitate help-seeking, perceived and assessed need for care, and predisposing qualities. Social structures, health attitudes, and demographic traits are examples of predisposing characteristics that may influence an individual's decision to use health services. The practical components of accessing care, such as the availability of health insurance, the accessibility of health care professionals, and socioeconomic position, are referred to as enabling resources. The choice to seek care is influenced by the patient’s perception of need and the assessment of medical professionals. Finally, contextual factors, such as public policies and the community environment, can influence each of these individual-level domains [7,11].

The way people receive, exchange, and act upon health information is greatly influenced by the communication infrastructure, which consists of networks, information technologies, and channels for both interpersonal and mediated communication [12]. In the Anderson model, the communication infrastructure can be viewed as a contextual factor that affects the predisposing qualities, enabling resources, and need factors. For instance, the information gleaned from many media sources may influence one’s opinions about health, thereby influencing the likelihood of seeking medical care [13,14].

The communication infrastructure also has an impact on enabling resources, which includes the practical aspects of receiving treatment. People can choose when and where to seek care by having access to trustworthy communication channels and accurate health information [14]. The quality and accessibility of health information has a direct bearing on the demand for health care, which motivates the actual use of services [12]. Unmet health care needs might arise from underestimating health needs or from mistrusting the health care system due to inaccurate or incomplete information [14]. Furthermore, poor communication can worsen gaps in access to health care, especially for racial and ethnic minorities who do not have easy access to reliable health information [15]. By integrating the communication infrastructure into the Anderson model, we can gain a deeper understanding of the complex relationship among health literacy, information availability, and health care use.

Misinformation and Disinformation

The term “infodemic” describes the ongoing dissemination of false information and health-related falsehoods, and it has emerged as a significant public health issue [16]. Research has shown that the spread of misinformation, which refers to spreading inaccurate health information, and disinformation, which refers to intentionally disseminating false information, has had a negative effect on health-related behaviors and attitudes [17-21]. The dissemination of misinformation and disinformation via social media introduces a novel extension to the traditional Anderson model by representing a contextual factor influencing individual-level predisposing, enabling, and need factors. The frequency of social media use can either facilitate or hinder health care use, depending on the quality of information accessed and the user’s level of engagement [22-24]. Disseminating inaccurate information can undermine public confidence in the health care system, which is crucial for accessing health care services and can influence the decision to seek medical attention [25-28]. Moreover, the intentional dissemination of inaccurate information and misinformation to communities of color and ethnic minorities worsens preexisting disparities and perceptions of bias in the health care system, potentially leading to higher levels of unaddressed health care needs within these populations [20,29,30]. By integrating the communication infrastructure with the Anderson model, we can understand the communication barriers that prevent people from accessing health care and develop effective ways to mitigate the negative effects of health misinformation on public health [31].

Mechanisms Linking Mis- and Disinformation and Health Care Use

The exposure to misinformation and disinformation on social media platforms can be connected to unmet health care requirements through various pathways, some of which have not been thoroughly studied in existing literature. The first mechanism pertains to the relationship between the use of social media and the level of engagement with material, specifically pertaining to false or misleading information [23]. Increased use of social media correlates with a higher probability of individuals interacting with different forms of content, including deceptive or inaccurate information [32]. Exposure to this digital information environment can modify a person’s beliefs and decision-making processes surrounding health care, potentially resulting in unmet need for medical care [33,34].

The second mechanism pertains to the confidence individuals place in health care institutions and professionals. Trust has a crucial role in influencing individuals’ decision to seek health care services and adhere to the treatments recommended by health care providers [35]. Certain groups are less likely to trust credible health institutions like government health agencies or health systems, but misinformation can influence the level of
trust regardless of group membership [36-38]. Nevertheless, the spread of false information on social media undermines this trust, leading individuals to lose confidence in the health care institution and not seek medical attention when necessary [39-42].

The third mechanism emphasizes the problem of unmet need for care among racial and ethnic minority groups, which is frequently worsened by discrimination inside the health care system [43]. Social media platforms can serve as conduits for disseminating health-related disinformation, often aimed at these specific communities [44]. The deliberate dissemination of false information, specifically aimed at racial and ethnic minoritized populations, along with experiences of discrimination, can further erode trust in health care and result in higher levels of unmet health care needs [45-50]. These mechanisms collectively contribute to additional barriers in obtaining medical treatment. To effectively tackle the challenges presented by misinformation and its influence on public health, it is crucial to better understand how these mechanisms influence health care seeking and use.

Research Objective

By integrating the communication infrastructure into the Andersen model of health care utilization, our study aimed to investigate the effects of the infodemic on the use of health care, proposing that exposure to health misinformation is associated with a higher probability of delaying or avoiding necessary medical care. In addition, our goal was to investigate the underlying mechanisms of this association, such as the frequency of social media use, the degree of trust that individuals have in the health care system, and the occurrence of racial or ethnic discrimination in health care settings. Our hypothesis is that people who perceive substantial false or misleading health information and use social media daily, have little trust in the health care system, and experience racial or ethnic discrimination when trying to use health care services are more likely to have unmet health care needs.

Methods

Data

This cross-sectional study analyzed nationally representative data from the Health Information National Trends Survey 6 (HINTS 6), which periodically surveys noninstitutionalized adults in the United States about information seeking, health communication, as well as cancer prevention attitudes and behaviors. The most recent version of the data, HINTS 6, were collected through mail- and web-based surveys from March to November 2022 with a response rate of 28.1% (34,827,468/124,058,843) [51].

Ethical Considerations

The data were publicly available and deidentified; therefore, the human research protection program of the University of Texas Southwestern Medical Center decided it did not require review by the institutional review board. Further details about the survey methodology are available from the National Cancer Institute [51]. We excluded participants who did not report using social media or had not used social media in the past year. We analyzed data from 3964 adult social media users that had used social media in the past year, with complete data on the study variables.

Measures

The binary outcome variable was unmet need for medical care, defined as delaying or not getting the medical care that the participant believed was medically necessary. The primary predictor variable was perceptions about health mis- and disinformation on social media, which was assessed by the following question: “How much of the health information that you see on social media do you think is false or misleading?” The response categories were dichotomized for ease of interpretation as substantial (ie, “a lot”) versus less than substantial (ie, “none, a little, some”), consistent with past studies [20,29,37]. The secondary predictor variables suggested by the theoretical framework included frequency of social media use, trust in the health care system, as well as perceived racial and ethnic discrimination [11-13]. Social media use was categorized as daily use versus less than daily use (eg, “never, weekly, monthly”). For medical trust, participants were asked “How much of the health care system for example, hospitals, pharmacies, and other organizations involved in health care?” The response categories were dichotomized for ease of interpretation based on prior literature as high (“very” and “somewhat”) versus low (“not at all” and “a little”) [52,53]. For medical care discrimination, participants were asked “Have you ever been treated unfairly or been discriminated against when getting medical care because of your race or ethnicity?” and the response categories were “yes” or “no.”

The Andersen model for health care utilization was used to select the control variables [11]. Predisposing factors included age (18 - 49, 50 - 64, ≥65 years), sex (male and female), marital status (married or cohabiting, formerly married, and never married), self-reported race and ethnicity (non-Latino White, non-Latino Black, non-Latino other, and Latino), education level (college degree or higher vs less than a college degree), and residence in a metropolitan versus nonmetropolitan county as designated by the US Department of Agriculture in 2013. Enabling factors included full-time employment status, feelings about household income (finding it very difficult/difficult to get by on present income, or getting by on present income, living comfortably on present income), health insurance status (any insurance or uninsured), and frequency of health provider visits (0 - 3 or more than 4 annual visits). The need for health care was measured by self-rated overall health status (excellent, very good, good, fair, and poor).

Statistical Analysis

The descriptive statistics for the study sample were calculated as survey-weighted percentages. The bivariable relationships between the outcome and the predictor variables were calculated using column percentages and adjusted Wald P values. Then, we used a stepped multivariable logistic regression model to test for the main effects of perceptions of social media health mis- and disinformation on unmet health care needs. The first step adjusted for secondary predictor variables of social media use, medical trust, and medical care discrimination; the second model further adjusted for predisposing, enabling, and need
results were calculated between the primary predictor variable and the secondary predictor variables to estimate the theorized pathways between mis- or disinformation on social media and unmet need for health care in multivariable logistic regression models. The interaction results were converted into predicted probabilities using the margins command in Stata (StataNow 18.5; StataCorp) for ease of interpretation. Sensitivity analyses were conducted that excluded non-Latino White individuals for the interaction effect of medical discrimination and perceived mis- or disinformation. All analyses accounted for survey weights and design using jackknife replicate weights for variance estimation. Statistical significance was set at $\alpha < .05$.

### Results

Table 1 provides the unadjusted sample size and survey-weighted percentage points for all study variables among adult social media users in the past year. Most participants (n=2655, 67%) reported that their health care needs were met; a third of the participants (n=1309, 33%) reported unmet need for health care. More than a third of participants (n=1397, 36%) perceived substantial social media health mis- and disinformation. Most participants (n=2803, 73%) used social media daily. Low trust in the health care system was reported by 17% (n=596) of the participants, and 8% (n=348) reported experiencing racial or ethnic discrimination when seeking health care.

Table 2 shows the survey-weighted bivariable column percentages, showing statistically significant relationships between the outcome and predictors. Among participants who perceived substantial social media mis- and disinformation, 42% reported unmet need compared to 33% who reported their need for care was met. Among participants who reported daily social media use, 77% reported unmet need compared to 71% who reported their need for care was met. Among participants who reported low trust in the health care system, 24% reported unmet need compared to 14% who reported their need for care was met. Among participants who experienced racial or ethnic discrimination in receiving medical care, 13% reported unmet need compared to 5% who reported their need for care was met.

In Table 3, the first model of the multivariable logistic regression shows the association of the predictor variables without adjustment for covariates. Perception of substantial social media health mis- and disinformation (OR 1.41, 95% CI 1.09 - 1.81), daily use of social media (OR 1.43, 95% CI 1.09 - 1.87), low health care system trust (OR 1.74, 95% CI 1.29 - 2.35), and perceived discrimination (OR 2.34, 95% CI 1.47 - 3.73) were all associated with a higher likelihood of unmet need for medical care. In model 2, after adjusting for predisposing, enabling, and need factors, perception of substantial social media health mis- and disinformation (OR 1.40, 95% CI 1.07 - 1.82), daily use of social media (OR 1.34, 95% CI 1.01 - 1.79), low health care system trust (OR 1.46, 95% CI 1.06 - 2.01), and perceived discrimination (OR 2.24, 95% CI 1.44 - 3.50) remained significantly associated with a higher likelihood of unmet need for medical care.

Table 4 shows the results from multivariable logistic models, adjusted for predisposing, enabling, and need factors, in which perception of mis- and disinformation was multiplied by predictors hypothesized to be mechanisms of the relationship between unmet need for care and perception of mis- and disinformation. There was a statistically significant difference in the probability of reporting unmet need for care among adults who did not use social media daily and who did not perceive substantial mis- and disinformation (24%; 95% CI 19% - 30%) compared to daily social media users who perceived substantial mis- and disinformation (38%; 95% CI 32% - 43%). Adults who perceived substantial mis- and disinformation and had low trust in the health care system had the highest probability of reporting unmet need for care (43%; 95% CI 38% - 49%) compared to the other three groups. Adults who perceived substantial mis- and disinformation and experienced medical care discrimination had a statistically significant higher probability of reporting unmet need for care (51%; 95% CI 40% - 62%) compared to adults who did not experience medical care discrimination and did not perceive substantial mis- and disinformation (29%; 95% CI 26% - 32%). We conducted a sensitivity analysis for the interaction of medical care discrimination and perception of mis- and disinformation, excluding non-Latino White participants, which resulted in a sample size of 1692. The results were replicated. Racially and ethnically minoritized adult social media users who perceived substantial mis- and disinformation and experienced medical care discrimination had a statistically significant higher probability of reporting unmet need for medical care (52%; 95% CI 38% - 67%) compared to racially and ethnically minoritized adult social media users who did not experience medical care discrimination and did not perceive substantial mis- and disinformation (24%; 95% CI 20% - 28%).
Table. Unadjusted sample size and survey-weighted percentages for study variables from the 2022 Health Information National Trends Survey 6 (N=3964).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values, n (weighted %)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong></td>
<td></td>
</tr>
<tr>
<td>Unmet need for medical care (no)</td>
<td>2655 (67)</td>
</tr>
<tr>
<td>Unmet need for medical care (yes)</td>
<td>1309 (33)</td>
</tr>
<tr>
<td><strong>Predictors</strong></td>
<td></td>
</tr>
<tr>
<td>Perception of social media health mis- or disinformation</td>
<td></td>
</tr>
<tr>
<td>&lt;Substantial</td>
<td>2567 (64)</td>
</tr>
<tr>
<td>Substantial</td>
<td>1397 (36)</td>
</tr>
<tr>
<td>Frequency of social media use</td>
<td></td>
</tr>
<tr>
<td>&lt;Daily</td>
<td>1161 (27)</td>
</tr>
<tr>
<td>Daily</td>
<td>2803 (73)</td>
</tr>
<tr>
<td>Health care system trust</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>3368 (83)</td>
</tr>
<tr>
<td>Low</td>
<td>596 (17)</td>
</tr>
<tr>
<td>Medical care discrimination</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3616 (92)</td>
</tr>
<tr>
<td>Yes</td>
<td>348 (8)</td>
</tr>
<tr>
<td><strong>Predisposing factors</strong></td>
<td></td>
</tr>
<tr>
<td>Age group (years)</td>
<td></td>
</tr>
<tr>
<td>18 - 49</td>
<td>1751 (59)</td>
</tr>
<tr>
<td>50 - 64</td>
<td>1185 (27)</td>
</tr>
<tr>
<td>≥65</td>
<td>1028 (14)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1475 (47)</td>
</tr>
<tr>
<td>Female</td>
<td>2489 (53)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td>2208 (57)</td>
</tr>
<tr>
<td>Formerly married</td>
<td>928 (10)</td>
</tr>
<tr>
<td>Never married</td>
<td>828 (33)</td>
</tr>
<tr>
<td>Rural and urban designation</td>
<td></td>
</tr>
<tr>
<td>Nonmetro</td>
<td>489 (12)</td>
</tr>
<tr>
<td>Metro</td>
<td>3475 (88)</td>
</tr>
<tr>
<td>Race and ethnicity</td>
<td></td>
</tr>
<tr>
<td>Non-Latino White</td>
<td>2272 (61)</td>
</tr>
<tr>
<td>Non-Latino Black</td>
<td>613 (11)</td>
</tr>
<tr>
<td>Latino</td>
<td>725 (18)</td>
</tr>
<tr>
<td>Non-Latino other</td>
<td>354 (10)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Not a college graduate</td>
<td>1915 (64)</td>
</tr>
<tr>
<td>College graduate or higher</td>
<td>2049 (36)</td>
</tr>
<tr>
<td><strong>Enabling factors</strong></td>
<td></td>
</tr>
<tr>
<td>Full-time employment</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Values, n (weighted %)</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Feeling about household income</td>
<td></td>
</tr>
<tr>
<td>Finding it very difficult/difficult to get by on present income</td>
<td>773 (19)</td>
</tr>
<tr>
<td>Getting by on present income</td>
<td>1453 (37)</td>
</tr>
<tr>
<td>Living comfortably on present income</td>
<td>1738 (44)</td>
</tr>
<tr>
<td>Covered by any health insurance</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>337 (11)</td>
</tr>
<tr>
<td>Yes</td>
<td>3627 (89)</td>
</tr>
<tr>
<td>Number of health care provider annual visits</td>
<td></td>
</tr>
<tr>
<td>0 - 3</td>
<td>237 (65)</td>
</tr>
<tr>
<td>≥4</td>
<td>159 (35)</td>
</tr>
<tr>
<td>Need factors</td>
<td></td>
</tr>
<tr>
<td>Fair or poor general health</td>
<td>634 (15)</td>
</tr>
<tr>
<td>Excellent, very good, or good general health</td>
<td>3330 (85)</td>
</tr>
</tbody>
</table>

Table. Survey-weighted bivariable column percentages for unmet medical care needs, perception of health mis- and disinformation on social media, frequency of social media use, trust in the health care system, and experience of racial or ethnic discrimination when receiving medical care among adult social media users in the past year from the 2022 Health Information National Trends Survey 6 (N=3964).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Met need for medical care, weighted %</th>
<th>Unmet need for medical care, weighted %</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptions of social media health mis- and disinformation</td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>&lt;Substantial</td>
<td>67</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Substantial</td>
<td>33</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Frequency of social media use</td>
<td></td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td>&lt;Daily</td>
<td>29</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>71</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Health care system trust</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>High</td>
<td>86</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>14</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Medical care discrimination</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>No</td>
<td>95</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>P values were calculated using the adjusted Wald $\chi^2$ test.
Table. Multivariable odds ratios (ORs) and 95% CIs for unmet need for medical care among social media users in the past year derived from the 2022 Health Information National Trends Survey 6 (N=3964). Logistic regression models were adjusted for survey weight and design. Model 1 shows the association of the predictor variables without adjustment for covariates. Model 2 added adjustments for age, sex, marital status, urban or rural designation, race and ethnicity, education, employment status, feelings about household income, health insurance coverage, number of health care provider visits, and general health status.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1, OR (95% CI)</th>
<th>Model 2, OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception of social media health mis- and disinformation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;Substantial</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Substantial</td>
<td>1.41 (1.09 - 1.81)</td>
<td>1.40 (1.07 - 1.82)</td>
</tr>
<tr>
<td>Frequency of social media use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;Daily</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Daily</td>
<td>1.43 (1.09 - 1.87)</td>
<td>1.34 (1.01 - 1.79)</td>
</tr>
<tr>
<td>Health care system trust</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Low</td>
<td>1.74 (1.29 - 2.35)</td>
<td>1.46 (1.06 - 2.01)</td>
</tr>
<tr>
<td>Medical care discrimination</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Yes</td>
<td>2.34 (1.47 - 3.73)</td>
<td>2.24 (1.44 - 3.50)</td>
</tr>
</tbody>
</table>

Table. Multivariable-adjusted percentage points for unmet need for medical care and the interaction effect between perceptions of health mis- and disinformation on social media and frequency of social media use, health care system trust, and medical care discrimination from the 2022 Health Information National Trends Survey 6 (N=3964). Predicted marginal effects were calculated from multivariable logistic regression models that were adjusted for survey weight and design, age, gender, marital status, urban or rural designation, race and ethnicity, education, employment status, feelings about household income, health insurance coverage, number of health care provider visits, and general health status.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Perceptions of health mis- and disinformation on social media, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;Substantial</td>
</tr>
<tr>
<td>Frequency of social media use</td>
<td></td>
</tr>
<tr>
<td>&lt;Daily</td>
<td>24 (19 - 30)</td>
</tr>
<tr>
<td>Daily</td>
<td>33 (28 - 37)</td>
</tr>
<tr>
<td>Health care system trust</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>30 (24 - 36)</td>
</tr>
<tr>
<td>Low</td>
<td>31 (27 - 34)</td>
</tr>
<tr>
<td>Medical care discrimination</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>29 (26 - 32)</td>
</tr>
<tr>
<td>Yes</td>
<td>49 (37 - 61)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

The primary objective of this study was to apply an extension of the Anderson model of health care utilization by evaluating the communication environment as a contextual determinant of health care seeking and utilization, which had not been specifically explored in prior studies [11-13]. Toward this end, we examined the association between perceptions of health mis- and disinformation on social media and unmet need for health care. We found that perception of substantial social media mis- and disinformation was associated with a higher probability of reporting unmet health care need compared to those who did not perceive substantial mis- and disinformation. This result is consistent with a growing body of literature demonstrating the negative impacts of social media health mis- and disinformation on health outcomes [16-19]. Our findings suggest that the communication environment serves as a contextual determinant of seeking health care by exposing individuals to false health information that is then associated with people delaying or avoiding medical care, even when it is needed, resulting in worse health outcomes [14]. Thus, perceptions of substantial health mis- and disinformation on social media may serve as one possible underlying mechanism for unmet need for health care.

In addition, we evaluated whether the relationship between perceptions of health mis- and disinformation on social media and unmet need for health care depended on the frequency of
social media use, trust in the health care system, and experience of racial and ethnic discrimination in health care. We found that unmet need for medical care was lower among adults who did not use social media daily and did not perceive substantial mis- and disinformation compared to daily social media users who perceived substantial mis- and disinformation. This finding is consistent with our theoretical framework that integrated the Anderson model of health care utilization with the frameworks that view the communication environment as a social driver of health [11,12]. Our empirical findings and theoretical framework suggest that more time spent on social media increases the exposure to and engagement with mis- and disinformation, which can lead frequent daily social media users to be more likely to have false beliefs that lower their likelihood of seeking needed medical care [32-34].

For the second mechanism, adults with low trust in the health care system and those who perceive substantial mis- and disinformation had the highest probability of reporting unmet health needs. This finding is consistent with prior literature and our theoretical framework demonstrating a trend toward eroding trust in medical professionals and health institutions associated with social media health mis- and disinformation, which influences harmful health behaviors and poor health outcomes [36-42]. Trust in the health care system is important for patient compliance with recommended care and seeking necessary medical care, and trust is difficult to rebuild once lost, suggesting this mechanism is critical for reducing negative impacts of unmet need for medical care [4,35].

Finally, adults who perceived substantial mis- and disinformation and experienced medical care discrimination had a higher probability of reporting unmet need for medical care compared to adults who did not experience medical care discrimination and did not perceive substantial mis- and disinformation. This finding is consistent with a large body of literature and our theoretical framework demonstrating the negative impacts of racial and ethnic discrimination experiences for seeking needed health care [43-50]. Our results and recent studies suggest that experiencing discrimination may have a complex relationship with social media health mis- and disinformation by further reducing historical levels of low trust in health care institutions [43,52,53].

Limitations

There are several limitations to consider when interpreting the results. First, a nonresponse analysis of a prior iteration of HINTS from 2011 and 2013 suggested that estimates using measures on seeking health information may be higher than other surveys [54]. To mitigate this potential effect and increase the validity of the study results, we excluded individuals who did not use social media and those who had not visited a social media platform in the past year (during the study period). Of course, the findings are the result of a cross-sectional survey, and therefore, should not be interpreted to indicate a causal relationship. One final context for interpreting the findings is that this study was focused on social media health mis- and disinformation and the findings might not apply to mis- and disinformation in other health contexts.

Conclusions

Based on our expanded version of the Anderson health care utilization model, we found that unmet need for health care was higher among individuals who perceived a substantial amount of social media mis- and disinformation, especially among those who used social media daily, had a low trust in the health care system, and experienced racial or ethnic discrimination when receiving health care. Our findings justify an expansion of the Anderson model of health care utilization by conceptualizing the communication environment as a contextual determinant of seeking health care. Future research should continue to model the communication environment as a contextual determinant of health care utilization and further evaluate the mechanisms of exposure to health mis- and disinformation, such as trust and discrimination. By extension, another area of future research could be the role of digital literacy, which we did not measure in this study. Individuals possessing advanced digital literacy skills are more likely to have the ability to differentiate between reliable health information and misleading or false information, and therefore, digital literacy may influence the connection between social media use and unmet health care needs [55]. Another area of future research is the role that social media plays in the support networks for individuals with specific health issues, perhaps mitigating or exacerbating the adverse impacts of mis- and disinformation through peer-supplied information and emotional support, which could interact with trust and past experiences of discrimination [55].

To counter the negative effects of social media mis- and disinformation on unmet need for health care, tailored public health efforts on social media are needed. These efforts should specifically target prevailing misunderstandings and offer fact-based information that has the potential to diminish the spread and consequences of health mis- and disinformation. As a component of this effort, health care institutions could actively monitor social media patterns to swiftly detect and address the emergence of health-related false and misleading information, potentially in real-time [55,56]. By implementing patient education programs that focus on developing health literacy, especially in the assessment of web-based health information, patients could be empowered to make well-informed health care choices [57]. Finally, health care institutions should continue to work toward building trust and reducing structural racism to mitigate the effects of mis- and disinformation and encourage patients to access health services when needed.

Acknowledgments

This research was supported by the National Institute on Minority Health and Health Disparities (NIMHD) at the National Institutes of Health (NIH) under award number R01MD018727 to ANO. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIMHD. The funders had no role in the study design, data analysis, decision to publish, or preparation of the manuscript.
Data Availability
The data sets generated during and/or analyzed during this study are available in the National Cancer Institute repository [58].

Authors' Contributions
All authors contributed to the study conception and design. Data analysis was performed by JPS. The first draft of the manuscript was written by JPS. All authors contributed to subsequent drafts of the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest
None declared.

References


58. HINTS. URL: https://hints.cancer.gov/ [accessed 2024-07-04]

Abbreviations

HINTS: Health Information National Trends Survey
OR: odds ratio

© Jim P Stimpson, Sungchul Park, Fernando A Wilson, Alexander N Ortega. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 11.7.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.
Evidence From the China Family Panel Studies Survey on the Effect of Integrating the Basic Medical Insurance System for Urban and Rural Residents on the Health Equity of Residents: Difference-in-Differences Analysis

Yingying Meng, Prof Dr; Ran Yu, MA; Huixin Bai, PhD; Junqiang Han, Prof Dr

1School of Political Science and Public Administration, Center for Social Security Studies, Wuhan University, Wuhan, China
2School of Public Management, South-Central Minzu University, Wuhan, China
*all authors contributed equally

Corresponding Author:
Junqiang Han, Prof Dr
School of Public Management
South-Central Minzu University
182 Minzu Avenue
Hongshan District
Wuhan, 430074
China
Phone: 86 35 5409 4333
Email: Junqianghan@scuec.edu.cn

Abstract

Background: The fragmentation of the medical insurance system is a major challenge to achieving health equity. In response to this problem, the Chinese government is pushing to establish the unified Urban and Rural Resident Basic Medical Insurance (URRBMI) system by integrating the New Rural Cooperative Medical Scheme and the Urban Resident Basic Medical Insurance. By the end of 2020, URRBMI had been implemented almost entirely across China. Has URRBMI integration promoted health equity for urban and rural residents?

Objective: This study aims to examine the effect of URRBMI integration on the health level of residents and whether the integration can contribute to reducing health disparities and promoting health equity.

Methods: We used the staggered difference-in-differences method based on the China Family Panel Studies survey from 2014 to 2018. Our study had a nationally representative sample of 27,408 individuals from 98 cities. We chose self-rated health as the measurement of health status. In order to more accurately discern whether the sample was covered by URRBMI, we obtained the exact integration time of URRBMI according to the official documents issued by local governments. Finally, we grouped the sample by urban and rural areas, regions, and household income to examine the impact of the integration on health equity.

Results: We found that overall, the URRBMI integration has improved the health level of Chinese residents (B=0.066, 95% CI 0.014-0.123; P=.01). In terms of health equity, the results showed that first, the integration has improved the health level of rural residents (B=0.070, 95% CI 0.012-0.128; P=.02), residents in western China (B=0.159, 95% CI 0.064-0.255; P<.001), and lower-middle-income groups (B=0.113, 95% CI 0.004-0.222, P=.04), so the integration has played a certain role in narrowing the health gap between urban and rural areas, different regions, and different income levels. Through further mechanism analysis, we found that the URRBMI integration reduced health inequity in China by facilitating access to higher-rated hospitals and increasing reimbursement rates for medical expenses. However, the integration did not improve the health of the central region and low-income groups, and the lack of access to health care for low-income groups was not effectively reduced.

Conclusions: The role of URRBMI integration in promoting health equity among urban and rural residents was significant (P=.02), but in different regions and income groups, it was limited. Focusing on the rational allocation of medical resources between regions and increasing the policy tilt toward low-income groups could help improve the equity of health insurance integration.

(JMIR Public Health Surveill 2024;10:e50622) doi:10.2196/50622
Key words
medical insurance system integration; Urban and Rural Resident Basic Medical Insurance; URRBMI; urban and rural residents; health equity; China; difference-in-differences; DID; staggered DID

Introduction

Health Inequity

With the rapid development of the global economy and medical technology, human health is improving, but health inequity is growing rather than shrinking. Health inequity in China is prominently observed between regions, urban and rural areas, and different income groups. Regional health inequity is mainly characterized by marked differences between the eastern coast and the western regions. In terms of life expectancy, the eastern provinces (or municipalities) are in the lead (eg, Beijing at 82.49 years, Shanghai at 82.55 years, Zhejiang at 80.26 years, Jiangsu at 79.32 years, and Guangdong at 79.31 years), while the western provinces exhibit lower figures, especially Qinghai (73.96 years), Yunnan (74.02 years), and the Tibet Autonomous Region (72.19 years) [1]. Health disparities also exist between urban and rural areas. From the end of 2019 to the end of 2020, the mortality rates for infants aged <1 year and children aged from 0 to 4 years in rural China were 2.40‰ and 0.78‰, respectively; the corresponding figures for urban areas were 0.89‰ and 0.25‰ [2]. Research based on China’s national microdata has proved that inequity favoring the rich for health-related quality of life remains significant in China, even after controlling for demographic factors [3].

The factors influencing health inequalities have been well discussed in previous studies. Income levels greatly influence health [4,5]. Higher-income groups not only have better access to health care but also enjoy greater health coverage [6]. In addition, there are disparities in health and the accessibility of health services for rural and susceptible populations, [7-9], with economic, medical, and educational resources being the main causes of health inequalities [10].

Inequity in the Health Insurance System

However, relatively little attention has been paid in existing studies to health inequalities resulting from inequalities in health insurance systems. One study found that of the 5 key factors contributing to health inequities in the European region, income security and social security have the highest proportion of the burden of health inequities at 35%, including income security and social security, living conditions, social and human capital, access to health care systems and their quality, and employment and working conditions [11]. Many countries have achieved universal health insurance coverage by establishing health insurance for different groups of people [12], which can lead to fragmentation of health insurance systems, leading to inequalities in health service use and economic security [13], and this represents a major challenge to achieving health equity [14,15].

China has built the largest medical insurance network in the world, covering >1.3 billion residents but has yet to unify the system. Urban Employee Basic Medical Insurance (UEBMI) is designed to cover employed urban residents, whereas Urban Residents Basic Medical Insurance (URBMI) extends coverage to the unemployed and retired individuals, older adults, students, and children in urban areas. Meanwhile, the New Rural Cooperative Medical Scheme (NRCMS) is implemented to cater to the health care needs of rural residents. However, the 3 medical insurance systems vary in terms of coverage, fund-raising standards, security benefits, medical insurance catalogs, designated institutions, management systems, and the level of overall planning. This fragmented medical insurance system has affected the fairness of treatment access for urban and rural residents. Compared to the URBMI, the NRCMS has higher out-of-pocket medical costs but fewer drugs, services, and facilities that can be reimbursed. Urban and rural residents do not enjoy the same coverage when faced with the same medical needs. Thus, the 3 health insurance systems in China have different degrees of health promotion effects on participants. UEBMI has the greatest effect on health improvement due to the high level of security benefits. NRCMS has the least [16-18] or even insignificant effect [19]. More details about NRCMS and URBMI are provided in Table 1.
Table 1. Introduction to China’s basic medical insurance system (New Rural Cooperative Medical Scheme [NRCMS], Urban Resident Basic Medical Insurance [URBMI], and Urban and Rural Resident Basic Medical Insurance [URRBMI]).

<table>
<thead>
<tr>
<th></th>
<th>NRCMS</th>
<th>URBMI</th>
<th>URRBMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>Rural residents</td>
<td>Urban residents are not covered by the UEBMI, including the unemployed, retired, older adults, students, and children</td>
<td>Urban and rural residents are not covered by the UEBMI</td>
</tr>
<tr>
<td>Fund-raising standards</td>
<td>CNY 657 (US $90.7) per person per year in 2018</td>
<td>CNY 776 (US $107.1) per person per year in 2018</td>
<td>CNY 1020 (US $140.8) per person per year in 2023</td>
</tr>
<tr>
<td>Benefits</td>
<td>Lower</td>
<td>Higher</td>
<td>According to high standards</td>
</tr>
<tr>
<td>Management system</td>
<td>National Health and Family Planning Commis-</td>
<td>Ministry of Human Resources and Social</td>
<td>National Healthcare Security Administra-</td>
</tr>
<tr>
<td></td>
<td>sion of the People’s Republic of China</td>
<td>Security of the People’s Republic of China</td>
<td>tion, which integrates multiple functions</td>
</tr>
<tr>
<td>Level of overall planning</td>
<td>County-level integration</td>
<td>Prefecture-level integration</td>
<td>At the local and municipal levels, in princi-</td>
</tr>
</tbody>
</table>

aUEBMI: Urban Employee Basic Medical Insurance.

Policy Background

The Chinese government is attempting to break down the fragmentation of the health insurance system and promote health equity through integrating NRCMS with URBMI. The project began as a pilot exercise in 2009 and was officially implemented in 2016. In 2016, China’s State Council issued a document named Opinions of the State Council on Integrating the Basic Medical Insurance Systems for Urban and Rural Residents [20], which proposed to integrate URBMI with NRCMS to establish a unified basic medical insurance system for urban and rural residents (URRBMI). The opinions clearly set out 6 integration requirements, unifying the management of 6 aspects, including coverage of systems, fund-raising policies, social security benefits, medical insurance catalogs, the management of designated institutions, and fund management. In addition, the integration as a whole follows 3 principles: regardless of whether residents belong to urban or rural areas, URRBMI will be based on a lower standard for determining contributions, a higher standard for determining treatment, and a wider standard for determining the medical insurance catalog. Overall, 6 integration requirements have enabled urban and rural residents to enjoy a fair and unified medical insurance system and reduced the cost of running medical insurance. A total of 3 principles have raised the level of treatment and narrowed the treatment gap between urban and rural residents. Through the above mentioned policy design, the integrated URRBMI has improved the fairness of urban and rural medical insurance treatment. We compared the URRBMI with NRCMS and URBMI in Table 1.

Several empirical studies have explored the impacts of URRBMI integration and found different evidence. URRBMI integration is beneficial in increasing the number of health service utilizations among rural residents [21] and in mitigating rich inequalities in outpatient benefit probabilities and benefit levels [22]. However, some research finds that the integration has a negative impact on reducing inequalities in inpatient health services [23]. Although the integration significantly increases the use of inpatient care, it has a limited impact on health outcomes for rural middle-aged and older adult residents [24].

Overall, previous studies have focused on the impact on health service use and health care costs. While a small body of literature has explored the health impacts of URRBMI integration [23,25], there is a lack of focus on health equity. The relationship between URRBMI and health equity remains uncertain. Limited health care resources in rural areas and poor areas may reduce access to health care services, while the abundance of economic and health care resources in urban areas and wealthy households may increase health care use [26], so better-off areas and households benefit more from the integration. However, the price elasticity of demand for health care services is higher in the low-income group, and the degree of response to the price of health care use changed by health care insurance will be greater, so health care insurance can promote the improvement of health care use and thus health in this group [27]. The aim of this study was to explore the differential impact of health effects across urban and rural areas, different regions, and groups of different income levels and assess whether URRBMI integration is effective in improving health equity for Chinese residents.

Methods

Data

This study used open data from the China Family Panel Studies (CFPS), designed by the research team at Peking University. These data are the first large-scale, comprehensive, and academically oriented social tracking survey in China. The data were made publicly available in 2010 as a baseline survey result, followed by a biennial full-sample tracking survey covering all household members in the 25 provincial-level administrative regions of mainland China [28], which represents 95% of the Chinese population. The data focus on the economic and noneconomic well-being of China’s population, including economic activity, educational attainment, family relationships...
and family dynamics, population migration, and physical and mental health, among many other research topics.

This study used the CFPS data from 2014, 2016, and 2018 database, which was because most cities (50/98, 51%) were consolidated in the period from 2015 to 2017. To obtain a sample suitable for the study, the raw data were further processed. At the individual level, only those individuals whose participation types were NRCMS, URBMI, or URRBMI were retained; samples with missing or abnormal information were deleted. At the city level, we deleted cities that had already achieved URRBMI integration before and in 2014, such as Changsha, Shaoquguan, Yangjiang, Qianxinan Buyei and Miao Autonomous Prefecture, Dongguan, Maoming, Meizhou, Shantou, Jiangmen, Yunfu, Zibo, Qingyuan, Jieyang, Zhanjiang, Hangzhou, Chongqing, Chengdu, Wuxi, Zuhuai, and Tianjin. We also deleted cities with unclear implementation of the integration policy (judged by whether the local government has issued an official document on integrating URBMI and NRCMS). Finally, the unbalanced panel was treated into a balanced panel. Through the aforementioned processing, this study finally retained 27,408 individual samples and 98 city samples. A total of 55 (56%) city samples were in the integration group, and 43 (44%) were in the nonintegration group. The integration time of the city samples is presented in Textbox 1.

Textbox 1. Years the Urban and Rural Resident Basic Medical Insurance (URRBMI) was integrated in each city.

<table>
<thead>
<tr>
<th>Year</th>
<th>Cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Ningde, Rizhao, Zaozhuang, Yantai, Ganzi Tibetan Autonomous Prefecture, Guangzhou, Laiwu, and Meishan</td>
</tr>
<tr>
<td>2016</td>
<td>Xinzhou, Ningbo, Shanghai, Ganzhou, and Yibin</td>
</tr>
<tr>
<td>2017</td>
<td>Yuncheng, Anyang, Linfen, Yongzhou, Shijiazhuang, Guangyuan, Honghe Hani and Yi Autonomous Prefecture, Pingdingshan, Jiaozuo, Fuzhou, Xinyang, Xuchang, Taiyuan, Qinhuangdao, Guilin, Zhengzhou, Hengshui, Cangzhou, Zhanjiakou, Luoyang, Lianyungang, and Yibin</td>
</tr>
<tr>
<td>2018</td>
<td>Baiyin, Harbin, Hegang, Liangyungang, Xiangfan, Lanzhou, Wuwei, Yichang, Wangzhou, Linxia Hui Autonomous Prefecture, Qingyang, Liangshan Yi Autonomous Prefecture, Luliang, Daqing, Jixi, and Daxinganling District</td>
</tr>
<tr>
<td>2019</td>
<td>Tianshui, Jinzhoub, Anqing, Xucheng, and Pingliang</td>
</tr>
<tr>
<td>2020</td>
<td>Chaoyang, Xi'an, Qiannanbu, Siping, Chishui, Pu'er, Songyuan, Qiandongnan Miao and Dong Autonomous Prefecture, Tieling, Dalian, Zunyi, Anshan, Shenyang, Dandong, Yulin, Liaooyang, Yingkou, Weinan, Fuxin, Hefei, and Tonghua</td>
</tr>
</tbody>
</table>

Ethical Considerations

As a human-involved research project, CFPS regularly submits ethical reviews to the Biomedical Ethics Review Committee of Peking University and carries out data collection work upon receiving review approval (approval number IRB00001052-14010). All written informed consent forms are provided by participants aged ≥15 years or by their parents (for participants aged ≤15 years). The personal information and privacy of the interviewees are strictly protected by CFPS according to the ethics rules.

Variables

Measurement of the Dependent Variable: Health Status of Residents

“Self-rated health” is the subjective evaluation of the respondents’ health status. Among many health assessment indicators, “self-rated health” is widely used in health measurement and evaluation. On the one hand, “self-rated health” takes into account the subjective knowledge of the respondents, which is scientific. On the other hand, it comprehensively evaluates the physical, psychological, and social health of individuals, so it can be more comprehensive [29]. Therefore, this study used the self-rated health score to measure the health status of residents as the dependent variable. This variable was based on the question on the CFPS questionnaire: “How would you rate your health status?” The answers were given 5 grades from “poor” to “excellent,” assigned 1 to 5 points, respectively.

Measurement of the Independent Variable: Whether the Sample Cities Have Integrated NRCMS With URBMI Into URRBMI

The existing research had mainly used the type of insurance participation in the questionnaire to identify the treatment group and the control group. The method of individual definition directly using the self-reported data has serious inaccuracies because ordinary rural residents are usually only concerned about whether the health insurance benefits are improved and do not care about the change of the type of insurance participation from “NRCMS” or “URBMI” to “URRBMI” because the change does not affect their medical care and reimbursement processes. Therefore, this study obtained the exact integration time of the URRBMI according to the official documents issued by local governments, such as the opinions and plans on the implementation of URRBMI integration, which...
was the method of “city definition.” If the integration time was earlier than the data access year \( t \), the policy variable was assigned a value of 1; otherwise, it was assigned a value of 0. This setup produced a “treatment group” and a “control group,” as well as “pretreatment” and “posttreatment” double differences.

**Controlled Variables**

In order to alleviate the problem of missing variables, this study controlled other important variables affecting the health of residents as far as possible, according to previous relevant studies and survey data [30-32]. Controlled variables included not only the individual characteristics variables such as gender, age, marital status, type of household registration, education level, family size, annual household income per capita, and region but also the health characteristics variables such as smoking status, frequency of alcohol consumption, frequency of exercise, and chronic disease status. The definitions of the main variables and description of the mean are presented in Table 2.
Table 2. Definition of key variables and description of statistical analysis (N=27,408).

<table>
<thead>
<tr>
<th>Variables and definition</th>
<th>Value</th>
<th>Values, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Health levels</strong></td>
<td></td>
<td>2.8863 (1.2670)</td>
</tr>
<tr>
<td>Poor</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td><strong>Policy variables (dummy variable)</strong></td>
<td></td>
<td>0.2156 (0.4113)</td>
</tr>
<tr>
<td>The integration time is earlier than the year ( t ) of the data access</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Otherwise</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Gender (dummy variables)</strong></td>
<td></td>
<td>0.4646 (0.4988)</td>
</tr>
<tr>
<td>Man</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Woman</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Age in years (continuous variable)</strong></td>
<td></td>
<td>49.9984 (14.1595)</td>
</tr>
<tr>
<td>Age of the participant</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Marital status (dummy variables)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td></td>
<td>0.0462 (0.2098)</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Married or cohabiting</td>
<td></td>
<td>0.8863 (0.3174)</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Divorced or widowed</td>
<td></td>
<td>0.0675 (0.2509)</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Household registration (dummy variables)</strong></td>
<td></td>
<td>0.0968 (0.2957)</td>
</tr>
<tr>
<td>Nonagricultural household</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Agricultural household</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Education level (dummy variables)</strong></td>
<td></td>
<td>0.3426 (0.4746)</td>
</tr>
<tr>
<td>Unschooled</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Primary school</strong></td>
<td></td>
<td>0.2562 (0.4365)</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Middle school</strong></td>
<td></td>
<td>0.2871 (0.4524)</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>High school and above</strong></td>
<td></td>
<td>0.1142 (0.3180)</td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Family size (members; dummy variables)</strong></td>
<td></td>
<td>0.3398 (0.4737)</td>
</tr>
<tr>
<td>Small</td>
<td>1-3</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
### Staggered Difference-in-Differences Model

On the basis of the fact that the timing of URRBMI integration varies in different regions, this study used the staggered difference-in-differences (DID) model to examine the impact of the integration on residents’ health level and equity. The basic idea of the DID model is to construct the treatment group and the control group and to identify the health level of residents before and after policy implementation. The staggered DID model is applicable to a special case of policy implementation, that is, policy implementation from the beginning of the pilot to the gradual extension. There are differences in implementation time in different regions [33]. Our specific models were as follows:

<table>
<thead>
<tr>
<th>Variables and definition</th>
<th>Value</th>
<th>Values, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>4-5</td>
<td>0.3773 (0.4847)</td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Large</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>≥6</td>
<td>0.2829 (0.4504)</td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Income (in ten thousands of yuan; continuous variable)</td>
<td>Annual household income per capita</td>
<td>1.3693 (1.4146)</td>
</tr>
<tr>
<td><strong>Region (dummy variables)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>East</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>0.2851 (0.4515)</td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Central</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>0.3590 (0.4797)</td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>West</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>0.3559 (0.4788)</td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Smoking (dummy variables)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>0.2904 (0.4540)</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Frequency of alcohol consumption (dummy variables)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have drunk alcohol 3 times a week in the past</td>
<td>1</td>
<td>0.1587 (0.3654)</td>
</tr>
<tr>
<td>Otherwise</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Frequency of exercise (per week, dummy variables)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 times</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>0.6349 (0.4815)</td>
</tr>
<tr>
<td>Otherwise</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2-4 times</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>0.1210 (0.3262)</td>
</tr>
<tr>
<td>Otherwise</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>&gt;4 times</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>0.2441 (0.4296)</td>
</tr>
<tr>
<td>Otherwise</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Chronic disease (dummy variables)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>0.1969 (0.3976)</td>
</tr>
<tr>
<td>Otherwise</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
In light of the prominent health disparities observed across different regions, urban and rural areas, and various income groups in China, this study aimed to investigate the differential impact of the system integration on the health status of these subgroups. Specifically, through grouping estimations in subsequent sections, we sought to assess the extent to which policies contributed toward achieving health equity for Chinese residents.

The Impact of URRBMI Integration on the Health Equity of Chinese Residents

In light of the prominent health disparities observed across different regions, urban and rural areas, and various income groups in China, this study aimed to investigate the differential impact of the system integration on the health status of these subgroups. Specifically, through grouping estimations in subsequent sections, we sought to assess the extent to which

over time, $\lambda_t$ represents the time-fixed effect, and $\varepsilon_{ijt}$ is a random disturbance term. The estimated coefficient $\beta_j$ measures the effects of policy implementation.

### Results

#### Regression Results of the Impact of URRBMI Integration on the Health Level of Chinese Residents

Table 3 reports the regression results of the effect of the URRBMI integration on the health level of Chinese residents.

<table>
<thead>
<tr>
<th>Model</th>
<th>Policy variables (SE; 95% CI)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DID$^{ab}$ (N=27,408)</td>
<td>0.066 (0.031; 0.006-0.125)</td>
<td>2.15 (27,291)</td>
<td>.03</td>
</tr>
<tr>
<td>DID$^c$ (N=27,408)</td>
<td>0.068 (0.028; 0.014-0.123)</td>
<td>2.45 (27,291)</td>
<td>.01</td>
</tr>
<tr>
<td>PSM-DID$^{de}$ (n=14,410)</td>
<td>0.090 (0.038; 0.015-0.166)</td>
<td>2.35 (14,293)</td>
<td>.02</td>
</tr>
</tbody>
</table>

$^a$DID: difference-in-differences.

$^b$City-fixed effect and time-fixed effect were controlled.

$^c$City-fixed effect, time-fixed effect, and control variables were controlled.

$^d$PSM-DID: propensity score matching combined with the difference-in-differences.

$^e$City-fixed effect, time-fixed effect, and control variables were controlled.

We controlled for fixed effects and control variables in turn. The second row showed that the coefficient of the policy variable was 0.068, which was significant at the 5% statistical level, indicating that the URRBMI integration policy had significantly improved the health level of residents.

In addition, this study further used the propensity score matching combined with the DID method to address the endogenous issues that might arise from the URRBMI policy implementation “selection bias.” The role of the propensity score matching method is to look for the nonintegrated group whose characteristics are similar to those of the integrated group, that is, to select the nonintegrated group that is only in the range of the cosupport interval with the integrated group as the control group, to enhance comparability between integrated and nonintegrated groups. In this study, the nearest neighbor matching method was used to pair each sample in an integrated group with a sample in a nonintegrated group whose score was closest to each other, while limiting the propensity score gap to 0.01 (the caliper is not >0.01). Regression estimates were made using the matched samples, and the results are presented in the third row of Table 3. The regression coefficient was 0.090 after treatment, which was significant at the level of 5%. The regression result indicated that the conclusions above remained robust.

As mentioned in the Introduction section, due to the split medical insurance system between rural and urban areas, URRBMI integration will have a different impact on rural and urban residents. This study identified the impact by dividing the sample into 2 groups: urban residents (nonagricultural households) and rural residents (agricultural households), based on the residents’ household registration.

The results showed that the integration had a significant contribution to the health of rural residents (B=0.070, 95% CI 0.012-0.128; P=.02), while it did not have a significant effect on urban residents (B=0.063, 95% CI -0.103 to 0.229; P=.46). This indicated that the integration significantly reduced health inequalities between urban and rural areas and achieved the policy objective better.

#### Regression Results of the Impact of URRBMI Integration on Health Equity in Different Regions

As mentioned in the Introduction section, due to the regional differences in economic and health care resource allocation, URRBMI integration will have a different impact in different regions. This study identified the impact by dividing the Chinese mainland into 3 major economic regions, namely, the eastern, central, and western regions, according to the division criteria of the National Bureau of Statistics of China. The eastern region includes Shanghai, Beijing, Tianjin, Shandong, Guangdong, Jiangsu, Hebei, Zhejiang, Fujian, and Liaoning; the central region includes Jilin, Anhui, Shanxi, Jiangxi, Henan, Hubei,
Hunan, and Heilongjiang; and the western region includes Yunnan, Sichuan, Guangxi, Gansu, Guizhou, Chongqing, and Shaanxi. Table 4 reports the impact of URRBMI integration on regional health equity.

Table 4. Regression results of Urban and Rural Resident Basic Medical Insurance impact of the integration on health equity in different regionsa.

<table>
<thead>
<tr>
<th>Region</th>
<th>Policy variables (SE; 95% CI)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>East (n=7814)</td>
<td>0.016 (0.049; −0.081 to 0.112)</td>
<td>0.32 (7766)</td>
<td>.75</td>
</tr>
<tr>
<td>Central (n=9840)</td>
<td>0.090 (0.055; −0.018 to 0.199)</td>
<td>1.63 (9779)</td>
<td>.10</td>
</tr>
<tr>
<td>Western (n=9754)</td>
<td>0.159 (0.049; 0.064 to 0.255)</td>
<td>3.28 (9706)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aAll regressions included the full set of control variables and fixed effects.

The results showed that the integration had a significant contribution to the health level of residents in the western region (B=0.159, 95% CI 0.064-0.255; P<.001), while the effect on the eastern (P=.75) and central (P=.10) regions was not significant. This suggested that the integration had, to some extent, alleviated health inequalities between regions, but the effect was limited.

Regression Results of the Impact of URRBMI Integration on Health Equity for Different Income Groups

As mentioned in the Introduction section, because different income groups have different health needs and different levels of access to health services, URRBMI integration will have a different impact for different income groups. This study analyzed the impact by dividing the entire sample into 4 groups: low-income, lower-middle-income, upper-middle-income, and high-income groups, based on annual household income per capita. Table 5 reports the impact of the integration on health equity for different income groups.

Table 5. Regression results of the impact of Urban and Rural Resident Basic Medical Insurance integration on health equity for different income groupsa.

<table>
<thead>
<tr>
<th>Income groups</th>
<th>Policy variables (SE; 95% CI)</th>
<th>t test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (n=6860)</td>
<td>−0.036 (0.067; −0.168 to 0.095)</td>
<td>−0.54 (6747)</td>
<td>.59</td>
</tr>
<tr>
<td>Lower-middle (n=6444)</td>
<td>0.113 (0.056; 0.004 to 0.222)</td>
<td>2.03 (6330)</td>
<td>.04</td>
</tr>
<tr>
<td>Upper-middle (n=6852)</td>
<td>0.088 (0.058; −0.024 to 0.201)</td>
<td>1.54 (6739)</td>
<td>.12</td>
</tr>
<tr>
<td>High (n=6852)</td>
<td>0.050 (0.052; −0.051 to 0.152)</td>
<td>0.97 (6737)</td>
<td>.33</td>
</tr>
</tbody>
</table>

aAll regressions included the full set of control variables and fixed effects.

The results showed that URRBMI integration had a significant health promotion effect on the lower-middle-income groups (B=0.113, 95% CI 0.004-0.222; P=.04), but not on the other income groups (including low-income, upper-middle-income, and high-income groups). This indicated that URRBMI integration had, to some extent, alleviated the health inequity gap between different income groups, but the effect was limited.

Robustness Tests

Counterfactual Test

To justify the common trend hypothesis in this study, we conducted a counterfactual test to ensure that there is no systematic difference in the health level of residents in the integrated and nonintegrated groups before URRBMI integration. This study drew on existing research to test the “counterfactual” hypothesis [34]. This was done by applying the baseline model (1) to the period when the region had not yet integrated medical insurance (removing data from years after the integration of the integration group), that is, introducing a new variable “whether the region was an urban and rural resident’ health insurance integration group,” and we assigned a value of 1 if the region was integrated later than the time of data access and otherwise assigned a value of 0, thus constructing a hypothetical treatment group and control group about “urban-rural resident’ health insurance integration.” Since the hypothetical “urban-rural health insurance integration” was not expected to have a significant impact on the health of urban and rural residents during this period, the hypothetical control group could be considered the “counterfactual” result of the treatment group. The results show that integration did not have a significant (B=0.017, 95% CI −0.115 to 0.149; P=.80) impact on the health of urban and rural residents before the implementation of the integration, indicating that the “counterfactual” hypothesis was largely satisfied.

Placebo Test

In this quasi-natural experiment, the significant improvement in the self-rated health of urban and rural residents might be due to some random factors. In this study, we used some scholars’ approach to construct a placebo test to determine whether the health promotion effect of the integrated URRBMI was caused by other random factors [35]. The specific method was to use Stata software (StataCorp LLC) to make the time of the URRBMI on specific cities become random, and then repeat the random process 500 times to obtain a nuclear density distribution map of , as shown in Figure 1.
concentrated around 0, which proved that the unobserved influence factors had little influence on the estimation results. The dotted line indicated the coefficient size of the policy variable in column 3 of Table 3. Because the dotted line was positioned far away from the distribution of the 0, it meant that the real policy effect was different from the policy effect of the placebo test. It indicated that the coefficient of the policy variables in column 3 of Table 3 was statistically significant and stable.

Figure 1. Placebo test.

Substitution of Variables
In addition, we conducted robustness tests by replacing variables. The ability to perform activities of daily living was chosen as a measure of objective health. This indicator measured the ability of people aged >45 years to take care of themselves. It included 7 activities in the CFPS, such as going outside for outdoor activities, eating, performing kitchen activities, taking public transportation, going shopping, cleaning, and doing laundry. If an activity could be completed independently, it was assigned a value of 1; otherwise, it was assigned a value of 0. If there were n activities that could be completed independently, the cumulative total would be n. The larger the n was, the higher the objective health level. The results show that the integration policy still produced a significant (B=0.065, 95% CI 0.015-0.115; P=.01) positive contribution to residents' objective health, which indicated that the results were basically robust.

Discussion
Main Findings
This study used nationally representative data from China to demonstrate that URRBMI integration has a positive effect on improving health standards and promoting health equity. The results showed that URRBMI integration can improve the health level of rural residents, lower-income western regions, and lower-middle-income groups. However, the impact of reducing health disparities is still limited and has not improved the health of the central region and low-income groups. This section provides an in-depth discussion and analysis of the findings, adding additional evidence to further explain the plausibility of the findings. At the same time, this paper studied the impact path of medical insurance system integration on health, providing experience for global health insurance system reform and promoting health equity.

The Positive Impact of the Policy
The positive health effects of medical insurance are related to policy design. First, the integration has significantly increased the level of benefits, reducing the medical burden on residents accessing health care. The reimbursement rates and cap lines of URRBMI have generally increased in all provinces and municipalities, in line with the integration principles. For example, in Shanghai, before the integration, the minimum payment rate for URBMI was 60% (for those aged <60 years), while the minimum payment rate for the NRCSMS was only
50%. After the integration, the medical insurance payment rate for URRBMI was unified at 70% [36]. Second, patients’ choice of medical treatment has been increasing, which has improved the quality of medical treatment for the residents. On the one hand, the number of designated medical institutions in each province has increased after the integration. The original designated hospitals of NRCMS and URBMI have all been incorporated into the designated hospitals of URRBMI, which are located in various streets, towns, and communities (villages) in the city’s urban and rural areas, so the insured can reduce the inconvenience caused by residents seeking medical treatment in different places across counties (districts). On the other hand, the medical insurance catalog has been expanded by more than a thousand kinds of reimbursable drugs. After the integration, Inner Mongolia’s drug catalog of NRCMS has increased from 1988 to >2600, an increase of >30% [37]. Third, the process of reimbursement of medical expenses is more convenient, reducing the time and cost of medical treatment for residents. The reimbursement will be settled instantly through the social security card without the need for advance payment, eliminating all the hassles and inconveniences of medical treatment.

As mentioned in the previous paragraphs on impact pathways, rural residents have benefited more from the integration in terms of treatment levels and patients’ choice, so that URRBMI integration has promoted health equity between urban and rural residents. The increase in the drug catalog of the URRBMI is much less than that of the NRCMS, and so is the increase in the level of treatment. In addition, the quality of medical care for rural residents has improved due to the increased level of overall planning after the integration. Before the integration, URBMI was mostly coordinated at the prefectural and municipal levels, while the NRCMS was generally coordinated at the county level, so when rural residents went to a municipal hospital for medical treatment, they were seeking medical treatment in a different place, which had a lower reimbursement rate and complicated reimbursement procedures [38]. After the integration, URRBMI will in principle be coordinated at the municipal (prefectural) level and even maybe at the provincial level, breaking down the systemic and geographical differences between urban and rural residents’ access to health care and improving equity in the use of health services for rural residents [39].

The Negative Impact of the Policy

However, the URRBMI integration has played a limited role in promoting regional health equity, mainly in that it has not contributed to the improvement of health standards in the central region. This may be related to the level of regional economic development and public health care resources. The impact of health insurance and economic factors on health has a substitution effect; that is, the health performance of medical insurance diminishes as the level of regional economic development increases [40]. Those living in regions with higher levels of economic development are likely to have less impact on health care accessibility and, thus, lesser impact on health than on factors such as income, education, and environment [41]. For those living in less developed areas, their income levels and quality of life will also be relatively low, so health insurance can significantly improve access to health care, promoting health level in poorer areas [24]. As shown in Table 6, the economic development of the eastern regions is higher than that of the central and western regions, and therefore the health promotion effect of the integration is limited for the eastern regions.

The income level of the residents in the western region is lower than that of the residents in the eastern region, and the probability of “illness without treatment” is higher. Moreover, the population in the western region is relatively small, so the per capita medical resources are relatively sufficient, as shown in Table 6. In this case, the health promotion effect of the integration is more significant (P < .001) in the western region, while in the central region, it may be that the reform of the system has been neglected by the policy design due to its “middle zone” of economic development and medical resources.

We also found that URRBMI integration has played a limited role in health equity for different income groups and has not improved the health of low-income groups. A number of studies have shown that there is significant heterogeneity in the health performance of health insurance according to individual income, education, age, and household registration [27,42]. The heterogeneity is particularly evident in economic and social factors [43]. In general, low-income groups are more sensitive to price changes in health services than higher-income groups [44]. However, compared with the lower-middle-income groups, the increase in funding levels brought about by health insurance consolidation has resulted in a higher financial burden of health care for the low-income groups [45], which may still discourage health needs due to the risk of poverty, leading to a lack of access to health care. To verify this, the 2-week sickness rate and the rate of not seeking medical treatment were calculated using the questions in the CFPS questionnaire, “During the past two weeks, have you felt any physical discomfort? “ and “Have you seen a doctor?” and grouped according to income. The results shows that the 2-week sickness rate is higher in the low-income group (low-income: 0.3672; lower-middle-income: 0.3417; upper-middle-income: 0.3141;
high-income: 0.2877), and the phenomenon of not seeking medical treatment is more common (low-income: 0.8233; lower-middle income: 0.8090; upper-middle income: 0.8074; high-income: 0.7935). Therefore, the integration did not alleviate the equity of health care use among low-income groups, and the health promotion effect on low-income groups was not significant.

Mechanism Analysis

From the regression results in the Table 3, it can be seen that the integration of URRBMI has a promoting effect on the health of rural areas, western regions, and low-income and middle-income groups; therefore, in order to further verify the mechanism of the impact of health insurance integration on health equity, we chose “choice of medical treatment” and “reimbursement ratio” as possible mechanisms to be analyzed. In the CFPS, residents’ choices of medical care included 5 categories, such as general hospitals, specialty hospitals, community health centers or township health centers, community health service stations or village health offices, and clinics, which were assigned a value of 1 to 5, respectively. The reimbursement rate was expressed as the ratio of reimbursed medical expenses to total medical expenses, where reimbursed medical expenses were equal to the difference between total medical expenses and out-of-pocket medical expenses. Table 7 reports 2 mechanisms by which health care integration affects health.

Table 7. Mechanism analysis: regression results of the impact of Urban and Rural Resident Basic Medical Insurance integration on the medical choices and reimbursement ratio*

<table>
<thead>
<tr>
<th>Medical choices</th>
<th>β coefficient (SE)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural</td>
<td>-0.156 (0.067)</td>
<td>.02</td>
</tr>
<tr>
<td>Western</td>
<td>-0.118 (0.060)</td>
<td>.05</td>
</tr>
<tr>
<td>Lower-middle</td>
<td>-0.036 (0.017)</td>
<td>.04</td>
</tr>
<tr>
<td>Reimbursement ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>0.039 (0.014)</td>
<td>.004</td>
</tr>
<tr>
<td>Western</td>
<td>0.036 (0.017)</td>
<td>.04</td>
</tr>
<tr>
<td>Lower-middle</td>
<td>0.081 (0.046)</td>
<td>.08</td>
</tr>
</tbody>
</table>

*All regressions included the full set of control variables and fixed effects.

The first panel of Table 7 shows the regression results of the impact of the integration on the medical choices. It can be found that the coefficients of the policy variables were significant (rural: P=.02; western: P=.05; lower-middle: P=.04) for rural and western residents and lower-middle-income groups in the first column. This suggests that the URRBMI integration promoted the movement of residents to higher-ranking hospitals and improved the quality of care, thereby improving the health of rural and western residents and lower-middle-income groups. The second panel of Table 7 shows the regression results of the impact of the integration on the reimbursement ratio. It can be found that the reimbursement rate was expressed as the ratio of the reimbursed medical expenses to the total medical expenses, where the reimbursed medical expenses were equal to the difference between total medical expenses and out-of-pocket medical expenses. It can be found that the coefficients of the policy variables were significant for rural residents, western residents, and lower-middle-income groups (rural: P=.004; western: P=.04; lower-middle: P=.08). This suggested that the URRBMI integration increased the reimbursement rate of health insurance, which to a certain extent eased the medical burden of rural and western residents and lower-middle-income groups, thus improving the health of the population.

Conclusions

This study found that URRBMI integration has contributed to the improvement of residents’ health, mainly due to the improvement of medical treatment and the reduction of the cost of access to health care. In terms of the impact on health equity, this paper finds that integrated URRBMI can promote health equity between urban and rural areas. However, to a limited extent, it can alleviate health disparities between regions and between different income groups. Furthermore, we also explore 2 mechanisms affecting health equity, on the basis of these findings, the policy implications of this study are as follows: first, rationally allocate medical resources among different regions; do not neglect the development of regions with a moderate level of economic development, such as the middle region; improve the problem of insufficient resources in the middle region; and at the same time, increase support for regions with poor economic development, such as the western region. Second, the integration of medical insurance should increase

Limitations

There were still several limitations to this study. Due to data limitations, we used only self-rated health as an indicator to assess the health level of the insured. Although we chose objective health to replace the dependent variable, the indicator could only measure the health of people aged >45 years, resulting in a somewhat reduced sample size. However, due to the relatively large span of the research sample, at present, China’s official data have not found relevant data to measure the health level of the whole sample of indicators. In the future, we will further seek to use a more comprehensive evaluation indicator to further test the conclusions of the study.

The table shows the regression results of the impact of the integration on the medical choices and reimbursement ratio. The coefficients for rural and western residents, as well as lower-middle-income groups, are significant at the 0.05 level. The reimbursement rate for rural residents is 0.039 and for western residents is 0.036. The third column of the table shows the significance levels for these results.
the policy bias toward low-income groups to promote the use of health care services by low-income groups. This study provides experience and thoughts for China and other countries around the world in breaking down fragmented health care systems and establishing a unified health care system.

Acknowledgments
The authors thank the National Health and Health Council of China for cooperation and organizing field data collection. They acknowledge the financial support from Major Project of the Key Research Base for Humanities and Social Sciences of the Ministry of Education of China (project 22JJD630014) and the National Social Science Foundation of China (project 22BGL210). The authors would also like to thank all study participants for their time to be interviewed. YM would also like to thank Luo Jia Talents Plan for Outstanding Young Researchers.

Data Availability
The data sets were derived from China Family Panel Studies survey from 2014 to 2018. The data used in this paper can be downloaded from the official website of the China Family Panel Studies (http://isss.pku.edu.cn/cfps/index.htm).

Authors’ Contributions
YM is responsible for conceptualization, writing—original draft, project administration, and funding acquisition. RY is responsible for conceptualization, formal analysis, data curation, and writing—original draft. HB is responsible for writing—reviewing and editing and data curation. JH is responsible for methodology and supervision.

Conflicts of Interest
None declared.

References


20. Opinions of the state council on integrating the medical insurance system for urban and rural residents. State Council. 2016. URL: http://www.gov.cn/zhengce/content/2016-01/12/content_10582.htm [accessed 2022-12-12]


36. People get more from unified basic medical insurance system for urban and rural residents. Ministry of Human Resources and Social Security of the People's Republic of China. URL: http://www.mohrr.gov.cn/SYrlyshshbz/bdongtaixinwen/ buneiyaowo/201608/20160808_245052.html [accessed 2022-12-18]

38. Feng J, Cen XX, Hu B. The medical insurance based on efficiency and equity evidence from the integration of the national basic. Econ Res J 2021:154-172 [FREE Full text]


41. Social determinants of health. World Health Organization. URL: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1 [accessed 2022-12-27]


45. Ye H, Liu BT. Study on benefit equity of basic medical insurance for rural residents with different income levels: based on the survey of poverty-stricken ethnic minority counties in Hubei province. China National Knowledge Infrastructure. URL: https://shorturl.at/pU5ml [accessed 2022-12-20]

Abbreviations

CFPS: China Family Panel Studies
DID: difference-in-differences
NRCMS: New Rural Cooperative Medical Scheme
UEBMI: Urban Employee Basic Medical Insurance
URBMI: Urban Residents Basic Medical Insurance
URRBMI: Urban and Rural Resident Basic Medical Insurance

©Yingying Meng, Ran Yu, Huixin Bai, Junqiang Han. Originally published in JMIR Public Health and Surveillance (https://publichealth.jmir.org), 30.05.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on https://publichealth.jmir.org, as well as this copyright and license information must be included.